

GENERATIVE MUSICAL GRAMMAR – A MINIMALIST APPROACH

SOMANGSHU MUKHERJI

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF MUSIC

[Advisor: V. Kofi Agawu]

September 2014

© Copyright by Somangshu Mukherji, 2014. All rights reserved.

ABSTRACT

GENERATIVE MUSICAL GRAMMAR – A MINIMALIST APPROACH

SOMANGSHU MUKHERJI

As an undeniably cultural artifact, music has been subject to humanistic inquiry for centuries. How does this square with the equally ancient, yet conflicting, fascination with music as a scientific object – a fascination that has yielded important insights into the physics of musical sound and, more recently, the biology and evolution of musical behavior? This dissertation develops a cognitive, theoretical answer to this question by considering similar issues in language research, specifically ideas from the Minimalist Program in generative linguistics. In particular, it explores the unique, innate ability of the human mind to compute grammar in music in a manner likewise to that proposed for language by Minimalism. It proceeds from there to argue that the grammatical, musical mind is optimally suited to its various aesthetic functions, such as its ability to create meaning in both language and music. The dissertation makes this argument in two ways; first, by examining a deep, historical and philosophical link between the music theory of Heinrich Schenker and the generative linguistics tradition, and secondly, by using ideas and methods from Minimalism to explore various facets of musical grammar, including its computational structure, its cross-cultural invariance across Western tonal and North Indian Classical music and its ability to govern musical phenomena often considered extra-grammatical, such as musical meaning and rhythm. The dissertation explores these issues with many analytical examples, primarily from the works of Beethoven and Chopin, and from North Indian Classical instrumental performance.

TABLE OF CONTENTS

Abstract	iii
List of Examples	vii
Acknowledgements	ix
Prologue: The Universal Language	1
Part I: Minimalist Music Theory and (Grammatical) Description	
Chapter 1.1: Two Identity Theses for Music and Language	11
Chapter 1.2: Minimalist Musical Grammar: (i) Computational System	200
Chapter 1.3: Minimalist Musical Grammar: (ii) Lexicon	468
Part II: Minimalist Music Theory and (Phonetic/Semantic) Interpretation	
Chapter 2.1: Musical LF, Meaning and Agawu's "Schenkerian Poetics"	524
Chapter 2.2: Musical PF, Meter and Schachter's "Tonal vs. Durational Rhythm"	600
Chapter 2.3: From Description to (Internalist) Interpretation: Chopin Mazurkas, A Case Study	671
Epilogue: An Unanswered Question?	759
Bibliography	766

EXTENDED TABLE OF CONTENTS

Abstract	iii
List of Examples	vii
Acknowledgements	ix
Prologue: The Universal Language	1
Part I: Minimalist Music Theory and (Syntactic) Description	
Chapter 1.1: Two Identity Theses for Music and Language	11
1.1.1: The ‘Musilanguage’ Hypothesis	23
1.1.2: A Minimalist Program for Language <i>and</i> Music	73
1.1.3: Schenker, Humboldt, and the Origins of Generative Music/Linguistic Theory	111
1.1.4: The “Princeton Schenker Project” and GTTM revisited	154
Chapter 1.2: Minimalist Musical Grammar: (i) Computational System	200
1.2.1: The Computational System for Human Music (C_{HM}) – The Input	202
1.2.2: The Computational System for Human Music (C_{HM}) – The Output	224
1.2.3: A Brief Overview of C_{HL} in Linguistic Theory.....	243
1.2.4: The Relation of C_{HM} to C_{HL}	
i. Keiler on <i>Stufen</i> and binary-branching musical trees	327
ii. Pesetsky on cadence and Schenkerian “interruption” forms	365
iii. DOM-raising and parametric differences between Classical and Rock music ..	400
iv. Bare Phrase Structure and Schenkerian reduction	424
v. Empty categories and implied tones	429
vi. Principles of economy and voice-leading parsimony	439
vii. Musical grammar and ambiguity	445
1.2.5: Conclusion: Schenker, Minimalism and the Simplicity of Generative Theory	464
Chapter 1.3: Minimalist Musical Grammar: (ii) Lexicon	468
1.3.1: <i>Rāga</i> as neither “Scale” nor “Tune”	469
1.3.2: Melodic Structure and Hierarchy in North Indian vs. Western tonal music	480
1.3.3: Motivic Hierarchies in <i>Rāga Bihāg</i> : Some Analytical Examples	492
1.3.4: Conclusion: <i>Rāga</i> motives as an instance of a (harmonic) musical lexicon	517
Part II: Minimalist Music Theory and (Semantic/Phonetic) Interpretation	
Chapter 2.1: Musical LF, Meaning and Agawu’s “Schenkerian Poetics”	524
2.1.1: Internalism vs. Externalism in Musical Meaning	527
2.1.2: Logical Form and Internalism in Linguistic Meaning	530
2.1.3. Hanslick and Internalism in Musical Meaning	556
2.1.4. Schenkerian Poetics and its Relevance to Musical LF	568
Chapter 2.2: Musical PF, Meter and Schachter’s “Tonal vs. Durational Rhythm”	600
2.2.1: Schachter’s “Tonal” vs. “Durational” Rhythm	603
2.2.2: Rhythm, Binary Pitch Grammar and Quadratic Hypermeter	605
2.2.3: Durational Interpretation and Linguistic Phonetic Form	627
2.2.4: Conclusion: Musical PF - The Mapping of Structural to Rhythmic Accent?	657

Chapter 2.3: From Description to (Internalist) Interpretation: Chopin Mazurkas, A Case Study	671
2.3.1: Music Analysis as Experiment	671
2.3.2: Chopin: Formalist or Functionalist?	672
2.3.3. Internalism, Generation and Interpretation in Chopin’s Mazurkas	674
2.3.4: Interpretation vs. Interpretability in Tonal Music	721
Epilogue: An Unanswered Question?	759
Bibliography	766

LIST OF EXAMPLES

- 1.1-1. Wh-fronting in English
- 1.1-2. Mozart, Sinfonia Concertante, K. 364/ii: Melody of the theme and its reductions, mm. 8-16
- 1.1-3. Mozart, Sinfonia Concertante, K. 364/ii: Melody of the theme in the exposition/cadenza
- 1.1-4. Phyllotaxis in a sunflower
- 1.1-5. An architecture of C_{HL} according to the Minimalist Program (I)
- 1.1-6. The *Ursatz* (“Fundamental Structure”) in Schenkerian theory
- 1.1-7. Bellini, “Casta Diva” aria from *Norma*: Generation of the first phrase from *Ursatz*
- 1.2-1. Beethoven, Violin Sonata #9 “Kreutzer”, Op. 47/ii: Melody of the theme and its variations
- 1.2-2. Beethoven, Violin Sonata #9 “Kreutzer”, Op. 47/ii: Common tones in the theme/variations
- 1.2-3. Beethoven, Violin Sonata #9 “Kreutzer”, Op. 47/ii: Main theme, chord tones/motives
- 1.2-4. Beethoven, Violin Sonata #9 “Kreutzer”, Op. 47/ii: Tree diagram of theme, mm. 1-8
- 1.2-5. Beethoven, Violin Sonata #9 “Kreutzer”, Op. 47/ii: Voice leading structure, mm. 1-8
- 1.2-6. Beethoven, Cello Sonata #3, Op. 69/i: Harmonic structure of the main theme
- 1.2-7. Beethoven, Cello Sonata #3, Op. 69/i: Structure of transition in mm. 23-24
- 1.2-8. Beethoven, Cello Sonata #3, Op. 69/i: Development of the main theme
- 1.2-9. Beethoven, Cello Sonata #3, Op. 69/i: Voice leading sequence in mm. 119-123
- 1.2-10. Beethoven, Violin Sonata #5 “Spring”, Op. 24/ii: “Cohn Cycle” distribution of keys
- 1.2-11. Beethoven, Violin Sonata #5 “Spring”, Op. 24/ii: Harmonic structure of mm. 38-45
- 1.2-12. Beethoven, Piano Sonata #19, Op. 49 No. 1/ii: Sentence structure of main theme
- 1.2-13. The AP in English: The PSR “AP \rightarrow (AP) A” in tree-diagram form
- 1.2-14. The VP in English: The phrase structure of a VP in tree-diagram form
- 1.2-15. An architecture of C_{HL} according to the Standard Theory
- 1.2-16. VP structure according to X-bar theory
- 1.2-17. X-bar theory: The general XP model, and its NP, VP, AP, and PP manifestations
- 1.2-18. The X-bar sentence: (a) The general model, and (b) actual TPs with and without T
- 1.2-19. The X-bar S’: (a) The general model, and (b) actual CPs with and without overt C
- 1.2-20. The X-bar NP/DP: (a) with determiner, and (b) with construct genitive
- 1.2-21. An architecture of C_{HL} according to the P&P Theory
- 1.2-22. VSO order in Irish and the VP-Internal Subject Hypothesis
- 1.2-23. Word order parameters and Head-to-head movement: (a) Vata (b) French
- 1.2-24. Case marking, the Theta Criterion, and movement in an English sentence
- 1.2-25. Wh-movement and Bounding in English
- 1.2-26. An architecture of C_{HL} according to the Minimalist Program (II)
- 1.2-27. Keiler’s binary-branching tree representation of the *Ursatz*
- 1.2-28. Major/minor triadic relationships represented as “Circle of Fifths” features
- 1.2-29. Scale degrees represented as “Circle of Fifths” features
- 1.2-30. Merge-based derivation of a phrase from Bortniansky’s *Tebe poem*
- 1.2-31. Internal Merge in the full cadence from Katz & Pesetsky (2011)
- 1.2-32. Beethoven, Symphony #9, Op. 125/iv: Interruption form in “Ode to Joy” theme
- 1.2-33. Brahms, Lieder & Gesänge, Op. 32 No. 9 “Wie bist du meine Königin”: Analysis, mm. 6-20
- 1.2-34. Grammatical movement in music: (a) DOM raising vs (b) SUBD raising
- 1.2-35. Mozart, Symphony #35 “Haffner”, K. 385/iii: I-V / IV-II-V-I progression, mm. 1-8
- 1.2-36. Parameters in music: “Circle of Fourths” vs. “Circle of Fifths” settings
- 1.2-37. Parameters in music: A list of examples from Rock music vs Classical music
- 1.2-38. Ambiguity in language seen through tree diagrams
- 1.2-39. Beethoven, Piano Sonata #17 “Tempest”, Op. 31 No. 2/iii: ‘Uninterpretable’ analysis
- 1.3-1. The *thāt* system of *rāga* classification by Vishnu Narayan Bhatkhande
- 1.3-2. *Rāga Bihāg*: Phrase generation
- 1.3-3. *Rāga Bihāg*: Scalar and motivic structure
- 1.3-4. *Rāga Kāmod*: *Ang* structure
- 1.3-5. *Rāga Bihāg*: Tree diagrams of motivic hierarchies: A. Ustad Bismillah Khan
- 1.3-5. *Rāga Bihāg*: Tree diagrams of motivic hierarchies: B. Pandit Buddhadev Das Gupta
- 1.3-5. *Rāga Bihāg*: Tree diagrams of motivic hierarchies: C. Pandit Hariprasad Chaurasia
- 1.3-5. *Rāga Bihāg*: Tree diagrams of motivic hierarchies: D. Ustad Vilayat Khan

- 2.1-1. Tree for “Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo”
- 2.1-2. Beethoven, Violin Sonata #10, Op. 96/i: Apparent mazurka ‘topic’ in subordinate theme
- 2.1-3. Schubert, “Einsamkeit” from *Winterreise*, Op. 89 No. 12, mm. 41-48
- 2.1-4. Bach, Chorale “Ach Gott, wie manches Herzeleid”, BWV 153/ix: Score
- 2.1-5. Bach, Chorale “Ach Gott, wie manches Herzeleid”, BWV 153/ix: Generation of mm. 13-16
- 2.1-6. Beethoven, Piano Sonata #1, Op. 2 No. 1/i: Sentence structure of main theme
- 2.2-1. Tonal rhythm and durational interpretation
- 2.2-2. Tchaikovsky, Symphony #6, Op. 74/ii: Quadratic hypermeter in main theme
- 2.2-3. Tchaikovsky, Symphony #6, Op. 74/ii: ‘Limping’ waltz motive
- 2.2-4. Quadratic hypermeter in Rock: (a) Iron Maiden, and (b) Alice in Chains
- 2.2-5. Quadratic hypermeter in North Indian Classical music: *Pashto tāla*, theme and variations
- 2.2-6. Beethoven, Symphony #5, Op. 67/ii: Phrase expansion from Rothstein (1989)
- 2.2-7. Beethoven, Piano Sonata #17 “Tempest”, Op. 31 No. 2/iii: Harmony in mm. 1-16
- 2.2-8. Beethoven, Piano Sonata #17 “Tempest”, Op. 31 No. 2/iii: Schenker’s graph, mm. 1-16
- 2.2-9. Beethoven, Piano Sonata #17 “Tempest”, Op. 31 No. 2/iii: Durational reduction, mm. 1-8
- 2.2-10. Grammar-to-prosody mapping in English from Katz & Pesetsky (2011)
- 2.2-11. Mozart, Piano Sonata #11, K. 331/i: Main theme, mm. 1-4
- 2.2-12. Mozart, Piano Sonata #11, K. 331/i: Katz & Pesetsky’s reduction of mm. 1-4
- 2.2-13. Mozart, Piano Sonata #11, K. 331/i: Schenker’s graph, mm. 1-4
- 2.2-14. Well-Formedness analysis of syncopation by Lerdahl and Jackendoff (1983)
- 2.2-15. Brahms, Variations on a Theme of Haydn, Op. 56b: Metrical deletion in the theme
- 2.2-16. Mozart, “Champagne” aria from *Don Giovanni*, K. 527: Italian *Accento commune*
- 2.3-1. Chopin, Mazurka in C major, Op. 67 No. 3: Score
- 2.3-2. Chopin, Mazurka in C major, Op. 67 No. 3: Harmonic generation, mm. 1-8
- 2.3-3. Chopin, Mazurka in C major, Op. 67 No. 3: Interpreting the theme
- 2.3-4. Chopin, Mazurka in C major, Op. 24 No. 2: Pitch structure, mm. 21-36
- 2.3-5. Chopin, Mazurka in C major, Op. 67 No. 3: Metrical interpretation, mm. 1-8
- 2.3-6. Chopin, Mazurka in C major, Op. 24 No. 2: Metrical interpretation, mm. 5-20
- 2.3-7. Chopin, Mazurka in C major, Op. 7 No. 5: Tonal generation, mm. 5-20
- 2.3-8. Chopin, Mazurka in C major, Op. 7 No. 5: Metrical interpretation, mm. 5-8
- 2.3-9. Characteristic mazurka rhythms from Downes’ Grove entry “Mazurka”
- 2.3-10. Common mazurka rhythm from Rosen (1995)
- 2.3-11. The ‘Generic’ Mazurka
- 2.3-12. Rhythmic structure in Chopin Mazurkas Op. 7 No. 5 & Op. 67 No. 3
- 2.3-13. Chopin, Mazurka in C major, Op. 24 No. 2: Rhythmic structure, mm. 5-36
- 2.3-14. Chopin, Mazurka in F major, Op. 68 No. 3: Tonal generation, mm. 1-8
- 2.3-15. Chopin, Mazurka in F major, Op. 68 No. 3: Surface structure, mm. 1-8
- 2.3-16. Chopin, Mazurka in F major, Op. 68 No. 3: Prospective/Retroactive trees
- 2.3-17. Chopin, Mazurka in F major, Op. 68 No. 3: Alternative derivation of mm. 1-8
- 2.3-18. Chopin, Mazurka in F major, Op. 68 No. 3: Metrical analysis, mm. 1-8
- 2.3-19. Chopin, Mazurka in F major, Op. 68 No. 3: Rhythmic structure, mm. 1-8
- 2.3-20. Chopin, Mazurka in G major, Op. 50 No. 1: Tonal generation, mm. 1-4
- 2.3-21. Chopin, Mazurka in G major, Op. 50 No. 1: Voice exchanges, mm. 3-4
- 2.3-22. Chopin, Mazurka in G major, Op. 50 No. 1: Tonal generation, mm. 1-16
- 2.3-23. Chopin, Mazurka in C major, Op. 67 No. 3: B-section generation, mm. 33-40

ACKNOWLEDGEMENTS

Writing the acknowledgements for any large project such as this is a fateful event. It is the action that acknowledges the conclusion of the journey, the writ that summons the end of the mission. And if the journey has been a long one, one might have trouble accepting this

This certainly is the case for me, since this project really began some 20 years ago. It was during the period of my life, sometimes referred to as one's "salad days", when I came to realize that I wanted a life in music. This was when I was growing up in India, where the prospects for becoming a professional musician were not promising, certainly not for someone who wanted to make a career as a Western Classical concert violinist as I did. Despite the occasional despair to which this led, these early years were marked by several significant experiences, and many pleasant memories as well. I remembering hearing Itzhak Perlman play the Tchaikovsky violin concerto at the Indira Gandhi Indoor Stadium one time, and managing to finally acquire the score to that piece a few months later so that I could try my hand at playing it too – which was no mean feat, given that obtaining a score like that meant having it shipped from Boosey and Hawkes in London, and in British currency at that, which the rupees from the monthly allowance my parents gave me did not easily cover. But the most significant experience for me during this time is without doubt my initiation into the world of Indian Classical music. This is because this changed my experience with music from one of mere performance, to that of intellectual, and even spiritual, engagement – for learning two musical idioms simultaneously not only gave me a deeper appreciation for the music itself, it also sparked my curiosity about how different idioms such as these might be related, and in what all of us have in common, as musical beings. Perhaps my engagement with such questions in those early years did not amount to much more than an invocation of that age-old belief, best captured in the poet Henry Longfellow's declamation, that music is the "universal language" of humankind. But it did help me make the decision to dedicate my life, or at least the next several years of it, to exploring these topics – and this culminated, ultimately, in the present dissertation.

So, this dissertation is about music and its connection to human nature, and why this makes it language-like. This is also a specifically music-theoretic exploration of these questions. But its long and tortuous path to completion, which has not been without a few mis-steps, and even a few mishaps, bears the mark of all the other subjects I have encountered along the way – including philosophy, cognitive psychology, neuroscience, evolutionary biology, and linguistics. And needless to say, this project also shows the influence of all the people I have been privileged to work with along the way too. So even though I now face the bitter task of acknowledging the end of this journey, it is a task that is tempered with the pleasant prospect of being able to thank the countless kind souls who made this journey possible in the first place.

The petty constraint of space, however, makes it impossible for me to acknowledge more than just those individuals who have directly facilitated this project in recent years. So, I will start by thanking those teachers and mentors who helped me begin this project even before I started writing it officially as a graduate student at Princeton. I would like to begin by thanking my undergraduate philosophy and psychology teachers at Oxford – Paul Azzopardi, Fiona Ellis, Peter McLeod, Paul Snowdon, and especially Bruce Henning, who supervised my undergraduate research project, and Kim Plunkett, who was my unofficial mentor at Oxford, and who also wrote several letters of recommendation on my behalf (including one that helped me get into Princeton). I would also like to thank Carol Krumhansl, who co-advised my undergraduate research project, and in whose lab at Cornell University I worked for one delightful summer. Finally, my pre-Princeton work would not have been possible without the support and kindness of Emery Brown, and Marc Hauser, whom I sincerely thank for all their help.

Of course Princeton itself has been the location for most of my pleasant learning and writing experiences recently, and the group of people who deserve the first words of thanks for this are undoubtedly the unsung heroes of the third floor of the Woolworth building: Marilyn Ham, Cindy Masterson, Greg Smith, and Kyle Subramaniam. It is easy to forget the indispensable role these individuals play in just making

life in the Princeton music department manageable – so a sincere thank you to all of you. I must make a special mention of Greg Smith in this regard, who over the years must have sent at least 500 letters of recommendation from my referees on my behalf, especially during that long, hapless experience that constitutes looking for an academic music job in the 21st century.

Turning to the Princeton music faculty itself, I would like to express my heartfelt gratitude to Peter Jeffery, Noriko Manabe, Simon Morrison, and Robert Wegman in the musicology department for all their support and encouragement, and especially to Wendy Heller for her superb guidance over the years as Director of Graduate Studies. I have been exceptionally privileged to have been able to take classes with and teach for (sometimes both) all the members of the Princeton composition faculty as well – so another heartfelt thanks to Paul Lansky, Steve Mackey, Dan Trueman, Dmitri Tymoczko, and Barbara White, especially Paul and Barbara in this regard for their help and support during the post-graduate job hunting process. (And Paul – best wishes for your retirement!)

One individual who is not officially on the Princeton music faculty, but who played a formative role in my first year as a student here, is *tabla* maestro Zakir Hussain. My very first semester here was spent assisting him in teaching a large introduction to Indian music course – in fact, I received my teaching certification from Princeton's McGraw Center in order to do this even before I had received my official Princeton student ID. This was certainly a baptism by fire into graduate student life, and it also allowed me to revisit my childhood experiences with Indian music, particularly when I had the chance to accompany Ustadji by playing *lehra* for him on the violin. So, a heartfelt thanks to Zakir Hussain for that experience – which I hope will continue to be the first of many collaborations to come.

Despite the blessing of being able to work with the fantastic scholars and musicians just named, my Princeton experience would have been much poorer if it were not for the fair share of wonderful friendships with other graduate students that I formed along the way, both within the music community and without it. In the case of the latter, Patrick Purdon and Nikolai Slavov have been two of the best buddies a guy could ever want. But a special word of gratitude, affection, and thanks is owed to my partners-in-crime in the (tiny) Princeton graduate music theory community, viz. Jeffrey Levenberg, Christopher Matthey, and Daniil Zavlunov. Now that my time at Princeton has come to an end, I hope we will all be able to stay in touch, and that you will all prosper in your future lives as well.

It is one thing, however, to be a student of a certain discipline, and another thing altogether to build a career in it. I have had one of the richest student experiences one could desire, both as an undergraduate at Oxford, and as a graduate student at Princeton. But for me to develop, and build a career, as a professional music theorist, would have been a futile endeavor, were it not for the support of a number of esteemed scholars in the field, who have given me guidance, encouraged me, and shown enthusiasm for my work over the years, which has been indispensable especially given the absence of an official music theory program at Princeton. To this extent, I cannot but express my deepest appreciation and gratitude for the support of Richard Ashley, Poundie Burstein, Richard Cohn, Robert Gjerdingen, David Huron, Allan Keiler, Steve Larson, Peter Manuel, Elizabeth Margulis, Panayotis Mavromatis, Robert Morris, William Rothstein, Frank Samarotto, Janet Schmalfeldt, Joseph Straus, and David Temperley. Poundie and Joe deserve a special mention in this regard for reading large parts of this dissertation and offering me helpful feedback, and in Joe's case, for even writing me a letter of recommendation on a tight deadline. Janet Schmalfeldt allowed me to audit her Schenker class at Harvard in 2004, which was the first music theory course I had ever taken. Despite my severe lack of preparation for that course, she patiently went through all my graphs, and even read my final paper for the course. The positive impression this had on me helped greatly with my transition into life as a music theorist, for which she deserves full credit.

I did not mention one particular name in the above list, and this is because this person deserves a separate mention all to himself. Fred Lerdahl was the first music theorist I ever interacted with, when I was still a student in India. Over the past 15 years, he has always been a source of support and inspiration, which has meant much to me particularly because his work on connections between music and language has not only been revolutionary, but because it is also in many ways the basis for my own work (albeit not always for positive reasons!). My decision to turn down an offer from Columbia to be Fred's doctoral

advisee, so that I could go to Princeton instead, is, and will remain, one of the hardest decisions I have had to make in my life. I sincerely hope that we will be able to stay in touch, and ultimately work together one day.

Which brings me finally to the four individuals with whom I have worked the most closely during the dissertation writing process. David Pesetsky very kindly agreed to serve on my dissertation committee initially, but could not participate in the end due to scheduling conflicts. Nevertheless, being able to bounce ideas off of him from time to time over the past few years has been a real privilege for me, and has provided me with a perspective not readily available from within the music theory community. As one of the leading generative linguists of the day, David is one of the few people in the world who really understands the issues at stake in researching connections between music and language, in the way I have been interested in doing since childhood. So I am really looking forward to being able to work with him more than our limited dissertation-related interactions have made possible so far, and I hope this will lead soon to a wider collaboration between music theorists and linguists, with exciting and far-reaching implications for the study of music-language relations, and their connection to human nature.

I first met Matthew Brown when he came to give a talk at Princeton in 2009. But what was meant to be a subsequent, brief, discussion with him about his talk, soon turned into a marathon two-day event, during which we spent hours sketching voice-leading progressions on napkins in Small World Café on Nassau Street, and discussing everything from delta Blues, the ideas of Willard van Quine, and the differences between attacks on civil liberties in the US versus the UK, to his (and to a large extent my) favorite subject, the scientific basis of Schenkerian theory. And in the process we discovered a lifelong friendship – particularly given our maverick status as two of the few theoretical scientists in the music theory community. So his agreeing to serve on my committee has been a boon for me, and I have profited from his advice and feedback on much more than just the dissertation. Needless to say, I hope our friendship and collaborations will continue for many years to come, and will lead to many more projects that explore the scientific aspects of music theory.

Which leaves me with the names of two people who have not only been my teachers and mentors at Princeton for the last several years, but who have been essentially father figures to me – Scott Burnham and Kofi Agawu. Kofi and Scott have seen me from my earliest days as a student of cognitive science transitioning into the world of academic music scholarship, to my current status as a, hopefully passable, professional music theorist. So, it would not be excessive to say that I owe essentially my entire existence as a music theorist to these two individuals. I think Scott would agree that our intellectual interests really only overlap when it comes to certain historical figures in an essentially German world of ideas, figures such as Goethe, Kant, Schenker, A. B. Marx, Hugo Riemann, Wilhelm von Humboldt and so on. However, my relationship with Scott has been much more than merely an esoteric intellectual one. As I said, he has been like a father to me, especially during times of frustration or financial hardship, which are almost inevitable realities for a junior scholar in today's tough economic times. His endless words of wisdom on matters both music-theoretic and otherwise, his tireless assistance with professional concerns (especially job applications and recommendation letters), and constant support not just for me, but the entire graduate student community, makes him not only indispensable for Princeton and the world of music at large, but also someone for whom I have come to develop the deepest respect and fondness. I was hugely honored when he asked me to do the musical examples for *Mozart's Grace*, and I hope we will continue to work together, and will continue to share our joys and sorrows together as well, both professional and personal.

And finally Kofi Agawu, my advisor. I will restrict myself to only a few sentences here, for to acknowledge Kofi properly would probably require a dissertation in and of itself. All I can say is that I am extremely lucky to have found a mentor, advisor, friend, and hopefully future collaborator and colleague in someone like Kofi. It is rare enough to find an advisor who shares so many of your intellectual interests as Kofi and I do – given our mutual regard for Schenkerian theory, the theory and analysis of non-Western musical idioms, and the study of music-language relationships (especially issues of musical meaning and its relation to linguistic signification) – although I am yet to get Kofi sufficiently interested

in matters of music cognition! But it is rarer still to have an advisor who shares so much of your personal background and history as well. Having been brought up in the ‘third world’ (in Ghana), subsequently educated in the UK, and then having established a career as a professional music theorist in the US, Kofi’s life has been an example for younger scholars like me to follow. And this makes Kofi’s influence on me not just a professional or intellectual one, but a personal one too. His unwavering support for me over the years, and his careful and sympathetic response to my work, even when in the crudest stages of development, is what has allowed me to grow into the scholar I am today, and in many ways the person I am today too. And even though I am sure our mutual interests will help us stay in touch, just the role he has played in my life over the past years at Princeton is enough to make me eternally grateful to him, and hopeful that we will continue to have a productive partnership for years to come.

Before I end, I should also mention that my studies, and especially dissertation work, at Princeton would not have been possible without the support of the American Council of Learned Societies, the Charlotte E. Procter Foundation, and Professor George W. Pitcher and the Edward T. Cone Foundation, all of whom honored me with fellowships, for which I am deeply grateful.

And finally to the three individuals without whom I would just not be: Ma, Baba, Alex – thank you.

Prologue: The Universal Language

“Music is the universal language of mankind, poetry their universal pastime and delight.”¹

- Henry Wadsworth Longfellow

“Music is the language of the spirit. Its melody is like a playful breeze which makes the strings vibrate
with love”²

- Kahlil Gibran

The metaphor that music is a language is an old and resilient one. It is a metaphor that has not only captured the imagination of poets, as the above quotations illustrate, but also a host of other scholars and sages who have wondered about music, and who have used this metaphor in their quest to understand it. This dissertation continues that endeavor.

But thinking about music in linguistic terms is of course just one of the many ways humans have attempted to understand this elusive phenomenon. For example, much of the history of ideas can be seen as an attempt to understand reality according to the laws of physics, and musical reality has been no exception to this either. Its various patterns and intricacies, several of which we will explore in this dissertation, have led many to conceive of music as a case study in the elegant designs of a physical universe – a universe that is seemingly incomprehensible, to paraphrase Albert Einstein, but which can be made comprehensible with equally elegant mathematical formulations.³ Indeed, as far back as a couple of thousand years ago, the Greek philosopher Pythagoras declaimed the shared physical basis for the musical

¹ Henry Wadsworth Longfellow (1883) *Outre-Mer: A Pilgrimage Beyond the Sea* (Rev. Ed.). Boston: Houghton, Mifflin and Company, page 197.

² Kahlil Gibran (1995) *The Eye of the Prophet*. (Trans. Margaret Crosland from *L'Oeil du Prophete* (1991)). Berkeley: Frog Books, page 97.

³ Einstein, no mean musician himself, is believed to have said, “The most incomprehensible thing about the universe is that it is comprehensible.” See Dennis Overbye, “The Most Seductive Equation in Science: Beauty Equals Truth.” *New York Times*, March 26, 2002. Also, “Einstein Letter on God Sells for \$404,000.” *New York Times*, May 17, 2008.

and astronomical aspects of the universe in his doctrine of the “Harmony of the Spheres”, according to which musical harmonies are governed by the same, simple mathematical ratios that seem to govern the speeds with which celestial bodies circle the earth. And the idea that music is subject to the laws of physics is borne out by the fact that certain physical conditions need to obtain for music to even exist – air needs to vibrate for musical sounds to occur, electrical currents need to propagate down hair cells in ears for humans to hear those sounds, and bones and muscular tissue need to move for people to perform those sounds. (That music requires human agency, or even the presence of sound, might be debated by some, especially those who find music in silence, or in the wind and in waterfalls, or in the sounds made by non-human creatures like birds and whales. Clearly the very definition of “music” is at stake here – but I will not belabor the point because the physical medium of sound, and the physical processes involved in human sensation and action, *are* clearly necessary for a large proportion of what we *do* consider music, viz. that which we play and sing, and hear in concert halls, and during the ceremonies, sporting events and the myriad other activities that make up human life on this planet.)

A peculiar property of music though, which distinguishes it from the other objects of natural study, is that in addition to being a manifestation of physical reality, it is also a manifestation of *mental* reality. To the extent that music requires human agency, it needs *thinking* individuals, who can acquire musical knowledge, use this knowledge to express novel musical thoughts, and respond to the meaningful musical thoughts of other individuals. To the extent that such thoughtful engagement with music has come to characterize all human cultures, across the world and since the beginning of human history, the mental reality of music is one of the things that has defined who we are as a cultured species – it is one of the things that defines human nature.

But if the very existence of music suggests the reality of a non-physical, mental universe, this poses a challenge for the grand project in intellectual history to understand all reality in physical terms. In consequence, many have attempted to explain mental reality away, by reducing it to some form of material reality. But there have been others who have been skeptical of this possibility, given that the principles that govern physical and psychological reality might not be the same – and might even be

incompatible or contradictory. The physicist and Nobel laureate Eugene Wigner once remarked how we might be unable to explain the mental universe because we do not even understand how the physical universe works, or more specifically why mathematics is able to describe it so elegantly – a thought that repeats Einstein’s sentiments about the most incomprehensible thing about the universe being its comprehensibility:

“A much more difficult and confusing situation would arise if we could, some day, establish a theory of the phenomena of consciousness, or of biology, which would be as coherent and convincing as our present theories of the inanimate world. Mendel’s laws of inheritance and the subsequent work on genes may well form the beginning of such a theory as far as biology is concerned. Furthermore, it is quite possible that an abstract argument can be found which shows that there is a conflict between such a theory and the accepted principles of physics. The argument could be of such abstract nature that it might not be possible to resolve the conflict, in favor of one or of the other theory, by an experiment. Such a situation would put a heavy strain on our faith in our theories and on our belief in the reality of the concepts which we form. It would give us a deep sense of frustration in our search for what I called “the ultimate truth.” The reason that such a situation is conceivable is that, fundamentally, we do not know why our theories work so well. Hence, their accuracy may not prove their truth and consistency.” (Wigner (1960): 13-14)

The above suggests that the very reality of music demands a special kind of inquiry – an inquiry that accounts for its psychological nature, even if one ultimately conceives of it as an object of natural science, as many have as well. And this brings us back to the music-language connection. For not only does language, like music, manifest both physical reality (e.g. in its speech-based dependence on physical sound) *and* psychological reality (e.g. in the way it informs how individuals express thoughts and respond to the thoughts of others), a certain paradigm in the study of language, viz. the field of generative linguistics, influenced by the ideas of the linguist Noam Chomsky and his colleagues, has been particularly associated with the above mode of inquiry suggested for music – i.e. one that explores certain aspects of psychological reality while situating this study within the broad purview of the natural sciences. The most recent research project in generative linguistics, the Minimalist Program, is specifically devoted to understanding the mental reality of human language, but in terms of general scientific principles that, as the linguist Cedric Boeckx says:

“...are uniformly considered essential items on the agenda of theoretical physicists ... That is, minimalists endorse the belief (held by all major proponents of modern science, from Kepler to Einstein) that nature is the realization of the simplest conceivable mathematical ideas.” (Boeckx (2006): 8)

In this light, a Minimalist approach to the study of music – and specifically the *generative* study of music, as influenced by ideas from generative linguistics – would not only be a way to continue exploring what Longfellow referred to as the “universal language” through the lens of the age-old music-as-a-language metaphor, it would also help situate this exploration within the grand agenda of the natural sciences that has so frequently informed the central narrative in the history of ideas. And this is exactly what this dissertation aims to do.

So this dissertation undertakes a comparative study of music and language from the perspective of a Minimalist theoretical science – one that is akin to physics, but also distinct from it in the way it accounts for, and is influenced by generative proposals about, the human mind. One of the reasons for why such a Minimalist approach to music is novel is because of its focus on musical and linguistic *grammar*. As is evident from its name, the first part of the dissertation – “Minimalist Music Theory and (Grammatical) Description” – describes this term in detail. In short, “grammar” is just a theory about the unique psychological aspects of music and language, particular those aspects that make music and language information processing, or *computational*, systems.

Now, a theoretical scientific approach, as exemplified by the current Minimalist one, is by no means the only one that can be, or has been, taken by those interested in the comparative study of music and language. So, the first chapter in Part I of the dissertation, chapter 1.1, begins by reviewing some of these alternative approaches. These approaches can be broadly classified into two categories. The first approach, often taken by ethnomusicologists and cultural anthropologists, is one in which the shared role of music and language within *human society* is taken to be an important factor in their comparison. In contrast, the second approach, often taken by psychologists and neuroscientists, is one that focuses on the shared *biological* aspects of music and language, using methods derived from the experimental, rather than the theoretical, sciences. What the chapter attempts to show is that these approaches ultimately lead to a rejection of any deep connection between music and language, and therefore a rejection of the music-as-language metaphor, because these approaches seem unable to account for the crucial grammatical, and

specifically computational, aspects of music and language. And given this dissertation's commitment to continuing the comparison of music and language, this is the reason why it rejects the anthropological and experimental scientific approaches to comparing music with language in favor of a Minimalist approach to (generative) musical grammar – hence the title of the dissertation.

As a result of this Minimalist approach to comparing music and language, chapter 1.1 makes two rather novel claims as well. First, it claims that music and linguistic theory are essentially the same discipline. This is based on several historical and technical convergences between the generative approach in linguistics, and the generative approach inherent in certain paradigms of music-theoretic scholarship, which the chapter discusses. In particular, some of the ideas within current music theory, particularly those belonging to the Schenkerian tradition, and some of the current debates within music theory, reveal a very similar, though implicit, understanding of music of the kind linguists have of language – even though this understanding has been absent from explicit music theoretic discourse. This implicit similarity is why music and linguistic theory seem to be essentially the same discipline. However, these convergences are not the result of any conscious collaboration between linguists and music theorists – so, chapter 1.1 also claims that music and language are identical, as the only explanation for why music and linguistic theory appear to be identical too, which has, in turn, compelled music theorists and linguists to develop similar conceptions of music and language, albeit independently of each other. These two claims are the two identity theses for music and language that give the chapter its title.

Chapter 1.2 then proceeds to examine the evidence for the thesis that music and language are identical by exploring several points of identity in musical and linguistic structure. It does this by showing how the components of linguistic grammar, as described by generative linguists, are mirrored in the components of musical grammar, some of which have already been described by certain music theorists, especially those working within the Schenkerian tradition. In the process, the chapter also explores aspects of musical grammar that have often been ignored or overlooked in music-theoretic scholarship, but which a Minimalist, generative perspective helps throw new light on. Some of these aspects of musical grammar include its lexical input, its phrasal output, its constituent structure, its transformational

and parametric organization, and its economical architecture. However, and importantly, the chapter – and the dissertation in general – falls short of proposing a full-fledged generative theory of music, mainly because such a project, apart from being beyond the scope of a doctoral dissertation, is beyond the scope of any individual researcher at this point. This is because such an endeavor would involve the examination of music across multiple idioms found in the world, if it is to be a general scientific theory of music, and if it is to have all the psychological richness, biological plausibility, and philosophical depth that current generative *linguistic* theories have – of the kind that have already been proposed by Noam Chomsky and his colleagues, as mentioned above. Music theory has not even begun to approximate this level of theoretical depth yet, since current music theory never even asks questions of the kind current generative linguistics asks about language – related questions of a biological or philosophical nature being essentially absent from the discussion.

The discussion of the next chapter, chapter 1.3, does however allow the dissertation to at least take some steps towards a more comprehensive generative theory of musical grammar, by extending the discussion of the previous chapters – most of which focuses on Western music – to the idiom of North Indian Classical music. Such cross-idiomatic extensions and comparisons are part and parcel of the trade of generative linguistics, and they also raise the possibility that the model of grammar under consideration is really a universal one. So by extending the discussion of musical grammar to another idiom in chapter 1.3, the dissertation takes an important step toward developing a broader Minimalist theoretical science of music of the kind already implicit in generative linguistics.

One of the major reasons for why the music-as-language metaphor has been so persuasive for so many thinkers is because music is unique in the language-like ability it gives humans to express themselves, and to understand each other. The very use of both music and language as mediums of expression and interpretation implies that they are inherently meaningful, and this allows them to facilitate communication between humans too, through speech and gestures (i.e. sign language). This, in turn, has led music and language to become indispensable components of human culture. This is one of the reasons

why the music-language nexus has been of particular interest to more culturally-oriented scholars, often those working within the humanities or social sciences. So, being able to account for the expressive and interpretive aspects of music and language seems to be of some importance, especially for those research paradigms that claim an interest in connections between music and language.

Even though Minimalism sees itself as a theoretical science, and specifically a theoretical natural science, the expressive and interpretive aspects of language play as important a role in its theorizing about language as it does in more humanities or social-scientific approaches to the subject. But given its grammar-centric approach to language, Minimalism understands these aspects of language in terms of how the grammatical components of the mind interface with certain extra-grammatical components, including the one responsible for meaning (i.e. the semantic or “conceptual-intentional” component), and the one responsible for speech and gestures (i.e. the phonological or “sensorimotor” component). In particular, Minimalism proposes that these extra-grammatical components *interpret* the outputs of the grammatical component, “interpretation” being a term that the dissertation will discuss in detail. This emphasis on interpretation has allowed Minimalist linguists to propose a robust psychological science of semantics and phonology in language too.

Given the importance of expression/interpretation in music, no theory of music would be complete without a consideration of this aspect of music, and the Minimalist approach to music being proposed by this dissertation is not exempt from this requirement either. Moreover, the two identity theses of music and language being defended by this dissertation would be significantly weakened if a Minimalist model of music were unable to account for this aspect of music, in the way Minimalist linguistics accounts for linguistic meaning and phonology. But the novelty of the Minimalist approach even in its consideration of musical grammar – which occupies the whole of Part I of the dissertation – makes thinking about the extra-grammatical, expressive/interpretive aspects of music in Minimalist terms quite a difficult task.

However, as is the case with much of the grammatical discussion in Part I, there are ideas implicit in extant music-theoretic scholarship, especially within the Schenkerian tradition, that contain clues as to how one might countenance *musical* interpretation in Minimalist terms. This, then, is the focus of Part II of the dissertation, which is titled “Minimalist Music Theory and (Semantic/Phonetic) Interpretation” – although the novelty and difficulty of this enterprise leads to Part II being much shorter than Part I.

The first chapter in Part II, chapter 2.1, focuses on the ‘semantic’ aspects of music, and how this can be thought of in Minimalist, grammatical terms. Some ideas about the connection between musical grammar and musical meaning can be found in proposals made by the music theorist Kofi Agawu, which builds on both Heinrich Schenker’s thoughts on musical meaning, and some of Roman Jakobson’s work on linguistic meaning within the structuralist tradition. Chapter 2.1 develops these ideas from a Minimalist perspective into an account of semantic interpretation in music.

A similar set of ideas about the connection between musical grammar and musical ‘phonology’ is developed in chapter 2.2. The discussion in this chapter is restricted to the issue of musical rhythm, as an analogue to speech rhythm (i.e. prosody) in language. Speech rhythm is not only an important aspect of phonology in language, it plays a major role in the Minimalist description of how the phonological system in language interprets the outputs of the grammatical system. So, a discussion of musical rhythm might not only help us understand the phonological aspects of music, it might also help us understand how musical phonology interprets the outputs of musical grammar – thus providing an account of expression/interpretation in music that is not only crucial for a comprehensive theory of music, but which also parallels the linguistic account of phonetic interpretation. Some interesting ideas about the connection between musical grammar and rhythm are already implicit in the Schenkerian music theorist Carl Schachter’s work on musical rhythm, so chapter 2.2 builds on these ideas to develop a Minimalist account of rhythm and phonetic interpretation in music.

This brings us to the final chapter of the dissertation, chapter 2.3, which reconsiders the notion of analysis in music. Music analysis is something many music scholars engage in, often with the aim of illustrating interesting motivic, rhythmic, formal, or other details in a piece of music. Within the scope of

a theoretical exploration of musical structure, such as the Minimalist one being proposed here, such analytical illustrations can serve as evidence for the hypotheses made by the theory – and if the theory is a scientific one, such analyses can even be seen as experimental confirmations of the theory. All of the discussions in Part II of the dissertation, on musical meaning and rhythm (which are the kinds of things that musical analysis often focuses on), therefore have implications for the wider Minimalist grammatical theory being developed in this dissertation. That is, the phenomena of semantic and phonetic interpretation in music, which is essentially equivalent to what one does in the analysis of a musical passage, can be thought of as experiments that confirm or falsify the hypotheses proposed by a music theory. So, chapter 2.3 focuses on these issues, by exploring how the theoretical ideas proposed in the dissertation can inform musical analysis, particularly through the metaphor of analysis as (scientific) experiment. The chapter pursues this goal through an analytical investigation of some mazurkas by Chopin.

Part I

Minimalist Music Theory and (Grammatical) Description

Chapter 1.1

Two Identity Theses for Music and Language

There seems to be one overarching reason for why so many people over the ages – from poets and musicians, to philosophers and scientists – have sensed an affinity between music and language. To my mind, this seems to be the fact that both music and language play a decisive role in making us who we are, as a thoughtful, expressive, cultured species. In other words, they both seem to be indispensable components of *human nature*. Our ability to think and communicate with each other, which are certainly central components of who we are, would be impoverished without music and language. This is reinforced by the fact that music and language are the only communicative systems founded primarily on sound, albeit with a secondary realization in sight (script and signing for language, notation for music). Furthermore, not only do music and language facilitate communication, they allow us to communicate in structured ways, by organizing our thoughts, and their physical manifestation as musical/speech sounds, into temporally-ordered patterns, i.e. phrases and sentences. This of course gives us a unique ability to structure our thoughts and communicate in time – but also *over* time since, as systems of communication, music and language have allowed humans to interact socially, create elaborate cultural and aesthetic traditions/idioms, and, ultimately, develop a history.

So, this dissertation is aimed at exploring the above phenomenon – the connection between music and language, and human nature. But as one might suspect, such a project entails grappling with complicated metaphysical issues regarding what music and language are, and what it means, more generally, to be human. Since so many thinkers have provided perspectives on the music/language nexus, equally diverse definitions have been proposed for the above terms too. Therefore, my goals for this dissertation must necessarily be simple and straightforward. I hope to defend the above belief that music and language share an intimate connection, as aspects of human nature – but only under a specific, psychological definition of these terms. Under this definition, I will also argue that music and language are not just

related but *identical*. To demonstrate this, I will compare some resilient issues in music theory with analogous ones in linguistic theory, specifically those within the Minimalist Program in current generative linguistics. This suggests an intriguing identity not only between music and language, but also between their respective theories – hence the two identity theses for music and language from which this chapter gets its title.

To begin, let me illustrate just how compelling it has been to describe music and language within the context of human nature. In fact, the belief that their ‘humanness’ makes music and language the universal cornerstones of civilization has often become a point of focus especially for those interested in the anthropological study of musical and linguistic societies. For example, in his book *How Musical is Man*, written within the, then relatively young, field of ethnomusicology, John Blacking famously said:

“The question, “How musical is man?” is related to the more general questions, “What is the nature of man?” and, “What limits are there to his cultural development?” It is part of a series of questions that we must ask about man’s past and present if we are to do anything more than stumble blindly forward into the future. ... There is so much music in the world that it is reasonable to suppose that music, like language and possibly religion, is a species-specific trait of man. Essential physiological and cognitive processes that generate musical composition and performance may even be genetically inherited, and therefore present in almost every human being. An understanding of these and other processes involved in the production of music may provide us with evidence that men are more remarkable and capable creatures than most societies ever allow them to be. This is not the fault of culture itself, but the fault of man, who mistakes the means of culture for the end, and so lives *for* culture and not *beyond* culture.” (Blacking (1973): 7)

Blacking’s assertion here of a connection between human musicality and human nature is clear. It is also clear that he thinks that music and language are related by being jointly “species-specific traits of man”, even though he does not explicitly equate music with language.

However, what’s not so clear in this statement is *why* language and music are species-specific traits of humankind, and the way Blacking answers this question later is very interesting not only because of the way it reveals his own peculiar understanding of human musicality/linguisticity and human nature but because his answer reflects a broader ideological approach to these issues, which have dominated the field of ethnomusicology ever since. Realizing that asking “how musical (or linguistic) is man” requires

dealing with the inherent metaphysical and terminological problems of what music and language *are* to begin with, Blacking, focusing on the musical aspect of this question, states that music is “humanly organized sound” (Blacking (1973): 10). His point is essentially simple: wherever there is music, there are people. In other words, music needs human agency to exist – it exists only because people behave in certain ways, whether as individuals or as part of society. So, music is just sound organized in ways that reflect patterns of human behavior, both individual and collective.¹

This point might seem trivial, but actually it has immense rhetorical force, especially when one considers that Blacking had political aims in defining music, directed specifically against those kinds of music scholarship that reject the musical status of certain cultures, and the musical aptitude of certain individuals, based on classist or racist convictions about the superiority of Western art music. In light of his musical experiences with the Venda of southern Africa, Blacking was particularly interested in emphasizing the complexity and sophistication of the indigenous music of various non-Western peoples, so as to reject the dogma of Western superiority, the belief that Western ‘art’ music is better than, say, African ‘folk’ music. “All music is folk music”, he says, “in the sense that music cannot be transmitted or have meaning without associations between people” (Blacking (1973): x).

The field of ethnomusicology had emerged a few years prior to Blacking’s text in the 1950s and as an alternative to the earlier “comparative musicology” of scholars like Erich von Hornbostel, Curt Sachs, Carl Stumpf and others from the so-called Berlin School. Comparative musicology was targeted by the new ethnomusicology as being founded on the questionable belief that the study of non-Western music can only begin when its relation to Western (art) music has been established, so that a “comparative musicology” really means establishing how a non-Western idiom is different from Western music. This implies a Eurocentric view of the musical universe, and an explicit ‘Other-ing’ of the non-West, which the later ethnomusicology came to see as racist, imperialist and outdated. Therefore, dismissing the centrality of Western music in the musical universe or, more accurately, asserting the significance of the non-West, not as the ‘Other’ but as a self-standing entity in a universe in which “all music is folk music”,

¹ Other ethnomusicologists such as Simha Arom have repeated this assertion, e.g. Arom (2001): 27.

became a non-trivial rallying cry in the agenda of the new ethnomusicology, especially for those who wanted to distinguish it from Western musicology (Bohlman (1992): 120-123, Brown, Merker and Wallin (2001): 17-18).

In light of the above, Blacking intentionally defines music in a way that includes not only the musical practices of the non-West but also the latent abilities of non-professional musicians and audiences in the West, upon whose very abilities the existence of Western musical culture depends, but whose musicality is denied by elitist, capitalist dogma (Blacking (1973): 9, 34-35, 106, 116). But although this egalitarian political goal might be admirable in itself, Blacking makes a larger intellectual statement about what kinds of musical study are acceptable, in light of his critique of comparative musicology, which I will argue is ultimately untenable.

To understand this larger statement, consider the following passage:

“The function of [musical] tones in relation to each other cannot be explained adequately as part of a closed system without reference to the structures of the sociocultural system of which the musical system is a part, and to the biological system to which all music makers belong. ... In order to find out what music is and how musical man is, we need to ask who listens and who plays and sings in any given society, and why.” (Blacking (1973): 30-32, see also 18-21, 49-58, 71-75, and 97-99)

In this passage, Blacking makes it clear that a legitimate study of human musicality must have at least two components. In keeping with his previous claims about the species-specificity of music, in addition to his thoughts about human musicality being genetically inherited and the result of “physiological and cognitive processes”, he states that a study of human musicality must be grounded in a study of human biology. But he also claims that such a study must have a sociocultural component – and the larger problem with Blacking’s position lies in how one defines this term.

On the one hand, a sociocultural system could be the set of formal, i.e. structural, features that characterize *specific* musical or linguistic cultures, rather than those that exist cross-culturally across idioms. Examples of such features are the words in human languages and the order in which they appear

in sentences, which are both features that differ across languages.² On the other hand, a sociocultural system could be the set of functional, as opposed to formal, features that characterize a musical/linguistic idiom. In other words, it could be the set of features that govern how that idiom is *used* within a given culture.

Now, there is no question that even to understand human musicality/linguisticity more generally one has to understand how specific musical or linguistic idioms work. This is why scholars who are interested in developing, for example, relatively abstract cross-idiomatic theories about human linguistic ability value the study of structural features in specific languages, and place great emphasis on the unconsciously-acquired intuitions of native speakers about such structural features, e.g. their intuitions about the grammaticality of particular sentences in that language. Related to this point, even Blacking says that Venda music making requires knowledge of certain unconsciously-acquired conceptual models from which actual melodies are generated, which can only be learned (and therefore understood by scholars) by a deep involvement in Venda society (Blacking (1973): 98-100) – and understanding such conceptual models is undoubtedly important for the larger study of human musicality that Blacking is interested in. This might suggest that Blacking takes a formal approach to defining the connection between music and sociocultural systems, especially since he also claims that he is primarily interested in what music *is*, rather than what it is used for – so that a study of the sociocultural functions of music is necessary only insofar as it can shed light on musical structure (Blacking (1973): 26).

But the fact that Blacking even thinks that the study of the sociocultural functions of music is necessary for a study of musical structure demonstrates the greater compatibility of his position with a functionalist approach to sociocultural systems. To appreciate this, take for example his claim that Venda musicality involves a “deep involvement in Venda society”. This is of course the level of involvement that Venda musicians themselves have in their society, by growing up in that society as native speakers of any language do. But for Blacking, growing up in Venda society also means learning about the intricate

² This point should be obvious but just to clarify, consider that English transitive sentences normally have a Subject-Verb-Object word order, whereas Japanese transitive sentences normally have Subject-Object-Verb orders. For example, compare *Kim ate the apple* with *Kim-ga ringo-o tabe-ta* (literally “Kim apple ate”).

social roles that Venda music and musicians play in that society, so a deep involvement in Venda society essentially means learning about what Venda music means to the Venda themselves.

But does one have to learn about what music means to a Venda musician to understand, especially the abstract properties of, Venda musical structure? Blacking certainly thinks that one does given the importance he places on the sociocultural functions of music in a study of musical structure. So, it seems that Blacking is saying that to truly understand the structure of Venda music one has to essentially be Venda. This is similar to saying that to understand the grammatical properties of sentences in a language not only does a linguist need to value the intuitions of native speakers of that language, s/he also has to be a native speaker of that language him/herself, one who is familiar with all the sociocultural uses of that language. Blacking's approach to this issue is clearly an emic one – following anthropologist Kenneth Pike's famous emic vs. etic opposition – in which the value of a descriptive statement about an idiom or culture depends on whether it is meaningful or of value to a native of that culture. That is, Blacking seems to believe that descriptions of Venda music are of value only if they are meaningful to the Venda musician, i.e. only if they are ways in which the Venda would describe their own music. His emic biases are revealed palpably in the following statement:

“In asking how musical is man, I am obviously concerned with all aspects of the origins of music, but not with speculative origins, or even with origins which a foreign historian thinks he can detect, but which are not recognized by the creators of the music.” (Blacking (1973): 58)

Needless to say, a close involvement with how music is made in a particular idiom – e.g. by paying attention to the intuitions of native practitioners of that idiom – allows one to avoid making the same errors that one might make when making claims about an idiom ‘from a distance’, which was of course a central problem with the earlier Eurocentric models of human musicality that Blacking was criticizing. But the important issue here is whether the native practitioners of an idiom even care about understanding the more abstract or formal properties of their music. In other words, is a study of Venda musical structure even of any value to the Venda musician, whose main interest might lie in just performing it and thus accomplishing the primary task they have learned how to fulfill in Venda society from an early age?

Since the study of a musical idiom is of value essentially to the ‘outsider’, or to one who is separated from the primary activity of performing/making music – since it is essentially a study of knowing *that* the structures of an idiom have *x* properties rather than knowing *how* to create those structures in performance – Blacking’s prescription that only a cultural insider can make valid claims about an idiom seems inappropriate, which is an observation echoed by the music theorist Kofi Agawu in his description of the emic biases in ethnomusicology (Agawu (1990): 229).³

So, one problem with Blacking’s invocation of the importance of sociocultural systems in the study of human musicality is that it imposes a cultural bias on the study of, what he himself claimed to be, a species-specific (and not culture-specific) trait of humankind. But an even bigger problem is what including sociocultural functions implies for an analysis of what music is more generally, as humanly organized sound. For Blacking, the sociocultural functions of music can include a variety of even extramusical phenomena, such as the political functions served by music within a culture. As an example, he mentions the study of a sacred vocal composition by Baroque composer Claudio Monteverdi, which would be incomplete in his opinion if it does not take into account, say, the liturgical framework of the composition or Monteverdi’s services to Vincenzo Gonzaga, the Duke of Mantua (Blacking (1973): 30). These are clearly examples of the religious or political functions Monteverdi and his music served in early 17th century Italian society, so their inclusion in a study of Monteverdi’s music amounts to much more than a study of the formal properties of Monteverdi’s melodies (for which a study of Monteverdi’s knowledge of counterpoint might be more appropriate).

But a study of the various cultural functions of music automatically makes the study of human musicality a study of differences. Not only is there so much music in the world, as Blacking himself noted in his statement at the beginning of this chapter, but all of this music is strikingly varied and diverse – and

³ This just reflects the broader idea that the knowledge of how music is made is separate from the knowledge of what music is – and by extension, how humans are musical. One could also couch this in linguist Noam Chomsky’s distinction between *performance* and *competence*, as I shall do later in the chapter. But for the present it should suffice to say that this distinction between the two forms of knowledge also makes it clear that there is a further distinction between their respective study and between the individuals who possess these two forms of knowledge – so that the greatest poets may not be the greatest linguists, or the greatest composers/performers the greatest music theorists.

this diversity is to a large extent the result of the diverse ways in which music is used in different cultures. To take a few examples, cultures in which music is primarily used to accompany dance is more likely to have meter, as opposed to cultures in which music is used primarily for calm reflection, such as in various religious chant traditions. Cultures in which music making is a group activity is more likely to be antiphonal or polyphonic, as opposed to the monophonic music in cultures in which musical performance is a solitary activity. Cultures in which music is primarily vocal and sung by women will necessarily have musical textures with a higher tessitura than in those cultures where vocal music is sung by men.

The problem is that in the face of all this diversity one is thrown off the deep end when it comes to the aforementioned metaphysical issue of defining what music *is*. If we accept Blacking's definition that music is humanly organized sound, we have to also reckon with the fact that humans organize sound in radically different ways across cultures.⁴ Moreover, in some cultures music is inextricably connected with 'nonmusical' activities like dancing, and in others it is inextricably connected with certain ceremonies and rituals – so that in these societies music does not even exist as an independent entity that can be labeled as such (i.e. as "music") to begin with. This, in turn, makes it incredibly difficult to define what "music" is cross-culturally. Since he wants to recognize the cultural functions of music in defining what music is, and in studying what makes humans musical, this consequently becomes a problem for Blacking's definition of music – unless we want to accept all humanly organized sound practices across cultures as species-specific traits of humankind, including all the various political, religious, ceremonial etc. uses of music, which is clearly an untenable proposition.

To summarize, Blacking seems to subscribe to a functionalist approach to sociocultural systems in the study of music because of his insistence that a knowledge of the uses of music within society, which comes from a deep involvement within society, is imperative for studying even the formal aspects of music. But a functionalist approach also forces one to reckon with the immense diversity and complexity of music across the world, which stems from the manifold uses of music in different cultures

⁴ Even though *within* a culture there must be a consensus about which forms of organized sound constitute music for that culture even to have a musical tradition (which Blacking asserts that all cultures have).

globally. This leads to problems in defining how music is a cross-cultural, species-specific trait of all humans – and in consequence this is the larger problem with Blacking’s view of human musicality.

In this light, it is not surprising that many culturally-oriented thinkers have abandoned the search for a ubiquitous human musicality, some even denying that there is such a thing as “music”. For example, the philosopher Roger Scruton has argued that music is not a natural kind; it is not a unique, homogeneous, identifiable object that can serve as the basis for a broad study of human musicality (Scruton (1999): 156), which is an idea the semiologist Jean Molino subscribes to as well (Molino (2001): 168-169). In other words, music is but a cultural artifact, eminently variable across human societies, and not subject to the laws that natural objects are susceptible to, such as evolution.

As mentioned earlier, this belief has come to epitomize the broader ideological approach within the discipline of ethnomusicology about how to define music and how to answer the question of human musicality. Ethnomusicologist Bruno Nettl states this succinctly when he says that a “typically-ethnomusicological view would provide for a world of music that consists of a large group of discrete musics, somewhat analogous to languages, with stylistic, geographical, and social boundaries” as opposed to “a single vast body of sound and thought, a kind of universal language of humankind” (Nettl (2001): 464).⁵ Given the political aims of ethnomusicology in asserting the significance of various non-Western musical idioms against the hegemony of Western art music scholarship, it is understandable why scholars in the field would want to describe the intricate and complex attributes of ‘ethnic’ musics in all their cultural totality, and as seen from the perspective of an insider. To put it another way, ethnomusicology’s political goals made it imperative for scholars to demonstrate that they had a deep, preferably native, understanding of any non-Western musical idiom they were studying, including the manifold facets of the cultures in which these idioms exist. The consequence of all this was that the emic, anthropological approach became the methodological tool of choice in ethnomusicology.

⁵ Notice here the reference to – and explicit rejection of – Longfellow’s poetic statement about human musicality, with which I began this dissertation in the Prologue.

But such an approach, as we have just seen, forces one to contend with the immense diversity of music across the globe, given the manifold uses of music in world societies, and which leads to acute problems in the search for a universal human musicality, as a species-specific trait of *all* humans. The impossibility of defining what music is in the face of this diversity makes it impossible to *compare* different musical idioms – in order to isolate those properties that are truly cross-cultural (i.e. species-specific) from those that are particular to a culture (i.e. culture-specific). And without a uniform definition of what music is, there ceases to be a yardstick for comparing music cross-culturally – making the search for a universal, species-specific human musicality practically impossible. As Philip Bohlman says, ethnomusicology saw a preponderance of methodologically-oriented publications in ethnomusicology when the discipline arose in the 1950s, as opposed to texts that reveal newer insights about a wider variety of world musics (i.e. more data about world music cultures) that would allow for different idioms to be compared cross-culturally (Bohlman (1992): 124-125). This suggests that ethnomusicologists were implicitly aware of the impossibility of a true comparative musicology from their chosen anthropological perspective. That is, the diversity of world music cultures as seen from this perspective makes the comparison of these cultures impossible – hence ethnomusicology’s emphasis for the last few decades on defending different methods for doing ethnomusicology depending on the culture one is studying, as opposed to an emphasis on more data collection for purposes of comparison.

I began this chapter by talking about the relationship between music and language, and their possible connection as facets of human nature. Since ethnomusicology is a discipline in which this connection has been of particular relevance, it was necessary to spend time exploring how ethnomusicology has treated it, especially in the anthropological perspective that has dominated the field since its inception. (In fact, see the review by Feld and Fox (1994), especially pages 25-26 and 38, for a striking illustration of this phenomenon.) But our discussion has only focused on the musical side of this issue so far, so let us now see what the above discussion implies for a study of human linguistic ability – or more specifically, what it implies for the connection between human musicality and linguisticity.

A functionalist perspective, which I am claiming has been the general orientation of ethnomusicology, not only poses problems for a broad description of human musicality, it also poses problems for a description of the connection between music and language. The main reason for this should hopefully be obvious now – music and language serve different functions in society, so a functionalist perspective cannot possibly reveal any deep connection between the two. In particular, language functions as the facilitator of quotidian discourse, among other things, whereas music functions as society's soundtrack to a variety of phenomena, whether in art, in rituals and ceremonies, in propaganda etc. – but not in the facilitation of direct, day-to-day referential communication between individuals. (Although see Charles Boiles' (1967) study of the quasi-linguistic, conversational use of music in the Tepehua community in Mexico for a striking counterexample to this.)

For those ethnomusicologists who see language in sociocultural, functionalist terms, as does Bruno Nettl in his statement above, this functional difference between music and language serves, paradoxically, as a point of identity between them – i.e. music and language are the same because they are both cultural artifacts that differ drastically across human societies. This has encouraged many ethnomusicologists to pursue joint studies of music and language from an emic, anthropological perspective, for example Steven Feld in his work with the Kaluli of New Guinea (Feld (1974, 1982)).⁶

Of course, this position also denies the possibility of music and language being structurally or formally similar because it does not accept a formal attitude towards musicality or linguisticity to begin with. But some scholars more recently have begun to question ethnomusicology's quasi-ideological commitment to functionalism. For example, Kofi Agawu has criticized the field's overemphasis on cultural differences (Agawu (2003): 64) and on cultural contexts in the study of music in general (Agawu (2000-01): 65). The composer Francois-Bernard Mâche says that the denial of cultural similarities and the isolation of each culture from every other amounts to a form of 'reverse racism' (Mâche (2001): 474). The cognitive scientists Steven Brown, Björn Merker and Nils Wallin even suggest rehabilitating the

⁶ At one point in his text, even Blacking asserts the difference between music and language (Blacking (1973): 21), despite his earlier claims to the contrary – perhaps unsurprisingly, given his ultimately functionalist attitude towards this issue.

comparative musicology of the Berlin School for the unique, scientific insights it offered, while rejecting the anti-universalist rhetoric of ethnomusicology (Brown, Merker and Wallin (2001): 3, 21-22). For such scholars, many of whom maintain an interest in the structural aspects of music and language, the differences between music and language, and their variation across cultures, do become problems – which has eventually led many to question the possibility of any deep connection between music and language.

One of the most famous statements of this skepticism was expressed by Harold Powers when he said:

“[There is] a fundamental deficiency in the general analogy of musical structuring with the structuring of languages. Put barbarously in terms of the analogy itself, the “linguisticity” of languages is the same from language to language, but the “linguisticity” of musics is not the same from music to music. To [semiologist Nicholas] Ruwet’s telling observation [that “all human languages are apparently of the same order of complexity, but that is not the case for all musical systems”] I would add only that musical systems are much more varied than languages not only as to order of complexity but also as to kind of complexity. For instance, no two natural languages of speech could differ from one another in any fashion comparable to the way in which a complex monophony like Indian classical music or the music of the Gregorian antiphoner differs from a complex polyphony like that of the Javanese gamelan *klenengan* or of 16th-century European motets and masses. In monophonic musical languages we sing or play melodic lines more or less alone, just as we talk more or less one at a time in conversation, and our hearers follow accordingly. We do not all talk at once, saying different things, and expect coherently to be understood. Yet in ensemble musics with varied polyphonic lines we can (so to speak) easily make beautiful music together, which can be as easily followed by those who know how.” (Powers (1980b): 38)

Powers is of course right that there seem to be differences in the types of musics that are found across the globe – polyphonic musics do seem to have a different kind of complex structure than monophonic idioms do. But is it possible that these differences are really just apparent ones, which take on greater significance only because of the different functions they serve in different musical systems – so that a *non-functional* approach to dealing with musical structure might just reveal various similarities, rather than differences, between them? Consider the fact that Powers points to the relative “structuring” of music and language, but then proceeds to describe *how* we talk versus how we sing or play, which are instances of how music and language function in society, how they are *used* in society. As I suggested earlier, the different ways in which musical idioms are used might lead to their having apparently different structures too, e.g. in the group use of polyphonic music vs. the individual use of monophonic music. Additionally, some of the most influential language-influenced models of musical structure have

focused on music that is primarily harmonic or polyphonic, as we shall see in later parts of this dissertation, despite Powers' suggestion that *monophonic* music more closely resembles human language.

Given language's universal function as a facilitator of quotidian discourse, it might seem as if it needs to have some basic components that remain unvaried across languages – it has to have meaning, for us to convey in discourse (i.e. a semantics); a set of procedures to glue together the units of expression, such as words and phrases, into the sentences that frame discourse (i.e. a syntax); and an articulatory system for realizing human discourse as a sequence of speech sounds (i.e. a phonology). Lacking this particular kind of discursive function music does not need to have such a uniform constitution, at least on functional grounds, which allows it to be much more varied in its manifestation across cultures. And this might be why, as Powers observes, musical systems seem to have a greater variety in the kinds of complexity they have, relative to language. However, this functional difference between music and language is only a difference if we consider function to play a role in determining structure – which we cannot do anyway if we want to continue the search for a deeper connection between music and language, as I have been arguing so far. So it is possible that music would cease to be so markedly different from language if we take function out of the equation, and that it has a much more uniform structure across idioms than it might seem from its description in the ethnomusicological literature.

The big question then is whether there *is* such a non-functionalist, *formalist* description of human musicality or linguisticity, which might reveal a uniformity between them in the face of their apparent differences? Or has Powers sounded the death knell for any further search for the “universal language of humankind”?

1.1.1. The ‘Musilanguage’ Hypothesis

For the rest of this dissertation I will argue that there is indeed a formalist way of describing both human musicality and linguisticity. This is because I will propose some formalist thoughts about human musicality – but these thoughts will be based on formalist hypotheses about language that, as I suggested at the outset, have been proposed by the linguist Noam Chomsky and the field of generative linguistics he

founded. I am by no means the first to describe music in Chomskyan terms, but many of the previous attempts are susceptible to the above functionalist critique, with some, most notably the celebrated work of Lerdahl and Jackendoff (1983), implicitly accepting this critique and consequently abandoning the search for any deeper connection between music and language. However, I will try to illustrate how some recent (and some not so recent!) developments in music theory and generative linguistics provide exciting *new* insights into the music/language nexus, and demonstrate that the search for the universal ‘musilanguage’ of humankind is alive and well.

It might be helpful to start this discussion by exploring a Chomskyan response to Harold Powers’ above statement. Consider the following pair of sentences:

(1a) Jürgen read a book.

(1b) What did Jürgen read?

These sentences are obviously related, as a question and its answer. This relationship can be further described as the relationship between the object in 1a, the noun phrase “a book”,⁷ and the question word “what” in 1b, since “what” Jürgen read is “a book”. However, “a book” appears at the end of 1a, whereas “what” appears at the beginning of 1b – so these two sentences seem to be different at least in the order in which words appear in them. But notice that both 1a and 1b are concrete, tangible versions of these sentences, i.e. the *articulated* versions that are pronounced by a speaker and heard by a listener. Generative linguists argue that these sentences have an abstract structure too, separate from their concrete, articulated versions, of which we are not even conscious – where these sentences are actually the same. So the difference between them really lies in just how they are ultimately articulated. In other words, in articulating the final version of the sentence, one could leave it with the object in a position after the subject (as in 1a), or one could replace the object with the associated question word, which is then *moved* to the beginning of the sentence (as in 1b), a process called *wh-fronting*, since it involves moving a

⁷ Why this is a noun phrase, and why it might also be considered a determiner phrase is an important matter in generative linguistics, which I shall deal with in detail in section 1.2.3, where I give a brief history of the generative linguistics enterprise.

question word (which normally begins with “wh” in English, e.g. what, who, where, which, when etc.) to the front of the sentence. And it is in their final articulation that (1a) and (1b) come across as separate sentences.

Example 1.1-1. Wh-fronting in English

What (did) Jürgen read _____ ?



The final articulated forms of sentences are generally known as *surface* forms of a common, unarticulated, *deep* structure (Chomsky (1965): 16). So, the differences between sentences like 1a and 1b are essentially surface differences that disappear at deep structure, once phenomena like movement have been accounted for. To put it slightly differently, the final, articulated form of 1b can be considered a *transformation* of 1a, since it can be transformed into 1b through *movement* of the words of 1a to the positions they occupy in the surface form of 1b – in this case with the movement known as wh-fronting.⁸ (As an important caveat, I should point out that “transformation” here does *not* refer to the actual, real-time, processes through which a surface structure is generated. Instead, it refers to the abstract *relationship* that exists between deep and surface forms, which can be illustrated figuratively through “movement” – the latter not being an actual process of motion either.)

Now, why one chooses to generate one version of the sentence rather than the other could depend on the ultimate function of the sentence, e.g. whether it is being used to ask a question or answer it – but take this function away and the sentences are, structurally, the same. Therefore, by positing a transformational theory of linguistic structure, one can show how the apparent diversity in the forms of a language – or, importantly, *across* languages – are merely surface differences.

⁸ There are some interesting cases in which movement happens, but does not get articulated – which is known as *covert movement*. I will ignore these cases for the moment.

The notions of surface and deep structure have evolved within generative linguistics over the years, and the current Minimalist Program does away with them altogether. But the basic idea that language is a transformational system has persisted, and this is what has allowed generative linguists like Chomsky to argue for the universality of human linguisticity – because of their belief that the diversity seen in the world’s languages is not an intrinsic structural feature of those languages but merely a consequence of the transformational character of human language in general, and the various functions it can be put to.

I just made a terminological leap in the previous paragraph from a description of specific languages to a description of human language in general. This deserves an explanation because terms like “music” and “language” have inherent metaphysical problems associated with them, as we have explored to some extent now. What is language in general, if it is not the specific languages with their multifarious functions that populate the world’s cultures? (Even if we do accept that languages serve a more circumscribed set of functions due to their use in quotidian discourse.)

Given the functionalist provenance of such a question, the answer to it might have to be the formalist one that I speculated about at the end of the last section – and this is exactly what generative linguistics provides us. To understand this, consider that transformational phenomena such as wh-movement are purely formal. That is, these phenomena occur because words are ‘moved’ around (remember that movement refers to an abstract phenomenon and not an actual process) in the deep structure of a sentence according to certain principles that govern such movement, and all of this is unaffected by the functions the resulting surface sentence will serve. So, a transformation such as wh-movement occurs because a question word is fronted in a sentence, owing to a principle that governs how certain kinds of sentences are generated (*viz.* interrogative sentences)⁹ – but this is not affected by the fact that such sentences are often *used* to ask questions. This gives transformational phenomena an abstract, quasi-mathematical, and specifically algorithmic quality – because the principled manner in which

⁹ Or more precisely, given the caveat about movement being abstract, the principle states what kind of abstract relationship an interrogative sentence has with its declarative form.

sentences are generated, through the abstract manipulation of words, is akin to the principled manner in which a computer generates a specific output by ‘crunching’ an algorithm, which similarly involves the manipulation of symbol-tokens according to certain (logical) rules (although in this case symbol-manipulation is actual and not merely abstract).

In other words, the generation of a linguistic surface structure is akin to the information processing activities of a computer. But for many years now, especially since the dawn of the computer revolution, various scholars have claimed that the human mind is a computational system too, so that those human activities that possess formal characteristics, and presumably require algorithmic processes, can be situated in the workings of the human mind. In fact, the very concept of an algorithm was developed by the celebrated computer scientist and mathematician Alan Turing based on the principled, stepwise character that human thinking is supposed to have, especially in those thought processes implicated in solving certain computational problems, and his famous “Turing Machine” was an attempt to describe an artificially intelligent computing device that simulates human thought through algorithm-based information processing (Turing (1936, 1950)).

So, under this view, the formal processes involved in the generation of linguistic surfaces are really just the processes that take place in our minds when we are engaging in linguistic behavior. This is why Noam Chomsky says that language is specifically a *psychological* system in which information regarding the generation of linguistic surfaces is computed, which he calls the *computational system of human language* or C_{HL} (Chomsky (1986b): 5, 43-44; Chomsky (1995a): 7) – even though how the computational processes of C_{HL} actually take place in mind, in real time, is not necessarily well understood, the ‘hardware’ of the brain in which these processes are implemented even less so (hence the description of these phenomena through abstract concepts like “movement”).¹⁰

¹⁰ This is why the mind (or least the linguistic part of it), though a computational system, is *not actually* a computer, or does not necessarily resemble anything that we know to be a computer in the ordinary sense of that word. This can be explained in the context of the three levels of description that the cognitive scientist David Marr (Marr (2010): 24-27) proposed for describing cognitive processes, (a) the computational level (which describes the relationship between the inputs and outputs of the process, (b) the algorithmic level (which describes the actual steps through which the mental ‘software’ converts input to output, and (c) the implementational level (which describes

This brings us to what is essentially just another term for this computational system – but which is also the single most important term of this dissertation – viz. *grammar*. “Grammar” is the formal, psychological system that lies behind human linguistic ability, specifically the ability to compute linguistic information and generate linguistic surfaces. In light of this latter role, this system can also be called *generative grammar*. (Although, properly speaking, “generative grammar” is really just a theory proposed by generative linguists about the nature and structure of this system – the system itself is what it is.)¹¹ In this sense, “grammar” does not refer to the prescriptive systems that many of us associate with the term, which govern how language is *used* in certain contexts, as seen in the prescriptive ‘grammatical rule’ that forbids the splitting of an infinitive. Such *prescriptive grammars* really apply only to the culture-dependent, conventional uses of language, which often depend on the functions of language in a society – and which are thus distinct from the formal, psychological system in our minds that, *ex hypothesi*, gives human linguisticity its universal character.¹²

I said at the beginning of this chapter that I intend to pursue a specific psychological definition of music and language as a simple way of dealing with the complex metaphysical issues of what music and language are. In light of the Chomskyan approach to language, we are now in a position to understand this psychological perspective more clearly, and also to understand why grammar is such an important concept within this perspective. Moreover, we can now attempt an answer to the functionalist critique of

how this software is ‘installed’ in the hardware of the brain). Within this framework, all that generative linguists claim we know about language is (a), viz. the inputs (words) and outputs (sentences) of the computational system, and the abstract relationships they have with each other. The transformational description of language is not a description along the lines of (b) and (c), since it is not a description of the actual processes in the mind, and implemented in brain, through which a speaker articulates a sentence. In contrast, how actual computers process algorithms and how these are implemented in the hardware of a computer *can* be described well, presumably by the computer scientists and engineers who build them!

¹¹ Also, “generate” does not refer here to the processes (and hardware) through which surfaces are articulated, as noted before – rather it describes the (computational) relationships between abstract structures and principles that, in some poorly understood algorithmic way, yield the surface structures that make up the substance of the world’s languages.

¹² One could also describe generative grammar as a *feasible* theory of language, rather than a *usable* one (cf. Uriagereka (1998): 96). A feasible theory is one that allows linguists to solve some of the psychological puzzles about linguistic structure being discussed above, without concern for how these structures are used – the latter being the focus of a usable theory of language.

universalist descriptions of language – and by extension, music – that we have been exploring so far. So, functionalist approaches to language describe entities found all over the world, and which serve numerous sociocultural functions across the globe, which we can call “languages” (with a small “l”). However, what the formal, and in this case generative, approach is interested in is the human psychological faculty of language, or “Language”, which is characterized by C_{HL} . “Language” in this sense might manifest itself in the various surface forms we find in the world’s languages, but is separate from all of them in the important grammatical sense discussed in the previous paragraphs. If we think of the world’s languages in terms of this internal, psychological form, we can refer to them as *internal* languages, or I-languages, given their location inside our minds, as opposed to their culturally- and functionally-determined manifestations seen in the world’s cultures, which we can refer to as *external* or E-languages (Chomsky (1986b): 20-36) – the latter being what people often take, incorrectly, to be the only sense of “language”.

It is when one confuses I-language with E-language that at least part of the metaphysical problem of defining “language” arises. This is because unlike the languages that exemplify E-language, I-languages *can* be described as a uniform, universal phenomenon, viz. as products of a grammatical system that generates surface structures, possessed by anyone who is able to speak or understand a language (meaning the vast majority of human beings), and *this* could be a species-specific trait of humankind, which all humans genetically inherit – not the functional E-languages that populate the world’s cultures, which (as emic-oriented anthropologists rightly point out) do not necessarily have any universal properties. In fact, as Chomsky argues, I-languages *must* be a genetically-inherited, species-specific trait of humans because essentially every human being is capable of gaining native fluency in some language or the other, despite the great diversity in E-languages, and this fluency is usually acquired at a very young age, even though most children are not given a thorough instruction in their native languages, especially in underprivileged communities (which linguists call “poverty of the stimulus” (Chomsky (1980a): 34, Legate and Yang (2002))). All of which implies that humans must have an innate, (i.e. genetically-inherited) capacity for language.

The above points about generative grammar and linguistic universality are important and controversial ones, which therefore merit the varied and detailed treatment I will give them throughout this dissertation. But let us not delay a consideration of how all of this applies to music any longer. That is, what does the above generative grammatical view of language have to do with music, and how does it help address the functionalist critique of (and especially Harold Powers' argument against) the music/language nexus?

Well, first of all, the functionalist approach to human musicality, as seen in the canons of ethnomusicology, focuses on the surface forms of music as they manifest themselves in specific cultures – which is really a study of “E-music”. But given the (functional) diversity and non-universality of E-music, this leads to various philosophical and practical problems in defining what “music” is. Therefore, the place to develop a universal theory of human musicality should properly lie in a study of “I-music”, i.e. in a study of the musical mind – and in the psychological aspects of musical information processing, as seen in the workings of an abstract computational system of human music or “ C_{HM} ”. Secondly, one could argue that the *structural* differences between, say, polyphonic and monophonic forms in music, which Harold Powers takes to be evidence for the non-universality and non-linguisticity of music, are just surface differences in forms that are generated by this C_{HM} – i.e. by a universal generative grammar of music that all humans, *ex hypothesi*, have hardwired in their minds.

This is the main premise I will try to defend from several angles throughout the course of this dissertation. Moreover, the description of generative musical grammar I will give will bear a striking resemblance to generative linguistic grammar, especially in the way the current Minimalist Program in generative linguistics describes it (see Chomsky (1995b, 2002) and Lasnik (2002) for overviews of this project). Given the importance of grammar in a universalist description of music/language, these similarities between musical and linguistic grammar pave the way for a description of an even more remarkable similarity, if not identity, between music and language in general – which is the underlying theme of this dissertation anyway, albeit an age-old one as I discussed at the beginning of this chapter. Therefore, my arguments will essentially amount to a defense of this theme, what I refer to in the title of this section as the “musilanguage hypothesis”.

Notice that this argument will also be a purely formal one since it will be an exploration of, for example, generative musical grammar in Indian and Western music (as I will explore in detail in chapter 1.3), and not of the functions of monophonic music in India and polyphonic music in the West (which are different for the precise cultural reasons mentioned earlier). On this note, it might also be useful to appreciate that the study of musical/linguistic generative grammar does not necessarily involve a study of whether certain deep structures are found across the world's languages/musics; rather, it involves, as we have seen before, a study of the quasi-mathematical, formal relationships (such as those seen in transformational phenomena), which *connect* deep and surface structures across idioms (and as has been demonstrated in a wide variety of languages now by generative linguists over the past several decades).¹³

However, there are some surface, articulated musical/linguistic structures that do seem to exist across idioms too, and often for good reasons. For example, constraints on the kinds of sounds we are able to produce with our vocal apparatuses limits the kinds of speech sounds one finds in the world's languages, so that a small set of phonemes predominate in them. Such *substantive universals* exist in music too. Francois-Bernard Mâche alludes to the frequently cited example of pentatonic scales in global musical traditions, particularly “pentatonic polyphony on a drone”, which he observes as occurring in the Nung An music of Vietnam, in the *Gerewol* song tradition in Niger, in the music of the Paiwan aborigines of Taiwan, in *Dondi* funeral music from Indonesia, in Sena choir songs from India, and in Albanian folk songs (Mâche (2001): 475). Also, Jeff Pressing has discussed common scale and rhythmic patterns in musical traditions from West Africa and Eastern Europe (Pressing (1983)).¹⁴ But since the study of a universal human ability for language or music, at least in the Chomskyan sense, is a study of grammar, its

¹³ If the study of generative grammar is the study of formal transformations in language, then a study of *musical* generative grammar should essentially be a study of transformational phenomena in music – especially if this study is being conducted as part of the musilanguage hypothesis. But some, such as the linguist Ray Jackendoff, have argued that music does not have a transformational grammar like language (Jackendoff (2009): 201) – which is in many ways the most serious threat to the musilanguage hypothesis. I return to this important issue in section 1.2.4 of the next chapter, but its importance makes it worth mentioning here, even in passing.

¹⁴ Pressing confusingly calls the shared patterns in West African and Eastern European music “cognitive isomorphisms”. Even though he gives evidence from cognitive psychological studies to explain why the scale and rhythm types he observes are so frequently articulated in their respective cultures, his focus is on these patterns themselves rather than on the grammatical (and therefore psychological) operations that generate them – which makes the “cognitive isomorphisms” label an inaccurate one.

focus tends to be more on *formal* universals, which are the abstract categories that participate in grammar, such as nouns, verbs, and (as we have seen in the case of wh-movement) direct/indirect objects and wh-phrases – and which manifest themselves through, but are separate from, the thousands of concrete, articulated words that make up the world’s languages.

Along these lines, a generative-linguistic approach to human *musicality* would focus on the abstract categories, rather than the specific melodic or rhythmic types (such as certain scale or metrical patterns), that participate in musical grammatical processes. An example of such a category, which I will discuss more fully in the next chapter, is that of *harmonic function*. A harmonic function is the abstract label that is applied to certain musical structures (such as chords) based on their grammatical role and position in a musical ‘sentence’ – so that harmonic functions like “tonic” and “dominant” are analogous to (but not the same as) abstract linguistic categories like “subject” and “predicate”. And like these linguistic categories, harmonic functions can be articulated by a variety of surface sounds, even across idioms as I will argue.

Despite his criticism of hypotheses about musical universals, even Harold Powers seems to accept the existence of such abstract categories in music. For example, in Western Classical tonal music – the idiom with which “harmonic functions” are most closely associated – these categories determine which pitches can occur in a surface musical structure. So for a group of pitches to be accorded tonic function in, say, C major, the pitches C, E and G all have to appear either directly in the surface structure or be strongly implied in it. This means that the surface must be polyphonic too or else these pitches cannot be sounded simultaneously – in other words, a harmonic function constrains how a polyphonic musical surface is articulated. Powers argues that such constraints exist in Javanese gamelan music, and European Renaissance and medieval polyphony too, although he suggests that these constraints are closer to the surface in these idioms than in Western tonal harmonic music (Powers (1980b): 39-42). He also speculates that such constraints are not to be found in various other polyphonic musics (albeit without providing any examples to support this), and concludes that these cross-idiomatic connections suggest the non-linguisticity of these idioms, since he believes that the polyphonic nature of these idioms makes them

unlike language. But if anything, the fact that these idioms are structurally constrained, in the way tonal harmonic music is constrained by abstract categories like harmonic function, suggests the exact *opposite* of Powers' claim – because much of the Chomskyan argument for a universal human *linguistic* ability is based precisely on the presence of such categories in language.

To summarize what we have discussed so far, the formal, as opposed to functional, approach to studying human musicality or linguisticity I am pursuing in this dissertation is primarily a study of an abstract, psychological grammar, and the grammatical principles and categories from which musical or linguistic surface structures are generated across idioms. It is not a study of the specific manifestations of these phenomena in particular idioms, and it is not a study of the communicative or artistic functions of either music or language that result from these manifestations in different cultures. This means that it is also a study that focuses on the *syntactic* aspects of language and music, and not on their semantic or (as we shall soon see) phonological aspects – although these aspects play a role in grammatical theory too, as I will discuss in the next section. This distinguishes the formal approach from a functionalist one, which does focus on the functional, semantic aspects of music and language, and which therefore endangers their connection to each other in the process too, as we saw in the last section.

This brings us full circle, since we can now return to an examination of John Blacking's description of music and language as species-specific traits of humans. That is, we can now say that music and language are species-specific traits of humans, and that how humans are musical/linguistic creatures is related to human nature – but not in the sociocultural, functional sense pursued by Blacking. Rather, music and language are, *ex hypothesi*, species-specific traits of humans because they are both based on the hardwired, computational aspects of the human mind shared by all humans across cultures – and this psychological basis for music and language is what makes them aspects of human nature too.

Restating this in terms of Blacking's original framework, we could say that all humans do have unconsciously-acquired knowledge of certain 'conceptual models' from which actual musical/linguistic surfaces are generated, and this is what constitutes the native intuitions of speakers – but contra Blacking,

the argument here is that this knowledge is really a knowledge of grammar. Moreover, the “deep involvement in society” required for this knowledge is not one that involves the functions of music/language in specific societies – if anything, it is part of the genetic inheritance (to use another of Blacking’s terms) for music/language that *all* humans have, if the arguments of the generative tradition hold strong.

In this light, the study of human musicality/linguisticity is a study of what kinds of surface structures are *possible* if the grammatical workings of the musical/linguistic mind are allowed to operate freely, unhindered by the sociocultural factors that ultimately determine their use. Such a study is better known as the study of human linguistic/musical *competence*, as opposed to the study of *performance* (Chomsky (1965): 3-10, Chomsky (2006): 102), the latter of which *can* describe music and language in terms of factors relating to their sociocultural use. In other words, a study of human musical/linguistic competence is the study of the knowledge that all native speakers of musical/linguistic idioms have of the innate grammar that generates musical/linguistic surface structures, which they unconsciously acquire as part of their genetic inheritance – as a species-specific trait of humankind.

So, the present study is one of human musical/linguistic competence, and by extension a study of our knowledge of musical/linguistic grammar – it is therefore essentially an *epistemology* of music/language. According to Chomsky such an epistemology must fulfill at least two criteria of adequacy. First, it should be able to describe the knowledge that native speakers of a language have about which sentences are grammatical in that language, i.e. it should be able to describe all, and only, the grammatical sentences of a language, in a way that reflects (as John Blacking sensed) the unconsciously-acquired intuitions of native speakers. Chomsky calls this the requirement that a grammatical theory have *descriptive adequacy* (Chomsky (2006): 24). A descriptively adequate grammatical theory would of course account for those aspects of Language or Music (i.e. I-language or I-music) that manifest themselves in diverse ways across idioms, such as the different orders in which words appear in grammatical sentences in different languages. In other words, a descriptively adequate theory of universal grammar must account for the particular grammars through which it manifests itself. But since this entire

enterprise is a formal one, the theory needs to account for only the formal aspects of particular languages/musics, i.e. how the surfaces of particular languages/musics are generated from a set of abstract categories and principles – which I had described earlier as the formal approach to sociocultural systems.

But even a descriptively adequate theory of particular grammars would not amount to a true epistemology of human music/language if it does not also explain why humans even have grammar to begin with. That is, if this psychological system is the locus for human musicality/linguisticity, an epistemology of human language has to explain *why* we have such a psychology; i.e. what is the structure of the mind such that it allows us to have this kind of psychology? Chomsky calls this the requirement that a theory of universal generative grammar have *explanatory adequacy*.

In addition to the two above criteria for adequacy in grammatical theory, the current Minimalist Program in generative linguistics suggests that a grammatical theory should also fulfill certain criteria that go beyond explanatory adequacy. This is because there seem to be aspects of human nature that have to do less with its specific psychological makeup and more with some general properties of natural systems. As Chomsky says, theorists should “seek a level of explanation deeper than explanatory adequacy, asking not *what* the properties of language are, but *why* they are that way” (Chomsky (2004): 105). If music and language are connected to each other as parts of human nature – an assumption of the musilanguage hypothesis that forms the cornerstones of this dissertation – then an adequate epistemology of music/language has to account for these aspects of human nature too, insofar as they relate to music and language. Therefore, a Minimalist approach can provide important philosophical foundations even for the kind of study of generative *musical* grammar, and of musilanguage, that this dissertation wishes to pursue – and hence the title of this dissertation.

Given how radical many of the above ideas are, a more detailed exploration of the Minimalist approach to language, and by extension music, is in order – which is why I will deal with it separately and in detail in the next section. But before we proceed to this important issue, it is worth stating again, as I did in the Prologue, that this dissertation does not aim to give a comprehensive description of human musical

competence, let alone human musilinguistic competence, which also fulfills the three criteria of adequacy stated above – even though it does aim to *justify* this enterprise from various (viz. historical, philosophical and technical) perspectives. A comprehensive description of musical competence would have to account for more than the Indian and Western idioms I will be discussing in this dissertation, and would have to give a more detailed account of musical grammar than is possible in a single document as this – and is therefore beyond the scope of this, or for that matter any, project at this point. Given the novelty of this enterprise, much more work needs to be done before a comprehensive description of human musilinguistic competence becomes tenable. But if this goal is a legitimate and justifiable one, as this dissertation contends, then the above considerations can provide good methodological guidance on the way there. This is why I will return to them time and again during the course of the dissertation.

It might also be worth our while to explore another methodological issue relating to human musical/linguistic competence before I proceed to a discussion of the Minimalist approach to language and music, primarily because I have been skirting it since the outset of this chapter. So, I will devote the rest of this section to this end.

This is the issue of the role of biology in the study of human musicality/linguisticity. After all, John Blacking did not just talk about the importance of sociocultural systems in the study of human musicality/linguisticity; he also stressed the importance of biological systems in explaining this phenomenon. This point is particularly relevant here given the above claim about human linguistic/musical competence being genetically inherited and species-specific – both of which Blacking invoked in his statement at the beginning of this chapter. Given the problems inherent in sociocultural approaches to this issue, could it be that a study of human musical/linguistic competence should really be a study of human biology?

It is possible that Blacking would have accepted this outcome himself. Of course he always foregrounded sociocultural systems in the study of human musicality, which is a commitment that never waned, since in later years he even attempted to develop a biological description of human musicality that

takes sociocultural factors into consideration (e.g. in Blacking (1992)). But as we have seen, the invocation of cultural factors only yields a description of musical performance and not a tenable cross-cultural description of musical competence, and Blacking accepted this position himself in yet another paper (Blacking (1990)). However, maybe he did realize implicitly that one might be able to describe the truly cross-cultural, species-specific nature of human musicality by foregrounding the biological over the cultural. At the very end of *How Musical is Man*, after arguing for the importance of culture in the study of music throughout his text, Blacking does an apparent about-face – and makes this cryptically-worded statement in defense of a *biological* approach to human musicality:

“Suppose we look at the social, musical, economic, legal, and other subsystems of a culture as transformations of basic structures that are in the body, innate in man, part of his biological equipment; then we may have *different explanations for a lot of things that we have taken for granted* [my emphasis], and we may be able to see correspondences between apparently disparate elements in social life.” (Blacking (1973): 112)

The possible biological basis for human musical and linguistic ability – particularly a joint biological basis for both abilities, in the face of the cultural diversity of musical and linguistic traditions – has intrigued several thinkers in the past. No less a personage than Charles Darwin surmised that the reason why all humans seem to have music is because we genetically inherited it, owing to “our semi-human progenitors having some rude form of music, or simply to their having acquired for some distinct purposes the proper vocal organs” (Darwin (1871b): 335). Moreover, the vocal basis for both music and language led Darwin to speculate that music and language have the same origin, i.e. from a shared system that some modern commentators have called, in a similar vein to that of this dissertation, “musilanguage” (Brown (2001a)) or “musical protolanguage” (Fitch (2006)). In fact, his convictions about their shared vocal origin also led Darwin to devote much of his discussion about human music/language evolution to a comparison with birdsong (Darwin (1871a): 53-62), and a later section on the evolution of birdsong is even titled “Vocal Music” (Darwin (1871b): 51-61). This is because birds are one of the few non-human species among whom ‘vocal music’ predominates, and part of whose vocalizations are acquired like human music and language – such vocalizations also being subject to a critical period of acquisition in

youth just like human language, and has been recently argued for human music (Trainor (2005)). Given the common (but contentious)¹⁵ belief that birds sing for reasons that can be explained in adaptive terms (e.g. to attract mates and defend territory, both of which might lead to reproductive success and thus survival of the species), Darwin took the similarity between human music/language and birdsong to be evidence for the evolutionary basis for both human music and language as well. Therefore, it is possible that music and language have a joint origin in our biology and evolutionary history, so that the search for human musilanguage should be a “biomusicological” or “biolinguistic” research program.

But one thing worth noticing in Darwin’s characterization of the origin of music/language is that music seems to originate before language in it, so that language originates *from* music rather than their both having a common origin in some proto-musilanguage – an idea that has had actually some popular appeal over the ages.^{16,17} Language is often believed to have originated after music because it is supposed to have added on that which music does not apparently have, viz. propositional or referential meaning. For example, even Wilhelm von Humboldt, whose ideas have strongly influenced modern generative linguistics, believed that “man, as a species, is a singing creature, though the notes, in his case, are also coupled with thought” (Humboldt (1999): 60). But as I have argued before, only a functionalist approach to explaining human linguistic ability requires a consideration of the referential aspects of language, not the formal description of human linguistic competence that I am pursuing here. This point is particularly

¹⁵ Since singing can attract not just mates but also rivals, it can possibly lead to a *decrease* in reproductive success. Also, that *female* birds have been found to sing frequently too (Morton (1996), Langmore (1998, 2000)) – which Darwin thought to be an aberration – is a problem for those who believe that male birds ‘serenading’ female birds is what leads to the adaptive advantage of singing. This is compounded by the fact that female birds appear to have the neural apparatus for singing even when they do not sing, which can be artificially triggered with hormones (Fitch (2006): 186), which matches the fact that *human* females are as musical as human males – all of which problematize the role of singing in sexual selection.

¹⁶ For example, it forms the basis for the ‘creation myth’ depicted at the beginning of *Das Rheingold* (Nattiez (1993): 53-60, Levin (1999): 42-43, Albright (1999): 51, Borchmeyer (2003): 218), the first opera in Richard Wagner’s monumental four-opera “Ring of the Nibelung” cycle. The opera opens with a sustained passage in E-flat major that lasts for over 130 measures – which depicts, in succession, the beginnings of music and language in the mystical waters of the Rhine; first harmony, then melody, then the musical protospeech of the Rhinemaidens, and finally the first complete lyrics of the opera. This evolutionary sequence, from musical protospeech to language, is exactly how Darwin envisioned the actual evolution of human music and language.

¹⁷ As further evidence of this consider how some advocates of the musilanguage hypothesis also say that “there is no *a priori* way of excluding the possibility, for example, that our distant forbears might have been singing hominids before they became talking humans” (Brown, Merker and Wallin (2001): 7).

relevant since Noam Chomsky, despite being influenced by Humboldt's ideas as just mentioned, has argued that reference does not necessarily exist in human natural languages (Chomsky (2000): 148-153).

Moreover, even if music did arise before language, there are significant differences between human music and the "vocal music" of other animals, including birdsong – and not for the trivial reason that birds don't sing operas and chimpanzees don't write symphonies. For example, the formal phenomena that underlie how the varied surface structures of music are generated are actually quite different from those found in the vocal communication systems of non-human animals – and have much more in common with human language. I will have reason to describe some of these in more detail in a bit, but just as an illustration consider this: one of the more important phenomena that governs how musical phrases are generated, and how we perceive musical structure, are the relationships that musical pitches and pitch collections have with each other. In fact, functional-harmonic relationships between two chords like that of "tonic" and "dominant" depend on the relationship between the chords, which is often thought of in terms of their distance from each other in a certain pitch space, e.g. the circle of fifths. This is similar to the relationship between different words, which helps determine their grammatical function in language.

Now, of the various relationships that musical pitches have with each other, the one between a pitch and the pitch one or more octaves above or below it is particularly important, since octaves are perceived as equivalent in almost all musical cultures (Brown, Merker and Wallin (2001): 14). But other relationships between pitches are of great structural significance too, such as the one between two pitches a fifth apart on the major scale – which comprises part of the aforementioned relationship of tonic to dominant. However, I am not aware of any experimental evidence that suggests that any non-human species can comprehend such a variety of pitch relationships, even ones with whom we share a good deal of evolutionary history. (Such as rhesus monkeys, although this species is notable for its ability to comprehend at least octave relationships (Wright et al. (2000)).) So, the vocal communication systems of non-human species are not necessarily based on the pitch relationships of the kind found in human music. In other words, human music is different from the "vocal music" of other animals.

In light of the above, there is no more reason to believe that language arose out of a ‘musical’ protolanguage than that music arose out of a ‘linguistic’ protomusic, at least from a formalist perspective. Rather, a more justifiable hypothesis seems to be that music and language both arose from a common origin, human proto-musilanguage, which is a system that might have been partially shared between our hominid ancestors and various non-human species – but which probably had some marked differences from the vocal communication systems of other species too.

Despite his belief that language originates from music rather than vice-versa, Darwin actually had some thoughts about such a proto-musilanguage, including the fact that it resulted from a general increase in the mental faculties of our hominid ancestors – which also allowed it to develop special properties that separated it from non-human vocal communication (Darwin (1871a): 54-57). This is an important point because human linguistic/musical performance is constrained not only by sociocultural factors, but also by biological factors that limit what sounds we are capable of producing or the amount of information we are capable of holding in memory when producing sentences. (Although musical/linguistic *competence* is not constrained by anything other than the principles that govern how grammatical structures are generated.) Moreover, it has been argued that a change in those parts of the nervous system that control air flow to the lungs also helped the joint evolution of music and language (MacLarnon & Hewitt (1999), Fitch (2006): 196), since an increased control over breathing greatly facilitates both the ability to sing music and speak language. Finally, evidence for the joint evolution of music and language comes from prehistoric musical instruments. Instrumental music seems to have been part of even the oldest musical societies (e.g. see Zhang, Harbottle, Wang and Kong (1999)), but some have argued that it existed even in the societies of our Neanderthal ancestors, based on a prehistoric ‘flute’ found in a Neanderthal burial site, which was made from the thigh bone of a cave bear (Kunej and Turk (2001)). This puts the origin of instrumental music back to a time that is at least as early as the

speculated origins of spoken language, maybe even at a time before the origin of modern *Homo sapiens* (given the flute's Neanderthal provenance).¹⁸

The above evidence notwithstanding, the joint evolution of music and language is hard to demonstrate in general since the products of music and language, and the organs that produce them, do not fossilize – so, it is hard to find past evidence of how music and language evolved. Even authors who have provided hypotheses about their origin, such as David Huron, readily accept the speculative, “just so stories”, nature of the enterprise (Huron (2003): 57-59).

In consequence, some authors have speculated on the origins of music and language by taking the functional uses of music and language into account. For example, David Huron himself has speculated that music in particular might have given early hominids a survival advantage in eight different ways: (a) by facilitating mate selection, (b) by increasing social cohesion and increasing the effectiveness of group activities like hunting, (c) by increasing the effectiveness of specific acts, like pulling a heavy object, (d) by facilitating perceptual development through constant ‘ear-training’, (e) by facilitating motor development, including in the coordinated vocal organs needed for speech, (f) by providing conflict resolution, e.g. in ‘campfire songs’, (g) by providing a harmless way of passing time, and thus keeping an animal out of danger, and (h) by facilitating intra-group communication, especially across generations or over long periods of time (a la the role of music in helping humans develop a history, mentioned at the beginning of this chapter). Huron gives particular importance to (b) here, i.e. the role of music in facilitating social bonding, for which he provides several pieces of evidence, e.g. the correlation between sociability and musicality in children with William’s syndrome (and the correlation between their mutual absence in children with Asperger’s-type autism); the role of music in group activities like singing “Happy Birthday”, popular across the world by now, or in activities like warring (or cooking, praying, story-telling etc., cf. Agawu (1995): 8-23) in various ‘indigenous’ societies; and finally, the role of music

¹⁸ Although whether this ‘bone flute’ is a true musical instrument is controversial because of the damaged state in which it was found (Fitch (2006): 197).

in mood regulation and in facilitating neurochemically- (such as oxytocin) based social behaviors, such as courtship and sex.

The role of social function in the development of music reappears in David Temperley's discussion of the 'evolution' of musical styles, which is of course not an example of true biological evolution but rather of cultural change (Temperley (2004b)). Building on some of Huron's work, Temperley argues that musical surfaces function to communicate musicians' thoughts to listeners, so that music (and specifically musical styles) must evolve to fulfill this social function – this “communicative pressure”. For example, some of the stylistic features associated with polyphonic composition in the Western art music tradition can be seen as specifically facilitating communication with the listener; the ban on parallel fifths and octaves exists in this idiom because such sonorities allow the melodies of a polyphonic piece to fuse into one, which prevents it from being heard as polyphonic – i.e. they prevent the piece's polyphonic structure from being communicated to the listener.¹⁹ Temperley extends these ideas to various other musical idioms too; for example, he talks about styles which have evolved to have very strict meters (such as West African and Western Rock music), in order for them to simultaneously have very complicated, syncopated rhythms, which stand out against these meters and are thus communicated, as stylistic traits, to listeners. He contrasts these styles with Romantic-era Western art music, which is less syncopated, since syncopations would not stand out against the relatively relaxed meter of much music in this era, which Temperley calls *rubato* (cf. his figures 1B and 1C).²⁰

¹⁹ Temperley suggests that this also explains the relative absence of (and the pedagogical ban on) small intervals between melodies in lower registers in a polyphonic piece – which tend to fuse into one due to interference between the overtones of pitches in the two melodies that fall within a certain critical band of frequencies.

²⁰ Temperley's use of the term *rubato* for a relaxed meter is slightly misleading here. In much Romantic music, the underlying meter is actually quite strict, so that uses of *rubato* really amount to another form of syncopation. For example, in the piano music of Chopin (which Temperley discusses, and which is often cited for its particularly explicit use of *rubato*), the pianist's left hand often plays regular sequences of chords (which Temperley notes) and this leads to such music having a fairly strict meter, against which the metrically-relaxed *melody* of the right hand stands out – essentially as a form of syncopation against the meter of the left hand. In fact, Karol Mikuli, one of Chopin's foremost students, writes in the Foreword to the Dover edition of Chopin's Mazurkas that the above is exactly how Chopin intended his music to be performed himself, and that a “metronome never left his piano”. So, the relative absence of syncopation in Romantic-era music might owe to the listener's difficulty in picking out such syncopations from the *other* syncopation of (right-hand) *rubato*, which was commonly used in this music.

As we have seen earlier, speculations about the connections between social function and the origins of music have been widespread in ethnomusicology, even before Blacking expressed his thoughts on the subject (e.g. in Merriam (1964)). This theme has also been a point of focus in speculations about the origins of language (e.g. Pinker (1994)). So in this light, both Huron's and Temperley's arguments about the origins of music are intriguing.

However, these hypotheses all invoke the functional use of music in human societies too – which as we have seen before, leads to the metaphysical question of what music *is* rearing its head again. When Huron talks about the role of music in social bonding, what music and whose music is he talking about? His arguments would certainly seem relevant for societies in which music making is a group activity, but what about societies in which it is not? In other words, invoking sociocultural or functional factors in even a biological study of music or language – giving the sparse evidence we have for proposing hypotheses about their origin – leads us back to the problem of whether this can tell us anything substantive about human musical or linguistic competence.

Moreover, since it is so hard to find a common, cross-cultural definition of music or language, evolutionary theorists are often forced to base their hypotheses about musical/linguistic evolution on the least common denominator of what music or language might be cross-culturally – unless, of course, one considers these systems in formal terms, as we discussed above. In other words, barring a formal approach to music/language, those surface aspects of these systems that seem to be the most prevalent across societies (and which thus comprise the least controversial definition of “music” or “language”), are often all that function-oriented evolutionary theorists are left with to base their hypotheses about music/language's origins on – and this runs the risk of seriously *under*-complicating what music and language really are. For example, the most prevalent surface feature of music across societies is that it is based on sound, hence Blacking's definition of music as “humanly organized sound” and Darwin's allusions to non-human “vocal music”. But as I have suggested before, the sound aspects of music and language are part of surface structure, since it is through sound that surfaces are articulated, even though

the deep-structural aspects of music/language are possibly quite distinct from their ultimate sonic articulation.

Now, if the sound aspects of music/language are properties of their surface structure, they do not shed much light on their deeper grammatical characteristics, which I have been arguing make up the essence of human musical or linguistic competence. And many of the extant evolutionary hypotheses about music/language's origins seem to suffer from this shortcoming, because in basing their ideas on sound they are often forced to reduce the complex structure of human music/language to its most trivial sonic form, e.g. the grunts and barks that allegedly constituted the speech or song of our non-human ancestors. After all, this is to a large extent all that is shared between human music/language and the hoots of chimpanzees, tweets of songbirds and whistles of dolphins (at least when considered in purely sonic terms). But since evolutionary theorists have so little to base speculations about music/language evolution on, such a sharing ends up being of much greater significance to them than it might be ordinarily. For example, in discussing the role of social bonding in music evolution, David Huron describes the role of music in stimulating group activity as similar to the cackling of geese about to take flight in a coordinated way (Huron (2003): 54). In his version of the musilanguage hypothesis, Steven Brown argues that human music and language both evolved from a common state that was based on discrete, meaningful word-like sounds called "lexical tones", which are frequently found in human languages, especially tone languages (Brown (2001a): 279-285). The discrete, pitch-based structure of lexical tones distinguishes them from the vocal communication systems of non-human species such as the unpitched hoots of chimpanzees or the non-discrete vocal glides of gibbons; but they are similar enough to the alarm calls of East African vervet monkeys for Brown to consider them the 'missing link' between human music/language and non-human vocal communication. Even Darwin's own hypotheses about human musilanguage suffer from an overemphasis on the sound aspects of music and language, and a neglect of the complex formal phenomena that underlie musical/linguistic competence, since for him a prominent example of human linguistic ability is the "murmur of a mother to her beloved child... more expressive than any words" (Darwin (1871a): 54), which he finds to be strikingly similar to "the inarticulate cries of the lower

animals”, the only difference being that humans can also articulate their thoughts and connect “definite ideas with definite sounds”.²¹

Though beautiful in its poetry, Darwin’s view has at least one significant problem. If being able to articulate definite ideas through definite sounds is a critical aspect of human language, any evolutionary theorist has to explain how this evolved from the inarticulate cries of non-human animals. As Tecumseh Fitch says, “this leap, from non-propositional song to propositionally-meaningful speech, remains the greatest explanatory challenge for all musical protolanguage theories”.²² This brings us back again to the idea that human musical/linguistic competence has to do primarily with the formal, grammatical processes that underlie our psychological ability to construct and comprehend surface musical/linguistic structures, and not with the communicative functions of these surface structures – to which we might now add that it does not have much to do with the sonic realization (i.e. the ‘phonology’) of these structures either. This is not to say that what we express through music/language, or how music or speech sounds, is irrelevant for a study of musical/linguistic competence; it is just that such a study cannot be based primarily on these aspects of music/language without either defining these two terms in contentious ways, on the one hand, or trivializing them on the other. Hypotheses about the origin of music and language must deal with the complex formal structure of music and language if they do not want to end up giving evolutionary explanations for phenomena that are basically caricatures of music and language.

The importance of a formalist approach to studying the origins of music and language becomes even more relevant in the light of the problems inherent in the *gradualism* that functionalist approaches often ascribe

²¹ Although infant-directed speech, or “Motherese”, is often considered to be quite musical, which adds to the list of similarities between music and language. Specifically, infant-directed speech exaggerates certain aspects of adult speech, e.g. it uses a slower tempo, larger pitch contrasts between syllables etc., and generally simplifies the speech input. Some have argued that this is essential for language acquisition (e.g. Elman (1993), Plunkett (1997): 150). But many of these features are exactly what has made some scholars compare Motherese to music, especially child-directed music (e.g. lullabies) (cf. (Trehub (2001): 437-439). Also, infant-directed music has a universal provenance too, and people seem to be able to recognize lullabies even when they are from foreign cultures (Trehub, Unyk and Trainor (1993)), which suggests that infant-directed music and language might have common foundations.

²² W. Tecumseh Fitch, “Musical protolanguage: Darwin’s theory of language evolution revisited”. Published online February 12, 2009 at <http://languagelog ldc.upenn.edu/nll/?p=1136>. Accessed March 13, 2013.

to the evolution of music and language too. This is the idea that the complexity of music and language both arose gradually from a simpler proto-musilinguistic state. One of the problems with such a gradualist perspective is that it cannot account for what is probably the most characteristic, formal, feature of music and language, viz. the hierarchical and recursive structure of their grammars.

To understand this, consider sentence (1a) again:

(1a) Jürgen read a book.

This sentence (or “clause” to be specific) can stand alone, or can be embedded in a larger clause like (1c):

(1c) Kwame said Jürgen read a book. (Often *embedded clauses* such as this are preceded by a *complementizer*, such as the word “that”, which would give us “Kwame said that Jürgen read a book”.)

Now, all of these clauses have a hierarchical structure, since they have constituents (i.e. words, or groups of words) embedded inside other constituents. For example, in (1a) the noun (or determiner) phrase “Jürgen” and the verb phrase “read a book” are embedded inside the larger clause “Jürgen read a book”. Moreover, this larger clause is embedded inside the even larger clause that is (1c).

Such hierarchical structure in sentences is not something arbitrary that has been made up by linguists. It is a necessary prerequisite for talking about the various kinds of transformation that take place in grammar. For instance, for the wh-movement transformation to occur, the noun/determiner phrase “a book”, which is embedded hierarchically within the larger verb phrase “read a book”, has to move out of this larger phrase to the front of the sentence, as we saw in the case of (1b) earlier. But this has important consequences for how far the moved phrase can be (in the front of the sentence) from the verb phrase that it came from, and also for the ‘gap’ that it leaves behind in the verb phrase (depicted with the horizontal line in Example 1.1-1) – all of which have to be dealt with for a sentence to be grammatical, and all of which would not happen if “a book” were not a hierarchically-inferior constituent within “read a book”.

But not only is the grammatical structure of language hierarchical, these hierarchies are often recursive too – i.e. they involve embedding constituents within *themselves*. In the case of (1a) a

noun/determiner phrase is embedded within a verb phrase, as we just saw – which is not recursive, since this involves a constituent being embedded within a different kind of constituent. But in (1c), the clause “Jürgen read a book” is embedded within another constituent of the *same* type, viz. the *clause* “Kwame said Jürgen read a book” – which makes this an example of a constituent embedded within itself, i.e. a recursive embedding.²³ Note that it is not the actual surface structure that is embedded within itself but rather the abstract grammatical category, in this case the category known as a clause. (Generative grammar is all about abstract categories and the formal principles that operate on them!)

Grammar in general makes our linguistic abilities *creative*, because it allows us to create new grammatical structures that have never been generated before from the finite set of words we learn as part of learning a language. (Such as the sentence “this is the first paragraph about linguistic creativity in Somangshu’s Princeton dissertation on Minimalist approaches to musical grammar”.) But its *recursive*, hierarchical structure makes our linguistic abilities *infinitely* creative. For instance, we can recursively embed (1c) within another clause to get (1d).

(1d) Yukiko heard (that) Kwame said (that) Jürgen read a book.

But we can continue to embed that clause in another, and yet another, self-similar clause *ad infinitum* so that our creative use of language is boundless – the only thing that prevents us from uttering such an infinitely long sentence being *performance* (as opposed to competence) related factors such as limitations on our memory, the muscular fatigue that would eventually affect our ability to utter such a sentence, and ultimately our mortality:

(1e) ... Selena loved (that) Yukiko heard (that) Kwame said (that) Jürgen read a book ...

Importantly, a hierarchical and recursive structure not only characterizes linguistic grammar; it characterizes musical grammar too. To understand this, consider Example 1.1-2, which discusses the

²³ Also, if noun and verb phrases are examples of a more general type of constituent, an idea we will explore in the next chapter, then even “read a book” might be considered an example of recursive embedding.

main theme from the second movement of Mozart's K. 364 Sinfonia Concertante for solo violin, solo viola and orchestra. The top stave in the example presents the second appearance of this theme from the end of measure 8 to measure 16, as played by the solo violin, after its initial presentation by the orchestra in mm. 1-8. A reduction of the orchestral accompaniment in mm. 8-16 is also shown, in the bass stave of this system.

As the Roman numerals under the system illustrate, the passage is in C minor, which is the initial harmony heard in the passage in measure 9 too. The passage ends on the downbeat of measure 16, on a C-minor chord in root position, which clearly reveals the C minor architecture of this passage. Something interesting happens at the beginning of the previous measure though. The chord here has the notes C, E-flat and G, just as in a C-minor chord – that has the harmonic function of tonic at the beginning and end of this passage. However, in this measure (i.e. measure 15) the G of the chord is in the bass, so the chord is a second inversion C-minor chord, unlike the root position C-minor chords that appear at the beginning and end of the passage. This means that the two other notes of the chord, E-flat and C, appear above this G at intervals of a sixth and fourth respectively – as shown by the Arabic numerals under the system. This 6/4 voicing makes the chord a rather unstable, dissonant sonority, especially because of the fourth, which must resolve downwards to B-natural to form the more consonant interval of a third with the bass G, just as the E-flat must resolve downwards to D to form the more consonant interval of a fifth with that note – both of which happen right at the end of the measure, as the Arabic numerals indicate there. Consequently, the second inversion C-minor chord at the beginning of the measure must resolve to a root position G-major chord, since this is what the chord at the end of the measure is (with the notes G, B-natural, and D). In other words, rather than having tonic function (as it does at the beginning and end of the phrase), the C-minor chord in measure 15 functions as a complex of accented passing tones over the *dominant*-functioning G-major chord that ends the measure – hence the notation $V^{6/4 - 5/3}$ under the system here.

Example 1.1-2. Mozart, Sinfonia Concertante, K. 364/ii: Melody of the theme and its reductions, mm. 8-16

The image displays a musical score for the second movement of Mozart's Sinfonia Concertante, K. 364, measures 8-16. The score is organized into three main sections:

- Full Score:** The top section shows the original orchestration. It includes a *solo violin* part with a melodic line and various ornaments (trills, mordents, grace notes) and a *strings* part providing harmonic support. The tempo is marked *Andante*. The key signature has two flats (B-flat and E-flat), and the time signature is 3/4. The score concludes with the instruction "etc." and a chord symbol $V^6_4 - 5 \ 3 \ i$.
- Reduction 1:** This section shows a simplified version of the melody. It consists of a single treble clef staff with a few notes, and a bass clef staff with a few notes. A chord symbol $V^6_4 - 5 \ 3 \ i$ is placed below the bass staff. The label "Reduction 1" is centered below the staves.
- Reduction 2:** This section shows a more detailed reduction of the melody. It features a treble clef staff with a melodic line and a bass clef staff with a few notes. A chord symbol $V^6_4 - 5 \ 3 \ i$ is placed below the bass staff. The label "Reduction 2" is centered below the staves.

Arrows indicate the relationship between the sections: a vertical arrow points from the full score down to Reduction 1, and a horizontal arrow points from Reduction 1 to Reduction 2.

The above harmonic structure of the passage implies that if we treat the music in measure 15 as a phrase, we might call it a dominant or V-phrase because the stable, controlling harmony here is G major. (Note that I am using “phrase” in an unusual manner here – and I will discuss this more in the next chapter.) Since the dissonant, second inversion C-minor harmony resolves to it, it can be considered as embedded within this larger V-phrase. But at the larger level of the entire passage, this G-major, V-phrase is embedded within the larger tonic ‘clause’ of this *C minor* passage. This clearly reveals the hierarchical structure of the passage. But moreover, the fact that we have a C-minor chord embedded within a G-major phrase that is, in turn, embedded within a larger C-minor clause here suggests that the C-minor chord is embedded within *itself* at the level of the clause – which suggests that the structure of this passage is recursive too.

I will develop this point more with the help of another example in just a bit, but before I go there, it is worth mentioning that the $V^{6/4}$ structure, within a V-phrase and embedded within a larger tonic structure, is an *extremely* common phenomenon in the grammar of Western common practice, tonal harmonic music, known as the “cadential 6-4” (Aldwell and Schachter (2011): 181-190; Kostka, Payne and Almen (2013): 135-136). So, if this is a true example of recursion in Western tonal music, it suggests that recursion is extremely common in this idiom as well.

Now, notice how the bass note C in measure 9 of Example 1.1-2 reappears in the bass in measure 13. But here it harmonizes the solo violin’s E-flat in the treble clef, as opposed to the solo violin’s C at the beginning of measure 9. Since the C to E-flat interval traced by the solo violin is part of the C-minor tonic harmony, we could say that in the course of mm. 9-13 Mozart “composes out” the C-minor harmony, especially since the tonic note C is sounded in the bass at the beginning and end of these measures too. This idea is portrayed more clearly in the simplified, abstract depiction of this passage in Reduction 1, in the middle system of the example. Since this depiction is an abstract one, it only depicts the pitch relationships of this passage, abstracted away from their temporal (and metrical) realization in the surface structure of the top system. (The barlines and note stems are shown here only for convenience in relating the reduction to the surface.) As we see here, the solo violin composes out the C – E-flat

interval by means of an ascending melodic line C – D – E-flat, all governed by the controlling C-minor harmony here, as shown by the arrow in the bass. This illustrates how this part of the passage can be considered a tonic-phrase in C minor. This tonic phrase is followed by a diminished triadic harmony in measure 14 on the second scale degree D (not shown in Reduction 1, but which has the notes D, F, A-flat, as can be seen in the top system). This harmony is considered to be a pre-dominant harmony as it leads to the dominant, V-phrase of the next measure, which we discussed above. The $V^{6/4}$ nature of the first sonority in this measure can be seen clearly now, as can the following $V^{5/3}$ sonority, in which the two notes that make the intervals of a fifth and a third with the bass are played successively by the solo violin (as the two bracketed notes illustrate). This dominant phrase is followed by the tonic chord that ends the passage, which yields the V – I cadential structure of the last two measures.

Finally, Reduction 2 at the bottom of the example gives an even more abstract depiction of the passage. Here, the initial tonic phrase is reduced to a single chord, viz. the C-minor chord in measure 13, since this chord represents the ‘initial tonicness’ that the phrase is all about, as does the head of a phrase in linguistics (more on this latter concept in the next chapter). This chord represents ‘initial tonicness’ because the bass note of the chord emphasizes the C-minor tonic character of the phrase, and the top E-flat represents the ‘initialness’ of the phrase, as it has to descend to the *final* tonic in measure 16 (after initially ascending from this tonic in mm. 8-13) to complete the phrase, via the D of the G-major dominant chord in measure 16. All of this can be captured by the abstract 3 – 2 – 1 descending scale-degree melody in the top voice of Reduction 2, harmonized by the I – V – I harmonic progression in the bass. In other words, Reduction 2 depicts an abstract, unarticulated (and therefore arrhythmic) deep structure from which the surface of the passage is generated.

Through all of the above we can see how a musical passage has a hierarchical grammatical structure, which can be represented by an abstract deep structure very similar to the similar structures proposed by generative linguists. This already shows how striking are the formal similarities between music and language. These similarities become even more striking when we see the recursive nature of musical grammar vis-à-vis linguistic grammar. So, on to Example 1.1-3 now. The top system of this

example just depicts the above passage from mm. 8-16 of the Sinfonia Concertante's second movement again, for ease of comparison with a later passage from this movement, viz. the cadenza that begins in measure 121, shown in the bottom two systems of the example. As before, the top staff of this latter passage is the solo violin part, with the solo viola's cadenza part shown in the lower staff.

Now, compare the top (solo violin) staves of these two passages, especially the first five measures of each until the downbeat of the sixth measure. These measures are quite similar in both passages. The only major difference between them is the pickup to each alternative measure in the first five measures of the passages, marked with an asterisk in both cases. In the upper passage, this pickup is made of a G – A-flat – G neighbor motive (with the A-flat further embellished by a B-flat grace note) – but this is essentially a more embellished version of the solitary G that acts as a pickup in the lower passage. In this light, the first five measures are essentially identical in both passages. (Notice also how in the lower passage, the E-flat of the initial ascent from the tonic C is reached in the fifth measure of the melody. Just as in the upper passage.) In the sixth measure, the lower passage also moves into the D-diminished predominant harmony as the upper passage does, although the melody of the solo violin here is different. The bass F-natural at this moment (i.e. the lowest note of the solo viola part) eventually moves to a G as happens in the upper passage, although this happens two bars later, in the eighth measure, since the F-natural moves to G via an F# in the intermediate measure.

Reaching this G might seem like the initiation of the cadential $V^{6/4} - 5/3$ phrase, as it did in the upper passage. But notice the chord right at the beginning of this entire (lower) passage with the fermata sign on top of it – it is a second inversion C-minor chord. This means that the whole passage up to the point where the solo viola plays the bass note G is *already* part of a $V^{6/4} - 5/3$ phrase. The C-minor melody we have been hearing, and which constituted a tonic phrase in the upper passage, is now part of a *dominant* phrase, and specifically the $V^{6/4}$ sonority that will eventually resolve to a $V^{5/3}$ harmony. So, in the eighth measure of this passage, a cadential $V^{6/4} - 5/3$ phrase does not have to be initiated – we are already in one. All that needs to happen is that the $V^{6/4}$ needs to resolve to $V^{5/3}$, which then has to resolve

Example 1.1-3. Mozart, Sinfonia Concertante, K. 364/ii: Melody of the theme in the (a) exposition and (b) cadenza, mm. 121-122

The image displays a musical score for Mozart's Sinfonia Concertante, K. 364/ii, focusing on the melody of the theme in the exposition and cadenza. The score is divided into two main sections: (a) exposition and (b) cadenza.

Section (a) Exposition: This section begins with the tempo marking "Andante". It features a solo violin part and a string ensemble. The solo violin part starts with a melodic line, and the strings provide accompaniment. The score includes dynamic markings such as *f* (forte) and *p* (piano). The exposition concludes with a fermata and the instruction "etc.".

Section (b) Cadenza: This section is marked "Cadenza" and is performed by the solo violin. It begins with a fermata and the instruction "Cadenza solo violin". The solo violin part is highly technical, featuring rapid sixteenth-note passages and trills. The score includes dynamic markings such as *f* and *p*. The cadenza concludes with a fermata and the instruction "etc.".

Instrumentation and Performance Instructions: The score includes parts for solo violin, strings, and cellos/basses. The cellos/basses part is marked "cellos/basses" and includes a fermata and the instruction "etc.". The score also includes performance instructions such as "Cadenza solo violin" and "cellos/basses".

to a root position C-minor chord, with the tonic note C in the melody as well, since this will end the passage as it did at the beginning of the movement. In a dazzling display of creativity, Mozart delays this $V^{5/3}$ for eleven measures until the last complete measure of the passage, where the two notes of the $V^{5/3}$ that form the fifth and third with the bass are played as a double trill by the two soloists. During the period of delay, the soloists engage in a touching back-and-forth that repeatedly hints at this approaching V (and the subsequent I), but none of these hinted sonorities can participate in the final, cadential V – I progression, since they are either not in root position or lack the appropriate notes in the melody to give the passage closure. When the $V^{5/3}$ finally arrives at the double trill, the passage quickly moves to the final tonic C-minor chord on the downbeat of the next measure, whose root is sounded by the orchestral cellos and basses, which brings the passage to an end.

Now, notice that the bass note G of the $V^{5/3}$ sonority is not sounded in the actual passage because it is implied by the *heard* G in the $V^{6/4}$ sonority at the beginning of the passage – as indicated by the arrow marks. As a result, this whole passage can be seen as an instance of a $V^{6/4 - 5/3}$ structure. This point holds up historically and stylistically too. Classical concerto cadenzas were often improvised by soloists in concert, and thus not written into the score – and usually the only indicator, in the score, for where the cadenza should appear is an orchestral $V^{6/4}$ chord (usually with a fermata), followed by a ‘trilled’ $V^{5/3}$ chord, leading to I. (Anyone familiar with the Classical concerto would also be familiar with the long trill played by the soloist at the end of the cadenza – a very common phenomenon.) Since the cadenza was meant to be played between the $V^{6/4}$ and $V^{5/3}$ sonorities, taken as a whole it therefore becomes essentially an instance of a $V^{6/4 - 5/3}$ phrase.

What is truly remarkable about this is that we could use the whole cadenza now, given its $V^{6/4 - 5/3}$ structure, to replace the $V^{6/4 - 5/3}$ part of the main theme, as the following schema shows:

	[main theme]
phrase structure:	[initial tonic phrase w/ ascent]	[predominant area]	[cadenza w/ cadence]
harmonic structure:	I	II	$V^{6/4 - 5/3}$ I

But since the cadenza sounds essentially the same as the main theme in its statement by the solo violin at the beginning of the movement – especially in the first five measures – the above schematic results in a particularly striking case of recursion in music, with the cadenza being embedded in the self-similar main theme. This recursion is striking because not only is it a recursion in the abstract sense, i.e. of a harmonic structure as $V^{6/4}$ embedded within a self-similar harmonic structure as I , it is also a recursion in a very surface sense, since it involves a sequence of actual notes (i.e. in the cadenza) being embedded inside the same sequence of actual notes (in the main theme) – a kind of recursion rarely seen even in language.

To be fair, such structures are normally not seen in music either. But this is usually due to reasons of musical design and form (as in “sonata form”) – there are no *computational* reasons why such structures should not occur in music. (In other words, such structures are fully possible within human musical competence, but performance factors relating to their stylistic use prevent them from appearing frequently.) But even if such surface recursion is rare in music and language, it frequently occurs in the more abstract sense, as I suggested before – and Example 1.1-3 just hammers in how striking even this is, given that abstract structures that are recursively embedded in each other might easily manifest themselves as actual surface structures recursively embedded in each other; a possibility the second movement of Mozart’s *Sinfonia Concertante* raises quite compellingly.²⁴

²⁴ This Mozart example is unique in that it shows us the different harmonic treatments of the same theme, as the main theme and as the theme of the cadenza; something that is rarely seen in the concerto literature – thus providing a strikingly vivid, and possibly unparalleled, illustration of musical recursion in Western tonal music. There are several reasons for why such an example should be rare. For one, cadenzas were rarely written down until the 19th century, which severely limits the number of extant examples one might find for them. Secondly, the written cadenzas of the 19th century concerto, and those that were added to 18th century concertos by later composers, are often highly virtuosic, complete pieces within themselves, which do not lend themselves well to being part of (i.e. embedded within) a larger tonal progression as the traditional Classical cadenza did. (But such Romantic cadenzas are really quite unfaithful to the original purpose of the Classical cadenza anyway, given their virtuoso provenance, exaggerated length, and relatively little use of thematic material from the concerto (Abert (2007): 898).) Finally, since the cadenza traditionally appeared towards the end of a movement, at a point when the main theme (especially in sonata form movements) had already been restated/recapitulated, composers usually chose secondary or closing thematic material for their cadenzas, if they used thematic material at all. Since such material is often presented in a different key initially (usually the dominant), it cannot realize the idea of a $V^{6/4}$ -harmonized theme being recursively embedded within itself in its *tonic* form. Which makes Mozart’s *Sinfonia Concertante*’s use of main theme material in a cadenza rare. Among the few other examples of this phenomenon is the cadenza written by Joseph Joachim for the finale of Beethoven’s *Violin Concerto*, Op. 61. Joachim uses the main, Rondo, theme of the movement in the cadenza – but even this example is not as succinct as the above Mozart one, since Beethoven treats the ending of the

If the above arguments are valid, then hierarchy and recursion are critically important aspects of both musical and linguistic structure.²⁵ The problem this poses for evolutionary theories about music/language is that such theories then have to show how recursive, hierarchical structure came about in music/language, especially if one believes that these came about gradually from a simpler, sonic, proto-musilinguistic state. As the linguist Derek Bickerton argues, a system can either be hierarchical or not – it cannot be partially hierarchical (Bickerton (2001): 158-159). This suggests that the hierarchical and recursive structure of music/language must have arisen in one fell swoop, rather than gradually – which, again, shows how problematic it can be to under-complicate musical/linguistic structure when speculating about their origins in purely sound-based terms.

The hierarchical, recursive structure of music/language, and the infinite creativity it allows, is also one of the reasons why Chomsky has argued for the innateness of language, i.e. the idea that it is hardwired and genetically-inherited, which we can extend to music given the above formal similarities between them. Just thinking about this intuitively, one can see why it would be impossible to learn an infinite number of surface structures – for this is what we would have to do if the ability to create an infinity of surfaces did not come hardwired in our minds.²⁶ But if music and language are genetically inherited because of their common hierarchical and recursive structure, this just adds to the argument that they have similar or joint biological foundations.

There is a complication here though, which is that the hierarchical and/or recursive aspects of musical and linguistic structure do not seem to be shared between humans and non-human species. And this complicates the issue of how music/language arose from an ape-call or birdsong-like proto-musilinguistic. To understand this, recall that attempts to understand music and language in terms of their social, and especially communicative, functions, and in terms of their sound structure have been popular

cadenza most unusually, modulating *away* from the tonic there by a series of written out trills, rather than ending it on the tonic by means of a straightforward V – I progression.

²⁵ For a justification of these issues from a different perspective, see Hofstadter (1979).

²⁶ Unsurprisingly, those scholars who reject the hierarchical and recursive structure of music (e.g. Narmour (1977)) and language (Elman, Bates, Johnson et al. (1996)) also reject their innateness, and argue that they have to be learned.

and pervasive, in both cultural and biological approaches to these systems. Interestingly, such attempts can be justified in light of the fact that various non-human species have been shown to have the ability to communicate, and process musical or language-like sounds, even though they do not have music or language – which suggests various pre-linguistic or pre-musical avenues from where human music and language might have arisen. For example, non-human animals have been shown to be sensitive to various perceptual features in speech sounds that are considered vital for speech perception, and to which only humans were thought previously to be sensitive. Patricia Kuhl and James Miller famously confirmed that even a squirrel-like rodent known as the chinchilla can perceive the difference between different categories of phonemes like /d/ and /t/ due to their sensitivity to the difference in the phonemes' voice-onset times (which is essentially a difference in how categorically-different phonemes are spoken) (Kuhl and Miller (1975, 1978)). Franck Ramus and his colleagues have shown that New World cotton-top tamarin monkeys can distinguish between the different ways in which bisyllabic words are stressed in different languages (e.g. stress-unstressed in English vs. unstressed-stressed in French and Italian), an ability critically important for human infants to acquire the prosodic patterns of their native language (Ramus et al. (2000)). And these results do not just pertain to language, since we have already explored some research that suggests that certain non-human animals are sensitive to perceptual features in *music* as well, e.g. in the results regarding rhesus monkeys' ability to comprehend certain (albeit restricted) pitch relationships, such as the octave (Wright et al. (2000)).

Moreover, those aspects of language that have to do with its use in communication, viz. its conceptual and intentional aspects (broadly “semantics”), also seem to be possessed by some non-human species, even though they do not have human language. Evidence for this is found widely in the group behavior of social animals like chimpanzees, and also includes this species' well-documented, thoughtful use of tools. Such evidence seems to be less available for music, primarily because aspects of musical meaning in the behavior of non-human animals has not been well-studied. But the emotive (e.g. alarm) calls of various non-humans might be evidence for a shared sensitivity to the emotive aspects of musical meaning in both humans and non-humans (Hauser and McDermott (2003): 666). Moreover, young male

songbirds often sing something called “subsong”, which is a kind of ‘practice’ for later adult performances, and even adult songbirds sometimes sing “whisper songs” quietly to themselves (Fitch (2006): 184), which could be further evidence for a shared ‘performative’ use for ‘vocal music’ in humans and non-humans.

But in marked contrast to all of the above, non-humans do not seem to possess the ability to process a recursive grammar. This has led some scholars to claim that this is an aspect of language that is uniquely human, and constitutes a special “narrow” faculty of language (or FLN) that can be distinguished from language in the more broad sense (FLB) that includes its sound and semantic aspects, and large parts of which might be shared with various non-human species (Hauser, Chomsky and Fitch (2002)).^{27,28} But it is exactly this ability to process recursive grammar that lies at the core of C_{HL} or C_{HM} , and which constitutes the basis for human music/linguistic competence. So, if this is unique to humans, it is also a unique part of human nature – and possibly what makes humans *human*, why music and language are *species*-specific traits of humans.

Now, given that a study of human musicality/linguisticity that sees these faculties as aspects of human nature has to focus on musical/linguistic competence, the uniqueness of this competence makes it very hard to find explanations for the biological origin of music and language, and more so for their *joint* evolution. Of course evolutionary theorists have mainly focused on the non-unique aspects of language we share with our non-human ancestors, since that is the only way one can explain language/music evolution; but unless such an explanation can explain why *human* language/music are the way they are, which we cannot without an exploration of the uniquely human ability for recursive grammar, such an

²⁷ Even though Hauser, Chomsky and Fitch concentrate on the recursive aspects of language, as part of their description of FLN, they accept that humans might have extended their ability to process recursive structures to other domains (Hauser, Chomsky and Fitch (2002): 1578). They are also open to music being one of these domains (Fitch, Hauser and Chomsky (2005): 182). However, they are cautious about this claim – but this seems to stem more from a neglect of the grammatical aspects of music, and an unfortunate (though common, as we know) emphasis on the sound, or “phonological” as they say, aspects of music (Fitch, Hauser and Chomsky (2005): 200).

²⁸ Timothy Gentner and his colleagues suggest that at least one songbird species, the European starling, can process recursive grammatical structures (Gentner, Fenn, Margoliash et al. (2006)), but Caroline van Heijningen and her colleagues suggest that just an ability to process certain sounds (e.g. phonemes), and possibly simple *non*-recursive grammatical rules, lies behind this ability of starlings, and other songbirds such as the zebra finch (van Heijningen, Vissera and Zuidemaa (2009)).

explanation will lack explanatory adequacy. It is considerations such as these that have prompted Noam Chomsky to even say that “language is based on an entirely different principle than any animal communication system, so it is a complete waste of time to ask how it arose from calls of apes and so forth” (Chomsky (1988): 183).²⁹

This statement of Chomsky’s really just repeats the point I made earlier, about how attempts to explain music and language must deal with the complex formal structure of these two domains if they do not want to end up being essentially caricatures of them. But does the inability to explain the evolution of recursive grammaticality imply, then, that we cannot explain the origins of music and language – and by extension the joint foundations of music and language in our biology? In other words, does this mean that we have to abandon a biological approach towards defending the musilanguage hypothesis, just as we had to abandon a cultural one in our earlier discussion of John Blacking’s ethnomusicological thoughts on music and language, as species-specific traits of humankind? Well, this would only be the case if we take a gradualist or adaptationist perspective to these issues, which we do not need to do for a start. We can begin instead from the premise that the explanation for musical/linguistic evolution must deal with the complex formal aspects of music and language, and that a different, non-adaptive, non-gradual evolutionary process might explain how such formal properties arose in music and language during the course of their evolution.

Such a perspective is consistent with proposals that linguists like Chomsky and Bickerton have made about language, and which we might extend to music now, which claim that language (and by extension music) arose in one fell swoop, complete with its hierarchical, recursive architecture, and that this did not emerge gradually from the vocal music of non-humans through the workings of natural

²⁹ This is also why some scholars who accept the complexity of language and music, but also wish to explain music/language origins in Darwinian adaptationist or gradualist terms, either reject, in the case of language, the distinction between FLN and FLB – and continue the pursuit of an evolutionary explanation for the complexity of language through FLB, and its sonic or communicative aspects (e.g. Molino (2001): 169-171, Pinker and Jackendoff (2005)) – or argue, in the case of music, that the observed complexity of the system is really the epiphenomenal result of the interaction of a variety of other systems like language, emotion, auditory perception etc., which one might call “auditory cheesecake” (Pinker (1998): 534).

selection. Instead, it might have already been there, in its complete form, as the result of a genetic mutation in our earliest human ancestors. Alternatively, the ability for language and music might have existed in a non-human species but was used for non-linguistic/non-musical purposes – and did not evolve, but only had its function changed to its current linguistic/musical role, in modern humans, a phenomenon some biologists have referred to as “exaptation” (Gould and Vrba (1982), Gould (1991)).³⁰

Arguments such as these suggest that the search for the origins of music and language – and a defense of music/language identity – can still be a biological research program, but just a different, more formalist one than the ones we looked at above. Indeed, this is why the current Minimalist Program in generative linguistics is often referred to as a research program in “biolinguistics” as well (Jenkins (2000), Berwick and Chomsky (2011)). As opposed to some of the more experimental approaches we have looked at above, this formalist, biolinguistic program of research is more of a theoretical science, given its foundation in ideas from theoretical, Chomskyan linguistics. This is not surprising, since it is only linguistic theory – and music theory, as we shall soon see – that pays any serious attention to the grammatical structure of language and/or music, which is necessary to address the issue of human musilanguage. And a theoretical-scientific approach to the issue of human musilanguage is what a joint Minimalist study of music and language really is anyway, so this is the approach I will continue to defend in the course of this dissertation.

But before we move on to this topic, there is one last issue within the experimental scientific approach to music and language relationships that I would like to discuss. This has to do with the fact that the formal, grammatical study of music and language I have been advocating is essentially a study of human psychology, i.e. of the computational system that underlies how grammatical structures are

³⁰ Charles Darwin was himself sensitive to this possibility, since he argued that certain biological structures such as wings in birds and bats were originally meant for cooling, which later had their function changed to that for flight. In this light, it is interesting to note that there have been proposals in evolutionary biology according to which the increase in hominid brain size, which, as suggested earlier, allowed for language to arise in humans, might have originally happened for cooling reasons too – a larger brain is easier to cool in the hot savanna areas that our hominid ancestors inhabited because it has more surface area for letting heat out (see Falk (1990)). But instead of giving this phenomenon a name like the modern “exaptation”, Darwin gave it the inaccurate name “preadaptation”, which is problematic, because it does not make sense from an adaptationist perspective for evolution to create structures that have no immediate adaptive function, but are merely steps toward a final function that might arise possibly millions of years later.

generated in music and language. So, experimental approaches to the study of music and language that focus specifically on the computational aspects of musical and linguistic grammar – independently of their cultural or evolutionary function, and independently of their connection to the sound and semantics of music and language – might shed light on the nature and origin of human musilanguage. In particular, any computational similarities between music and language that such studies reveal might be taken as evidence for the joint foundations and origins of music and language. And there has been some work done along these lines within the cognitive and neurosciences, so I would like to discuss some of these studies before turning to the more theoretical approach to these issues found within the Minimalist Program.

The first set of studies I would like to discuss represents some recent work in the neurosciences. All of these studies focus on the brain centers that are supposedly involved in musical/linguistic grammatical phenomena. For a long time, it was believed that music and language are processed in different parts of the brain, as exemplified by the popular myth that “music is in the right brain and language is in the left brain”. In recent years, this belief has received some experimental support from individuals with severe language deficits resulting from brain damage (such as aphasia) but who seem to have intact musical abilities – and who can be contrasted with other individuals with severe *musical* deficits (e.g. amusia) but who seem to have intact *linguistic* abilities – suggesting a neurological dissociation between music and language. The neuroscientist Isabelle Peretz suggests on the basis of this and other results that music and language are separate “domains” of the mind (Peretz (2006): 8-14).³¹

But neuroscientists like Aniruddh Patel have disputed these results, by suggesting that the differences in the above individuals are really the result of differences in the way music and language are coded in the brain, rather than in the grammatical processes that join these inputs into sentences. Moreover, neuroscientific studies that *have* focused on the grammatical processes involved in music and language have often revealed similarities, as opposed to differences, between music and language.

³¹ To be fair though, Peretz does not ascribe domain-specificity to music and language taken as wholes, but rather to specific components *within* music and language. That is, though there are specific parts of the cognitive system of music that might be music-specific, such as the encoding of pitch, there might be others that are shared between music and language.

Specifically, Patel and his colleagues have found that individuals with certain kinds of aphasia, i.e. those who have damage to language (and particularly linguistic syntax) processing areas, seem to have impaired musical grammar-processing abilities too (Patel, Iversen, Wassenaar and Hagoort (2008)) suggesting a common neural basis for these abilities. Moreover, ambiguities or deviations in grammatical structure seem to evoke similar responses in the electrical activity of the brain in both music and language (Patel, Gibson, and Ratner (1998)).³² Burkhard Maess and his colleagues have found that harmonically-inappropriate musical chords elicit electrical activity in the brain that seems to originate from Broca's area, a brain center widely believed to be the locus of syntactic processing in *language* (Maess, Koelsch, Gunter and Friederici (2001)), and these results have been extended to a larger cortical network in the brain, including Wernicke's area, which was previously thought to be specific to language processing too (Koelsch et al. (2002)). Similar observations made by Levitin and Menon (2003) and Brown, Martinez and Parsons (2006) led the former authors to say that such brain regions "may be more generally responsible for processing fine-structured stimuli that evolve over time, not merely those that are linguistic", and the latter authors, "music and language show parallel combinatoric generativity for complex sound structures (phonology) but distinctly different informational content (semantics)".³³ Finally, Aaron Berkowitz argues that there seems to be an overlap in the brain areas involved not only in musical and linguistic grammatical processing, but also in grammatical *production*, as seen in studies of speaking and musical improvisation (Berkowitz (2010): 150-151).³⁴

Despite the positive outcome such results might represent for the musilanguage hypothesis, Aniruddh Patel has said himself that musical and linguistic grammar involve *different* cognitive systems –

³² That is, in the "event-related potential" or "ERP" responses of the brain, specifically the ERP known as the P600, in which a positive spike in the brain's electrical activity peaks 600 ms after a relevant stimulus is presented.

³³ In light of the points made earlier though, Brown, Martinez and Parsons' claim about "combinatoric generativity" in music and language being about complex *sound* structures might be refined to being simply about "complex structures" given the problems associated with over-emphasizing the sound aspects of music and language.

³⁴ In a recent study, Evelina Fedorenko and her colleagues argue that the brain regions involved in high-level language processing (e.g. in sentence comprehension) are *not* shared between language and other cognitive processes like music (Fedorenko, Behr, and Kanwisher (2011)). However, their study seems to depend strongly on specifically linguistic stimuli, like words, which are clearly not shared with music. In other words, the study does not seem to address the grammatical, as opposed to lexical, aspects of music and language – which is where the similarity between music and language lies, at least according to the musilanguage hypothesis.

it is just that they share common neural resources in their workings (Patel (1998, 2003)). He has referred to this as the “shared syntactic integration resource hypothesis” for music and language, which is based on the above experimental results about music and language in the brain, and also the observation that musical grammar-processing tasks seem to interfere with simultaneous linguistic grammar-processing tasks (in a way that is not affected by similar non-musical/linguistic tasks, such as purely auditory ones) – suggesting that both these tasks make use of the same brain resources (Patel (2007): 282-298).

But a problem with some of these results is that they do not really deal with the computational aspects of musical and linguistic grammar, as a genuine musilinguistic study should. Rather they deal more with how music and language are *implemented* in the human brain. (This alludes again to David Marr’s three-level description of cognitive systems, referenced earlier.) Given that music and language involve inputs that seem to be different at least on the surface (i.e. words in language vs. pitches in music), it is quite possible that different neural systems are involved in processing these different kinds of stimuli – but that does not mean that the grammatical principles involved in the processing of these stimuli are different too. This is an important point because until we know what a music-grammatical computation is we cannot determine whether it is similar to or different from a linguistic one. Put in a different way, we know that wh-movement exemplifies a kind of computation in language. But how is this specific computation implemented in our neural hardware – where are the question words involved in wh-movement stored in the brain, for example? And does music have something akin to wh-movement – and where would this be implemented in the brain? The neuroscientific work discussed above has not addressed questions like this so far though – so, Patel’s claim that the overlap between music and language lies only in the brain resources they share, not between the systems themselves, does not seem to have much teeth.³⁵

³⁵ Even the assumption that the locus of music/language overlap lies in shared brain resources is problematic given that complex behaviors like music and language are usually spread out over the human body, and not just restricted to the brain. Single cell organisms, by definition, do not even have brains, yet are capable of complex behaviors like feeding and fleeing from danger. But how the brain connects with other bodily systems is poorly understood, since it is hard to tell whether a given cell is going to become a neuron or some other cell in the body during embryonic development, even when we have the complete description of the DNA of a fertilized egg. The process is even more

The other set of studies that appear to focus on specifically grammatical issues in music and language represents work in the field of cognitive psychology. For example, Jenny Saffran and her colleagues have focused on how we process strings of linguistic information, such as syllables, to understand how our minds parse such strings into words and the like (since this is relevant for speech recognition and sentence comprehension). They have found that even 8-month old human infants can reliably parse such strings based on statistical cues in the string, suggesting an innate psychological capacity for learning the statistical structure of words (Saffran, Aslin and Newport (1996)). However, they have found that humans use the same statistical learning mechanisms in parsing the structure of musical sequences too (Saffran, Johnson, Aslin and Newport (1999), Saffran (2003)), and that sequentially presented information, as found in language, facilitates our performance in musical behaviors, but not in behaviors that involve simultaneous (i.e. spatial, rather than sequential/temporal) information, such as vision (Saffran (2002)). This suggests that temporal information-processing systems like music and language are computational systems of a kind.³⁶ Finally, it has been argued that even though adult language and music might seem very different on the surface, they might appear to be very similar to an infant learning them, whether based on the above statistical mechanisms or not (McMullen and Saffran (2004)).

This last idea, that of musical and linguistic abilities being similar because of the common predispositions infants display when learning music and language, provides a different set of data for the musilanguage hypothesis, i.e. from studies of human development. The developmental psychologist

complicated if information about the environment in which the embryo develops has to be integrated into a description of how it will develop, if one accepts the common belief that an individual's DNA is structured in ways that allow for survival in different environments. (E.g. frog DNA is supposedly structured in a way that allows tadpoles to survive in different temperatures, which is not a problem mammals face and hence something not found in the structure of mammalian DNA.) Our lack of knowledge about how the brain connects with other bodily systems is a problem though, because understanding this connection is crucial for explaining various complex human behaviors, especially when the mental aspect of a behavior has significant bodily consequences. This can be seen in the complex behavior of pitching a baseball, where mentally rehearsing the pitching of the ball has been shown to improve one's actual, physical, throwing of the ball. (For more on this, see Uriagereka (1998): 53-61.)

³⁶ This is an important point because scientists interested in how humans process information have tried to develop computational models of a variety of human abilities ever since the birth of modern cognitive science, one of the most famous being David Marr's aforementioned work on vision (Marr (2010)). In this light, the idea that music and language have a peculiar computational system, not shared with other systems (like the visual one), is an important piece of evidence for the musilanguage hypothesis.

Sandra Trehub observes that there does not seem to be a huge difference between the musical abilities of infants and those of adults, in psychological terms. For instance, *both* adults and infants tend to focus on three properties of musical sequences when deciding whether one sequence is structurally different from another, viz. melodic contour, rhythm and grouping structure (Trehub (2001): 428-431). Adults and infants will uniformly judge a sequence of musical pitches to be invariant from another, as long as the contour of the sequence is maintained across them, and as long as they have the same rhythm – even if one of the pitches is changed in the other sequence, or if it is played faster or slower than the original. This is actually quite a sophisticated ability, which is implicated in some of the more complex, formal aspects of musical structure, such as transpositional equivalence between musical phrases.³⁷ This is in marked contrast with the abilities of non-human species, who as we saw earlier will not recognize two sequences as being invariant if the pitches are changed from one sequence to the other, the only exception being if the pitches are changed to those one or two octaves above or below the original (Wright et al. (2000)). What is even more remarkable though, is that even though these musical abilities shared between human infants and adults are not shared with non-humans, they are shared with human *linguistic* abilities, since “contour, rhythm, and perceptual grouping principles are important for perceiving and remembering spoken as well as musical patterns” (Trehub (2001): 431).

In addition to those aspects of musical and linguistic psychology that are shared between infants and adults, there are some aspects that are different and have to be learned. But even these aspects seem to be shared between our capacities for music and language. For example, infants are unable to distinguish pitches that belong to certain chords from pitches that do not belong to them, suggesting that chord structure is something that has to be learned. But chord structure is analogous to word structure in language as I discussed earlier, and children have to learn the specific words and word orders of their

³⁷ It is important to note that both adults and infants can detect changes of the above kind between sequences – it is just that these changes are deemed irrelevant for the identity of the two sequences, which is why they are judged as being invariant. The reason why this is a sophisticated ability is because it involves two steps: (a) detecting a change in pitch, e.g. through transposition, and then (b) deciding that the change in pitch is irrelevant. One could judge two sequences to be invariant because one could not even detect a change in the first place, which would be a less sophisticated ability (and which is in fact how infants perceive *non*-diatonic sequences, as long as the contour of the two sequences remains unchanged).

native linguistic cultures too – which is what I previously called the ‘formal’ aspects of sociocultural systems. But even if some aspects of both music and language have to be learned, it is well known that humans in general are innately equipped to learn any language in their infancy (as long as they do so within a critical period) – and even *this* seems to be true of music (Trehub (2003): 670).³⁸

Despite all of the above psychological data, many researchers in this field, Jenny Saffran and Sandra Trehub included, come to a conclusion that is equally disheartening for the musilanguage hypothesis, as was the neuroscientist Aniruddh Patel’s thoughts on the matter above. Instead of concluding that music

³⁸ All of these psychological results paint a picture of the human musical/linguistic mind as being a very sophisticated information-processing device, but one that even people without too much musical training or experience (such as infants) can have. In light of this, any claim that some kinds of music can only be processed by individuals who have a good deal of (cultural) experience of an idiom is worth re-assessing. This is not to say that specific idioms cannot have musical features that are complex and require extensive training; it is just that assumptions about the idiom-specificity of certain kinds of musical perception, stemming from an emic bias towards musicality, are worth questioning. John Blacking makes such an assumption when he criticizes the psychological aspects of the Berlin School comparative musicology for minimizing “the importance of cultural experience in the selection and development of sensory capacities” (Blacking (1973): 5-6). For example, he argues, Venda musicians conceive musical intervals harmonically, and would therefore be unable to distinguish harmonically ‘equivalent’ intervals like those of a fourth or fifth, which would make them appear to be “tone-deaf” in psychological experiments. Now, of course a psychological experiment that fails to correct for cultural differences is problematic, especially if it is informed by racial bias – as was the case with much Berlin School thinking. But is Blacking really saying that Venda musicians do not *hear* the difference between a fourth and a fifth, just because they do not *judge* them to be different? After all, someone might judge two sounds to be the same even if s/he clearly hears their difference, because this audible *structural* difference has no *functional* value for her/him. But to conclude from this that the listener does not even hear the structural difference between the sounds amounts to imposing an untenable functionalist (and emic) bias on human psychology – untenable unless we want to accept the equally untenable proposition that certain groups of people (in this case the Venda) have radically different brains and ears from us, which most scientists (or ear doctors for that matter!) would dispute.

Blacking reveals his emic biases again when he compares a Venda polyrhythmic pattern played by a single musician to an *identical-sounding* pattern played by two or more musicians (Blacking (1973): 29-30). Even though they sound identical, Blacking says that they should not be judged to be the same because the pattern played by two or more people serves a different function in Venda society – i.e. it signifies the relation between individuals in groups, which he believes to be an important social concept in Venda society. But musical patterns that can be played by a single musician or instrument are often played by more than one musician or instrument in other idioms too – e.g. when chords that can be played by one violinist (or group of violinists) in a Western orchestra are broken up and played by different violinists (or groups of violinists), in the technique known as *divisi*. In this case, however, the chords are still judged to be the same – and not for any functional reason. They are judged to be the same because even when the different notes of a chord are played by different musicians the resulting simultaneity that is heard is still, *structurally*, that same chord. But to insist that this chord becomes a different “sonic object” when its notes are played by different musicians, as Blacking says – just because this group activity signifies something different in a cultural context – amounts to imposing an emic, functionalist bias on even the analysis of a sonority’s structural status. In the context of Venda musical grammar, maybe those two polyrhythmic patterns have the same abstract structural status. But the truth of this matter will never be revealed to us if we accept Blacking’s emic approach. In consequence, this would derail any possibility of developing a genuine grammar of Venda music – and also a genuine understanding of human musicality, given the role of grammar in such an enterprise.

and language are different cognitive systems as Patel does, they interpret the above results as suggesting that music and language *are* similar cognitive systems, but that these similarities are shared by a host of other psychological systems as well – including some found in non-human animals (cf. Trehub and Hannon (2006), Saffran et al. (2007)). Rather than supporting the musilanguage hypothesis, this suggests instead that there is nothing special about the connection between music and language, especially as aspects of *human* nature.

However, this cognitive psychological conclusion is as problematic as Patel’s neuroscientific one. This is because *not all psychological phenomena are computational phenomena*, even if all computational phenomena *in music and language* happen to be psychological phenomena. So some of the psychological phenomena described in the above studies might have to do more with the *non-computational* aspects of the human mind – which might very well be shared with other, non-computational, cognitive systems, including those possessed by non-human minds. For example, consider the fact that the motor processes involved in playing an instrument, or moving one’s lips to sing (or speak), are definitely important aspects of musical behavior that have been widely studied by psychologists, but these do not necessarily require our minds computing grammatical information in the way we do when generating surface musical or linguistic structures. Similarly, the ability of our ears to hear sounds is indispensable for musical behavior, but our ears do not necessarily process algorithms when hearing sound in the way our minds do when processing musical information.³⁹ This means that a psychological study that relies on the motor or auditory aspects of music/language (the latter being something scientists are wont to do, as we have seen before) might reveal something about the computational aspects of music/language psychology, but might *also* reveal something about the non-computational (e.g. motor or sound) aspects of these systems. And if the latter is shared between music,

³⁹ In fact, much research on hearing suggests that our ears are biologically structured to hear sounds in a systematic way. For example, most auditory scientists believe that we can detect the frequency of a heard sound (and thus distinguish different kinds of sounds, which is necessary for all kinds of complex musical behaviors) either because the mechanical properties of the ear’s basilar membrane make different parts of it respond to different frequencies (an idea famously advocated by Hermann Helmholtz), or because cochlear neurons spike at different *rates* depending on the frequency of the heard sound (Kandel, Schwartz and Jessell (1991): 486-492). In other words, frequency detection by the ear is a *mechanical*, not a computational, phenomenon.

language and other psychological systems, it might lead one to conclude that the observed psychological properties of music and language are not specific to just music and language, or even to human psychology. Indeed, this seems to be why Sandra Trehub and Erin Hannon say that music is based to a large extent on “domain-general” psychological properties shared with non-human species (Trehub and Hannon (2006)) – since many of their experiments find similarities between music and language not on computational grounds (i.e. through experiments on grammatical processing) but on non-computational ones (i.e. through experiments that study general properties of pitch perception, much of which *are* shared between humans and non-humans, as we have seen).

Along similar grounds, the claim that music and language are not specific psychological domains, but rather general ones, because they are founded on statistical learning mechanisms shared with other cognitive domains (which is the conclusion Jenny Saffran and her colleagues reach from their above experiments) is misleading because such mechanisms might have very little to do with the computational foundations of music and language to begin with. The philosophers Jerry Fodor and Zenon Pylyshyn have argued that our minds need formal rules to process the unique computational structure of certain psychological systems like language, so that they cannot be learned from general, non-rule-based learning devices, such as those implicated in statistical learning mechanisms (Fodor and Pylyshyn (1988)). Moreover, there are languages in which certain grammatical phenomena occur only within circumscribed parts of the language (such as some kinds of noun pluralization in German and Arabic, which occur only within a small set of special nouns). This has led Gary Marcus and his colleagues to argue that learning such phenomena (i.e. how to pluralize a noun) has to happen according to certain grammatical rules, because a specifically statistical learning device *cannot* possibly learn them, given that such devices can learn successfully only when there is a high statistical preponderance of the stimulus to be learned (Marcus et al. (1996), but see Plunkett and Nakisa (1997)). In other words, the shared (statistically-based) psychological system that Saffran and her colleagues have described is not necessarily one that deals with

the grammatical aspects of language or music at all – but of course this is what one has to do if one wants to justify the musilanguage hypothesis on psychological or biological grounds.⁴⁰

The implication is that for a psychological study to really reveal anything about the possibly shared foundations of music and language, it has to focus on the specifically grammatical aspects of these systems – i.e. those aspects of the systems that form the basis for musical or linguistic *competence*. So the question is whether there is any such evidence that suggests that humans process the recursive or hierarchical aspects of both musical and linguistic structure? Well, the very fact that there are actual, observed utterances made by humans that can be described in recursive terms, whether they are made by language speakers or music makers as Examples 1.1-1.3 illustrate, should be sufficient evidence to answer this question in the affirmative. Moreover, such utterances are routinely treated as data by linguists and music theorists, on which to build scientific theories about musical/linguistic structure. This is why this dissertation takes a Minimalist, theoretical-scientific, approach to comparing musical and linguistic grammar. Fortunately, though, there has been (albeit quite limited) research in the experimental sciences that specifically explores issues of competence in music and language too, and with an emphasis on the specifically hierarchical and recursive aspects of musical grammar.

Perhaps not surprisingly, much of this work has been conducted in collaboration with theorists. So, the psychologist Carol Krumhansl, in collaboration with the music theorist Fred Lerdahl, has examined the issue of whether musical grammar is hierarchical or not, their conclusion being that musically trained listeners hear Western Classical tonal structure in a way that correlates more with Lerdahl's hierarchical music-theoretic description of this structure than with a non-hierarchical one (Lerdahl and Krumhansl (2007)). However, listeners' ability to perceive hierarchical structure decreases

⁴⁰ Interestingly, Saffran's statistical learning experiments have been replicated with cotton-top tamarin monkeys (Hauser, Newport and Aslin (2001)), which adds more evidence to the argument that Saffran's results do not really address the uniquely human, computational aspects of language. All of which correlates with the fact that the increasingly popular statistical study of musical phenomena within music theory is often done by scholars who deny the hierarchical, generative grammatical basis for musical structure (e.g. Temperley (2004a, 2007, 2009), Quinn and Mavromatis (2011), Tymoczko (2011)).

with the complexity of a piece, especially if it uses more dissonant, chromatic pitches in the surface. This raises the possibility that the mind can only process hierarchical information about musical structure through training (with increasingly complex pieces). But this contradicts the idea that it has an innate ability for processing hierarchical structure – at least for language in the way Chomskyan linguists describe the matter. However, Sandra Trehub notes that even though language is normally acquired early in life, much training through adulthood is required for skilled *linguistic* performance in activities like oratory or recitation (Trehub (2003): 669). So, musical training might facilitate more advanced skills in musical *performance*, but this does not imply that training is required for musical competence.⁴¹

Moreover, several experiments by Emmanuel Bigand and his colleagues have tested the possibility of our knowledge of musical grammar being innate like language too. For example, in one experiment they found that musically trained listeners did not perform better than untrained ones in implicitly learning a new artificial grammar of musical timbres (Bigand, Perruchet and Boyer (1998)) – suggesting that humans have a sophisticated innate capacity for acquiring musical idioms that is independent of musical training and experience. Another experiment tested the ability of listeners to recognize the common deep harmonic structure that underlies a variety of different surfaces, as exemplified by a musical theme and its (harmonically-similar) variations, which Bigand found even untrained listeners to be able to do at levels higher than chance (Bigand (1990)). And untrained listeners could also distinguish melodies that had different harmonic structures even when they appeared very similar on the surface, especially in terms of their rhythm and melodic contour – and these results were not significantly different from those observed with trained, musically experienced listeners (Bigand and Poulin-Charronnat (2006): 106-107).

This last result is particularly interesting in the light of Sandra Trehub and colleagues' above observations about the importance of surface rhythm and contour in the judgments of both infants and

⁴¹ This is especially important if a psychological experiment requires a participant to make explicit responses to a given stimulus. If knowledge of musical grammar is acquired implicitly or unconsciously, like language (as even John Blacking sensed), then explicit responses will really be testing a participant's musical performance and not their competence.

adults regarding the similarities or differences between melodies. But it also raises the issue of whether people rely on perceptual (as opposed to grammatical) cues to make such judgments, which was an issue raised regarding Trehub's results too. So to test for what kind of cues untrained listeners use in responding to musical stimuli, Bigand and his colleagues assessed whether listeners would judge the appropriateness of a target chord's presence at the end of a chord sequence based on the previous occurrence of that chord in the sequence (i.e. based on perceptual familiarity with that chord from its previous occurrence, a phenomenon known as "perceptual priming") or whether they would choose it because it is grammatically expected to appear there, irrespective of any perceptual cues to that effect. It was observed that both trained and untrained listeners were mostly influenced by the latter phenomenon, i.e. by "cognitive priming" (Bigand, Tillmann, Poulin-Charronnat and Manderlier (2005), Bigand and Poulin-Charronnat (2006): 111-112).

Finally, and again in collaboration with Fred Lerdahl, Bigand and his colleagues have found that there is no significant difference between trained and untrained listeners in their ability to comprehend the abstract nature of musical structures based on changing patterns of tension and relaxation in them, whether assessed in short chord sequences (Bigand, Parncutt and Lerdahl (1996)) or long ones (Bigand and Parncutt (1999)). This result applies even to large-scale musical structures, such as the exposition section of a Classical piano sonata, since it was observed that even untrained listeners were aided by a coherent presentation of sonata excerpts (i.e. one in which the excerpts were presented in the order they appear in the sonata exposition, thus preserving the abstract, large-scale structure of the sonata), when deciding whether a subsequent excerpt was new or old (Bigand and Poulin-Charronnat (2006): 114-115).

All of the above results lead Bigand and his colleagues to conclude that "an intensive explicit training in music is neither a necessary nor a sufficient condition to acquire a competence to produce music. ... These differences seem to be linked to the learning of motor skills specific to the playing of an instrument, to a greater familiarity to specific musical timbre in musicians, or to the development of very specific analytic perceptual processes" (Bigand and Poulin-Charronnat (2006): 121-125). In other words, musical and linguistic grammar seem to be similar at least in terms of how our minds are competent in

them, with the differences between musical and linguistic behavior often being a result of non-grammatical factors such as motor skill.

So far we have explored a number of empirical approaches to the music/language nexus as explored by ethnomusicologists and biological/cognitive scientists. As we have also observed, the vast majority of these approaches cannot be taken as evidence for or against the musilanguage hypothesis though, either because they avoid an exploration of musical or linguistic grammar in their methodology, or they focus on grammar but not from the requisite computational perspective – the one exception to this being the cognitive psychological research of scholars like Emmanuel Bigand and Carol Krumhansl.⁴²

So, what about a theoretical approach to the music/language nexus instead? Such a theoretical approach receives support from the linguist Derek Bickerton's claim that it is specifically an ignorance of *theory* that often prevents non-theorist scholars from accounting for the complex, computational aspects of systems like language (Bickerton (2001): 154-155), and by extension, music. Others have argued, in the case of music, that only psycho/biological studies are feasible at this point because of the paucity of music-theoretic views about the psychological structure of music (Hauser and McDermott (2003): 664) – implying, presumably, that the existence of such views would make other approaches unnecessary or less desirable.

As I have suggested several times now, I believe that the Chomskyan Minimalist Program in generative linguistics provides a way out of this conundrum, because it presents exactly the kind of theoretical, computational perspective that might give us a justification for the musilanguage hypothesis. So, it is now time to finally turn to a detailed exploration of this approach.

⁴² One could argue that even Bigand and Krumhansl's studies belong more to an algorithmic level of description than a computational one. This is because their experiments focus on the real-time processes through which inputs (the stimuli heard in an experiment) are converted to outputs (i.e. participants' responses in such an experiment), rather than on the computational structure of these processes.

1.1.2. A Minimalist Program for Language *and* Music

Let me start by reviewing the crucial points about music and language that have been the highlight of the arguments I have made so far. This is the fact that music and language are similar, and possibly identical, when (and only when) they are seen as a psychological system, marked with the uniquely human ability to compute information hierarchically and recursively. It is this psychological system – or more specifically, our theory of it – that is given the name “universal generative grammar” or just “grammar”, and which makes music and language aspects of human nature as well. Finally, it is the inability to account for musical/linguistic grammar that renders many current approaches to the issue of music/language identity impotent, especially when they focus on the evolutionary or sociocultural functions of music and language rather than its computational form.

The above focus on music and language as a psychological system implies a serious interest in the structure of the human mind, or what I just referred to as the “computational form” of the mind. In other words, it implies a serious interest in the structure of the computational systems of music and language (i.e. C_{HL} and C_{HM}), seen independently from their sociocultural and evolutionary functions. But such an interest, consequently, goes hand in hand with rejecting a purely ‘material’ approach to the study of music and language too. That is, such an interest resists reducing the study of the musical/linguistic mind to a study of the structure or evolutionary function of the musical/linguistic *brain*, for reasons we explored in the last section. For that matter, it also resists reducing the study of the musical/linguistic mind to a study of the sociocultural functions of the human *body*, as seen, for example, in the culture-specific study of the bodily gestures and rituals involved in dancing or musical performance, and as was inherent in John Blacking’s problematic ethnomusicological approach to studying the connection between music and human nature.

This, in a nutshell, captures the philosophical basis for the Minimalist Program, and for generative linguistics too, speaking more broadly. So, the basic premise of the MP is that a study of human language (i.e. I-language) should focus first and foremost on the structure of the linguistic mind, i.e. on the form (and not function) of C_{HL} , because it is this system that makes language uniquely human,

given its species-specific, recursive, computational properties. This, in turn, endows language (and the mind in general) with an infinite creativity that has no obvious neural or bodily locus, or evolutionary or sociocultural function.

There is an important historical basis for the above proposal in the Rationalist philosophy of Rene Descartes, who, after all, famously rejected a material approach to describing the mind, given his observation that it is the unique, infinite nature of the mind that enables our infinitely creative use of language, and which also separates us from both non-human animals, and finite machines controlled by mechanical principles. It is for this reason that Noam Chomsky has described the field of generative linguistics as really being that of a “Cartesian linguistics” (Chomsky (1966)). This connection can be seen especially in Descartes’ belief that only humans can express their thoughts by arranging words to form any new utterance they want – an idea we explored in the last section – whereas inanimate machines and non-human animals can only ‘parrot back’ what they have heard (like “Polly wants a cracker”):

“If any such machines bore a resemblance to our bodies and imitated our actions as closely as possible for all practical purposes, we should still have two very certain means of recognizing that they were not real men. The first is that they could never use words, or put together other signs, as we do in order to declare our thoughts to others. For we can certainly conceive of a machine so constructed that it utters words, and even utters word which correspond to bodily actions causing a change in its organs ... But it is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dullest of men can do. Secondly, even though such machines might do some things as well as we do them, or perhaps even better, they would inevitably fail in others, which would reveal that they were acting not through understanding but only from the disposition of their organs ... Now in just these two ways we can also know the difference between man and beast. For it is quite remarkable that there are no men so dull-witted and stupid ... that they are incapable of arranging various words together and forming an utterance from them in order to make their thoughts understood; whereas there is no other animal, however perfect and well-endowed it may be, that can do the like.” (Descartes (1988): 44-45)

An interest in explaining this uniquely human attribute that is the mind is what led to a renewed interest in its inner workings in the middle of the 20th century, in what would become the “cognitive revolution” (for a succinct, historical perspective on this, see Miller (2003)). Importantly, a critical aspect of this revolution, as we have discussed in some depth now, was the realization that the human mind has a computational structure, and that any attempt to describe and/or explain the mind would have to deal

specifically with this phenomenon. So, when Alan Turing took on Descartes' challenge of inventing machines that would be so 'humanly' intelligent that people would not recognize "that they were not real men" (i.e. when they take the so-called "Turing test"), he argued that such machines should simulate the way humans *compute information* while thinking.

For this reason, a computational perspective towards language has been the focus of generative linguistics since its earliest days too. So, instead of treating the linguistic mind as some sort of "black box", which just records and repeats (or 'parrots back') linguistic stimuli it has heard, as was inherent in Behaviorist descriptions of language until the 1950s, generative theory reconceived the linguistic mind as a computational system – an *intelligent* machine – that takes an active role in processing linguistic information according to psychological principles of some kind. This was expressed most famously in Noam Chomsky's critique of Behaviorism (in Chomsky (1959)), where he demonstrated that the Behaviorist attempt to characterize language as a set of learned responses to heard stimuli is unable to account for the creative use of language by people. Instead, Chomsky argued, an active, structured computational system – i.e. a *mind* – is required for an adequate description and explanation of human linguistic behavior.

In light of the above Cartesian roots, one might get the impression that generative linguistics is committed to a metaphysical split between mind and body. That is, its rejection of material approaches to the study of language, while simultaneously focusing on the internal workings of the linguistic mind, might suggest that generative linguists believe in the independent reality of mind and matter. Moreover, this might also suggest that the linguistic mind requires a different kind of study altogether than typically occurs when studying material phenomena – in other words, its Cartesian roots imply a *methodological dualism* between linguistics and the other, natural, sciences.

But generative linguists deny this. In fact, one of the novel features of the Minimalist Program (which also marks a change from earlier generative theory) is its insistence that linguistics is a natural science, no different in its methods and aims relative to any of the other natural sciences. That is,

Minimalist linguists subscribe to both a metaphysical *and* a methodological *naturalism* in their attitude towards the study of language (see Hinzen (2006): 33-65 for a more detailed discussion of this). As I have mentioned earlier, Minimalists see language as a biological system, and the study of this system a sub-discipline within biology they call “biolinguistics”. So, Minimalists do not see linguistics as being a unique paradigm, separate from the other, especially biological, sciences. It is just that the focus of this biolinguistic science is different from other extant biological approaches to language, notably in its formalist orientation.

The fact that Minimalists consider their research program to be just a kind of natural science has to do partly with the belief that we have countenanced since the beginning of this chapter, which is that language (and *ex hypothesi* music) is a part of human nature – and humans are biological entities, with species-specific traits (language and music being examples of such traits, according to John Blacking too). So the way we go about understanding language or music should not be different from the way we would study other parts of human biology. This, however, does not imply a return to a study of language or music in terms of their biological (and specifically, evolutionary) *functions*. The proper study of the human language faculty should be a study of its biological form, especially its *computational* form as I said earlier – but this itself can follow the methods and aims of a natural science.

In fact, in the history of the natural sciences, there have been certain paradigms that have been more interested in thinking about natural entities in terms of general principles of structure and form, rather than in terms of the function of these entities (for example, in physical rather than evolutionary descriptions of various biological phenomena) – and it is this sort of inquiry that the MP wishes linguistics to be.⁴³

⁴³ In this context, it is worth discussing Isaac Newton’s famous critique of Descartes’ ideas, which rejected Descartes’ mechanistic treatment of matter, given Newton’s description of matter in terms of his theory of forces and action at a distance. Importantly though, Newton’s critique did not extend to Descartes’ ideas about the mind. So, Descartes’ assertion of the mind, and Newton’s rejection of a purely mechanistic explanation for matter, essentially amount to the same thing – a rejection of the idea that important aspects of nature can be described in simplistic ‘cause and effect’ terms. (This actually made Newton very uneasy since he could not assign a cause to the

It might be useful in this regard to consider the philosopher Wolfram Hinzen's discussion of the three ways of thinking about the relationship between a biological phenomenon (i.e. an organism) and the form of that phenomenon, as proposed by the biologist George Williams (Hinzen (2006): 11-13). The first way is to think of an organism as a *document*. This is useful when one is interested in the role of history in building the organism's form, since organisms are records of the various historical events (such as gene mutation) in the past that resulted in their current form. The second way, in contrast, views an organism as an *artifact*, whose form serves certain functions, and is therefore only comprehensible when one understands what those functions are. This is analogous to the way the design of a machine (for example, a high speed train engine) only makes sense in the context of the function that design serves (e.g. reducing wind resistance, which enables better performance at high speeds). Finally, the third way of thinking about the relationship between an organism and its form views organisms as *crystals*, i.e. objects whose form is determined less by history or function, but by various laws and constraints on structure building instead. This forces an object to have only a few of the many possible forms it could have, so as not to violate any (often physical) laws of nature. (An example of this is the fact that organisms are restricted to certain sizes so that they do not collapse under their own weight.)⁴⁴

The first two ways of thinking about the organism/form relationship above, i.e. organism-as-document and organism-as-artifact, are quite common in the biological sciences, since they involve thinking about biological systems in terms of their evolutionary history and function respectively, which have both been popular avenues for research. But given the importance of understanding its

force of gravity, leading his critics to even describe his theory of forces as a return to the dark ages of Scholasticism.) The moral of the story being that the study of the mind, and mental phenomena like language, is not significantly different from the study of matter – they are both studies of some aspect of nature that cannot be adequately described in mechanistic terms. Whether we describe this as a study of “mind” or of “matter” is to some degree arbitrary.

⁴⁴ This is why paleontologists have long puzzled over how huge dinosaurs like *Diplodocus* could live on land (as the fossil record confirms), because such massive organisms would normally collapse under their own weight unless they happened to live in water-bodies such as lakes, where the buoyancy of the water would have given them some support. One recent proposal on this matter suggests that some large dinosaurs (like *Tyrannosaurus rex*) maintained their massive form by essentially shifting their weight back and forth between a large head and a large tail, like a giant see-saw – the downside to this being that their legs ended up being quite weak and light instead, which prevented them from running fast (Hutchinson and Garcia (2002)). Another proposal suggests that dinosaurs like the *Titanosaurus* had lighter, air-filled bones, and also special biomechanical features to help them cope with their weight (Wilson and Carrano (1999)).

computational structure – something that has no obvious evolutionary history or function, as we saw in the last section – these two approaches are less useful in a biological study of language. For this reason, Minimalists consider language as being analogous to a crystal – i.e. a natural phenomenon that is governed more by general scientific principles of form than by evolutionary function. So, it is the third perspective on the relationship between an organism and its form that the MP subscribes to, in its study of the biological ‘organ’ that is language. As Hinzen says:

“All three perspectives are equally legitimate when looking at a natural object, such as a human being. Humans are a mixture of history, artifactuality, and law. But whichever perspective leads to the most fruitful results in some particular case will depend on the object being studied, and the particular trait under consideration. It is the third perspective that [the Minimalist Program] emphasizes, for the case of human language. Although studying the *nature of humans* [my emphasis] under this perspective will necessarily entail studying *functions* they carry out, it will not involve viewing, as a matter of methodological principle, organic design as intrinsically *functional*, or as serving a purpose. This matters, as we shall see, because surviving conceptualizations of human nature (say, in evolutionary psychology) by and large depict it as intrinsically functionally designed. The human linguistic mind, in particular, is thought to be the result of external shaping of selective forces, acting on the communicative functions and other effects of subcomponents of the language system as a whole, eventually composing it, piece by piece, in a gradualistic fashion ... I note here that the intuitive notion of human nature as such does not invite a functionalist perspective in the sense of the second perspective above, despite the predominance of the latter in current revitalizations of the notion of human nature (e.g., Pinker (2002)). The point is that, intuitively, the *nature* of a thing is what it *is*, irrespective of what *happens* to it, or how it is *used*.” (Hinzen (2006): 12-13)

The above also points to another crucial aspect of the object-as-crystal approach within the natural sciences, which as we now see is the approach taken by the MP too. This involves the notion of *internalism* – which Noam Chomsky describes, quite simply, as a search to “understand the internal states of an organism” (Chomsky (2000): 134). The object-as-document or object-as-artifact approaches are not particularly inclined towards internalism, since in these approaches an organism’s external environment matters, often more than its internal states, since its history takes place in, and is often determined by, this environment (consider the much-discussed effect of ice ages or asteroids in the evolutionary history of dinosaurs in this regard), and the functions an organism’s design serves are a response to this environment too (for example, an insect’s wings will need to serve a cooling function only if the environment is hot). In contrast, when viewed as a crystal, only the organism’s internal structure, and how this structure obeys or violates natural laws, is of any particular interest – again, as Hinzen says,

irrespective of what happens to it, or how it is used. Of course, even something like a crystal is governed by external environmental factors – for example, a snowflake (possibly the prime example of a crystal) cannot form if the environment is too warm or dry. But, to quote Hinzen again:

“...to say that these external factors explain the internal structure would be as strange as saying that the water we give to a plant causes it to develop in the way it does. Just as the plant’s development will be a matter of what species of plant it is, and of internally directed principles of growth for which external factors are merely necessary conditions, the crystal is primarily explained internalistically by laws of form.” (Hinzen (2006): 12)

So, in sum, the Minimalist Program takes human language to be a natural system, specifically a biological organ, just like the heart, liver, kidneys etc., and it takes the study of this system to be a natural science, no different in kind from the other natural sciences – with the caveat that this science is specifically internalistic in approach, in order to account for the internal (computational) form of language, for which there is no obvious externalist (i.e. functionalist) explanation.

To put this in terms of our earlier discussion of John Blacking’s ideas about music and language, we could say that the MP treats language as a species-specific trait of humans – and thus an aspect of human nature – which is genetically-inherited, but which also has a specific (computational) form because of this genetic inheritance, similar to the way any biological organ has a specific form that is determined by its genes. Due to this genetic inheritance, the language organ knows how to grow and what form to take without requiring much external influence from its environment, just as a cell ‘knows’ how to become, say, a heart (or part thereof) because of its genes. Since it is innate and genetically-specified, this knowledge is of a different kind than the kinds of knowledge that have to be learned from the environment (such as cultural rules of etiquette), and it is this knowledge that is really our knowledge of grammar. I had defined “grammar” earlier as “the formal, psychological system that lies behind human linguistic ability, specifically the ability to compute linguistic information”. We can now fine-tune this definition to say that grammar is the system of psychological (or biological, assuming methodological naturalism) principles, which give rise to the form of the computational system of human language, i.e. C_{HL} – and knowledge of grammar is therefore the unconscious, unlearned knowledge we have of this

system, and which is coded into our genes as it is for any human organ. Finally, our language organ's genetically-specified form endows it with certain functions too – it gives us the ability to express thoughts and communicate with each other. But these functions do not determine the form of the language organ any more than the blood-pumping function of the heart determines the specific form of the heart. (Of course the fact that the heart pumps blood necessitates that it have certain features, such as hollow chambers for blood to flow through. But again, these are “merely necessary conditions” as Hinzen says, which are not *sufficient* to tell us why the heart has the *specific* form it does.)

The brief description I have given so far already shows us what a unique perspective the MP brings to our discussion of language (and music too, as we shall soon see) as an aspect of human nature, and a species-specific trait of humans. Much of this uniqueness has to do with Minimalism's uniquely internalist focus on the form of language, instead of on its functions as so many of the other approaches discussed so far do. Part of this focus comes from its Cartesian roots, as I discussed earlier, and part of it also has to do with its more general interest in the form of organic entities – which betrays the influence of the Romantic Rationalist tradition of thinkers like Goethe, Kant, and Wilhelm von Humboldt, who were also interested in this very notion, and were consequently an important influence on Minimalism as well. I will explore this historical link more in the next section.⁴⁵

But there is an important issue that arises out of all of this. That is, if the study of language should really be an internalist study of the form of the language organ, and of general principles of structure and growth, what are these principles? More specifically, are these just the principles of form and growth of the kind we see in biochemical or molecular biological descriptions of cells and genes (which would make linguistics a form of (bio-) chemistry), or of the kind we see in descriptions of the general physical

⁴⁵ Another name particularly worth noting in this regard is that of the famed British naturalist Alfred Russell Wallace, who co-proposed the theory of evolution through natural selection with Charles Darwin. Unlike Darwin, whose views on music and language we explored in the last section – and who was more ambiguous about whether he thought they evolved through natural selection or not – Wallace was quite unequivocal in saying that these faculties “cannot be accounted for in terms of variation and natural selection alone, but [require] “some other influence, law, or agency,” some principle of nature alongside gravitation, cohesion, and other forces without which the material universe could not exist” (Chomsky (2005): 3).

constraints on form and growth, as exemplified by the constraint that organisms not be too large so as not to collapse under their own weight – and which would make linguistics a branch of (bio-) physics?⁴⁶

To understand this problem more clearly, let us extend our earlier discussion of crystals to a biological phenomenon, such as a flower – specifically a sunflower. The face of a sunflower has a very intricate, crystal-like organization, similar to a snowflake. In fact, the little florets that cover a sunflower's face are arranged into specific patterns, which is a phenomenon called “phyllotaxis”.

The phyllotaxis of sunflowers, and other patterns in nature, were of immense interest to the early 20th century Scottish biologist D'Arcy Wentworth Thompson. Thompson pioneered the mathematical study of biological forms, and the explanation of how patterns are formed in plants and animals (a process called *morphogenesis*), as a result of which his ideas are cited frequently in the Minimalist literature, and were indeed an influence on a variety of thinkers interested in issues of form and design such as Alan Turing (e.g. in Turing (1952)), the anthropologist Claude Lévi-Strauss, and even artists like Henry Moore and Jackson Pollock. In his *magnum opus* “On Growth and Form”, Thompson described how many organic forms could be reduced to basic geometrical patterns, which also illustrates his belief in the organism-as-crystal metaphor. In particular, Thompson discussed how the phyllotaxis of a sunflower's face is that of several spirals, which correspond, quite amazingly, to successive numbers in the Fibonacci sequence (Thompson (1917): 635-651). (The Fibonacci sequence is made up of the numbers 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55 and so on, where each member of the sequence is the sum of the previous two members. So, if you count the spirals that seem to curve in a clockwise way, and then the ones that seem to curve in an anticlockwise way, you will always end up with a pair of numbers that correspond to successive numbers in the above sequence, such as 21 and 34.) Similar patterns can be found in the tail feathers of a peacock, or in the shell of a mollusk called the nautilus.

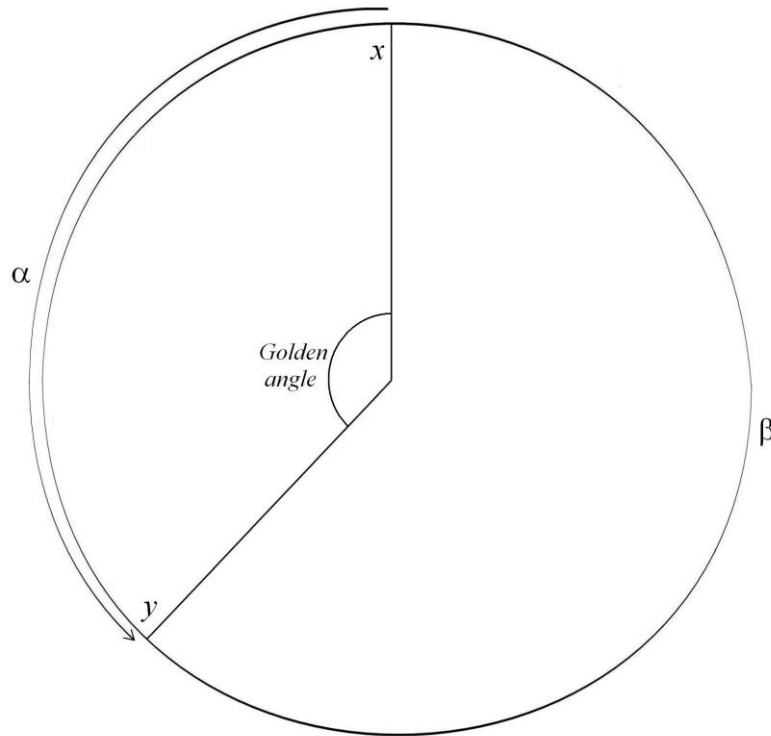
⁴⁶ The biologist Niels Jerne famously raised this possibility when he said, at the end of his acceptance speech for the 1984 Nobel Prize in Medicine (titled “The Generative Grammar of the Immune System”), that “this hypothesis of an inheritable capability to learn any language means that it must somehow be encoded in the DNA of our chromosomes. Should this hypothesis one day be verified, then linguistics would become a branch of biology.”

Sunflowers, peacocks and mollusks are not closely related in the evolutionary tree, and live in very different environments – so the similarity of the patterns in their forms cannot be the result of a shared evolutionary history or adaptive function. Perhaps unsurprisingly, this similarity therefore has more to do with general constraints on form and growth, i.e. with the organism-as-crystal view of biological forms. When a sunflower grows, the little florets in its face move outwards from a hypothetical central pole, where they are replaced by new florets – and as the older florets move outwards, the face of the sunflower gets larger and larger, which of course is how we can tell that the sunflower is growing. Now, if growth in a sunflower just involved the older florets moving out from the center, they could just move out in a straight line from the center, getting farther and farther away from the center as the flower grows older. But then the sunflower would end up growing in the form of a straight line, instead of having the round shape we associate with flower faces. So, to create this round shape, the florets have to move outward from the center pole *and also* curve, in order to create the circular shape of the sunflower’s face.⁴⁷ But moving in a curving fashion rather than in a straight line increases the chances of your ending up where you started too – at which point a floret will bump into the new floret that took its place after it moved, and this will impede growth. So, a floret has to move outwards from a central pole, in a curving fashion (to maintain the structural integrity of the flower), while simultaneously moving *away* from all other florets in the sunflower’s face (so as not to impede growth).

It turns out that the optimal angle that a floret needs to curve by, in order to meet the two above conditions, is the famous “golden angle”, which is approximately 137.5°. This is shown in Example 1.1-4.

⁴⁷ More accurately, the florets have to move in a curving fashion so that the flower can have a shape to begin with, or else it will just grow endlessly in one direction and fall apart, like an unraveled ball of string. There is no reason for why the flower needs to have a specifically *round* shape, however aesthetically pleasing that might be to us. Of course one could argue that having a round shape maximizes the surface area of the flower, thus increasing sunlight exposure, photosynthesis – and ultimately survival. This would be an organism-as-artifact type argument, since it explains the form of the flower in terms of the function this form serves, i.e. of increasing exposure to sunlight. But this begs the question of how the flower got its round shape to begin with – i.e. for it to realize the adaptive value of a round face, it has to grow a round face first, and one with a number of spirals that corresponds to a number from the Fibonacci sequence. So, the question is how the flower arrived at *that* for starters – the answer being that *formal* principles of growth took it there, or more specifically, made it unavoidable for it to have any other shape.

Example 1.1-4. Phyllotaxis in a sunflower



As the example also shows us, the floret will need to travel a certain distance from point x to point y , to realize the curve that corresponds to the golden angle – i.e. it will need to travel along an arc of length α . Now if the floret kept on curving it would end up where it started, so the distance α traveled by the floret is just a fraction of the total circular (or more accurately, *angular*) distance it could have traveled, with the remaining distance along this circle being represented by β . The ratio of distance β to distance α , i.e. β/α , is a famous mathematical quantity, related to the golden angle, known variously as the “golden ratio”, the “divine proportion”, or the Latin “*sectio aurea*” (i.e. the “golden section”), whose value is approximately 1.618. The golden ratio is an irrational number though, which is why 1.618 is just an approximation, and which is why the floret can never travel by the exact or ideal α needed to realize the actual golden ratio. (So, it cannot curve by the actual golden angle either, which is why it is only approximating the golden angle if it curves by 137.5° .)

But sunflowers are real and not ideal, as are their florets. So the best thing for the floret to do, to optimize its growth, is for it to move by a *real* number that approximates the α of the golden ratio, or curve by a *real* angle like 137.5° that approximates the golden angle. It so happens that the numbers of the Fibonacci sequence are examples of such real numbers – the ratio of a number in the sequence to the number preceding it in the sequence (such as 21:13, or 34:21) provide real-world approximations of the golden ratio β/α . This is why when a sunflower grows, its florets move outwards from the center in curving paths – which create the sunflower’s spirals – whose numbers correspond to successive numbers in the Fibonacci sequence, since this is the optimal way to derive a sunflower’s form during growth.

Speaking more generally, the above is what leads to the *discrete units of structure* we see in biological phenomena, as exemplified by the spirals of a sunflower. So, nature seems to bestow specific forms on organisms during growth, since these forms optimally solve certain problems of growth, in the way sunflower spirals – with their Fibonacci-based patterns – solve the problem of how a sunflower’s florets can be optimally arranged on its face.

Returning to language now, we know that language also exhibits discrete units of structure, viz. words and phrases (i.e. groups of words).⁴⁸ (In fact, this is one of the important formal similarities between

⁴⁸ In the case of sunflower spirals, these discrete units seem to be the result of an ‘approximating’ process, which takes the ideal, abstract form given by the golden ratio, and finds a real-world approximation of it in the Fibonacci sequence. This then gives rise to the sunflower’s real-world form, in which its spirals correspond in number to numbers from the Fibonacci sequence. So, there seems to be a meta-algorithm in nature that takes an ideal form-generating algorithm based on the golden ratio and transforms it into a real-world form-generating algorithm based on the Fibonacci sequence – which then gives rise to the form of real-world phenomena like sunflower spirals, peacock tail feathers, and nautilus shells. No one knows what this meta-algorithm is, or that it even exists – in fact, it could be the case that discrete units of growth are formed arbitrarily from the ideal form-generating algorithm. But the important thing is that discrete units do emerge in the process of growth that seem to have clear algorithmic properties, viz. of belonging to the Fibonacci series – and this is true of language too. To deny this, just because we have not figured out how these units arise, would amount to the mistake physicists actually made when they denied the discrete nature of atoms because they had not yet figured out how these discrete entities arise from continuous physical energy. This mistake went uncorrected for many years until Linus Pauling described the structure of the atom in physical terms in his Nobel Prize-winning work of the 1930s, unifying chemistry and physics in the process. The moral of the story is this: how the computational structure of language, with its property of discrete infinity, arises within human biology is not yet known (assuming that extant functionalist explanations of its origins are inadequate, as we have discussed in some depth now). However, this should not amount to a rejection of the study of language in computational terms, i.e. in terms of its form rather than its function (for more on this, see Uriagereka (1998): 69-76).

language and music, since musical units, such as the pitches of a scale, and groups of vertically-aligned pitches (i.e. chords) and linearly-aligned pitches (i.e. melodic/motivic structures) are all discrete too.) Recall that the discreteness of linguistic and musical structures plays a role in distinguishing human language and music, on the one hand, from non-human ‘vocal music’ (to use Darwin’s term) on the other (as we saw, for example, in Steven Brown’s description of “the unpitched hoots of chimpanzees or the non-discrete vocal glides of gibbons”). Moreover, our ability to combine the discrete units of language and music into more complex, hierarchically- and recursively-organized, structures (i.e. phrases) is what endows the human computational systems of language and music with their infinite creativity too.

Finally, there has been much research over the years that suggests, albeit controversially, that language and music show patterns of structure that can be understood in terms of the Fibonacci sequence as well. For example, the mathematician Marcia Birken and the poet Anne Coon argue that Fibonacci sequences play an important role in poetry, for example in limericks, which are 5-line poems with 13 stressed syllables, these syllables being arranged into groups of either 2 or 3 stressed syllables per line (Birken and Coon (2008): 59-61). In the case of music, the music theorist Ernő Lendvai’s description of Fibonacci patterns in the music of Béla Bartók (in Lendvai (1971)) is well known too, though this description has been debated by László Somfai (in Somfai (1996)), and the extent to which such patterns appear in the music of other composers, or in the music of other idioms is a highly contentious issue (see Livio (2008): 183-194) for a brief review).

So language (and music) have a computational form that makes use of discrete units of structure, in infinitely creative ways. And this form might be governed by general principles of formal optimization that have something to do with Fibonacci patterns, which would make the structure of language optimal in the way sunflowers and peacock tail feathers seem to be. Does this imply, then, that explanations for linguistic structure should be sought in the structure of flowers or other biological systems that reveal complex, Fibonacci patterns – or, as mentioned earlier, in general laws of chemistry or physics?

The answer to this has to be “no”, because linguistic structure seems to have two crucial properties in addition to its discreteness and infinite creativity, viz. (1) *underspecification*, and (2) *structural economy* – and no other biological or physical system seems to have all of these properties simultaneously (with the striking exception of music, as continues to be my contention in this dissertation). This is why general principles of biological growth, or general physical/chemical laws, cannot provide an adequate description or explanation of the structure of language (or music) – even though they come closer to doing so than the several approaches we have explored so far, such as ethnomusicology/anthropology, cognitive psychology, and neuroscience, which usually fail to countenance the peculiar computational structure of language and music to begin with, or attempt to explain away this structure in functional terms.

The underspecification of linguistic structure is what allows for the important phenomenon of linguistic variation. We know that one of the salient features of language is that it varies among different groups of people, which leads to the incredible diversity of languages around the world. The fact that language varies like this has often led to the rejection of its universality, and to the belief that it must be learned from a specific environment or culture – which has often also led to explanations for language in terms of the functions linguistic structure serves in those environments. We have seen how this is commonly the case with music too, especially in ethnomusicological theses about musical structure.

But it should hopefully be clear now that the environment- or culture-specific aspects of linguistic structure, including its functions, have more to do with the *external* aspects of language, also known as E-language. Language in its *internal* aspect, i.e. I-language – which is also the focus of the *internalist*, generative study of language – is no more dependent on its external functions than the structure of the heart is dependent on its blood-pumping function. Which is why it can be genetically-specified and possessed by all members of the species (i.e. it can be universal), as seems to be the case for the internal (computational) form of language, and as seems to be the case for other biological forms such as the phyllotactic form of a sunflower. In my review of some of Harold Powers’ thoughts about music and language earlier in this chapter, I discussed how language’s internal structure can generate a great

diversity of pronounceable surface forms, which I argued, in the case of music, to be the basis for the difference between monophonic and polyphonic musical surface structures too – and which gives both language and music the appearance of being idiom-specific rather than universal. But despite this universality, the variation seen in both music and language across idioms has to be accounted for. That is, we need to be able to describe how a universal, internal, C_{HL} or C_{HM} gives rise to a variety of surface forms. In the case of language, we have seen that part of this description has to do with the transformational nature of the computational system, in which diverse surface forms are generated by the system through transformational phenomena such as wh-movement.

The other part of this description involves the notion of grammatical “parameters”. A parameter is a binary variable that provides two, and only two, options for how a universal grammatical principle must be realized. To understand this, consider the universal grammatical principle called the *Extended Projection Principle*, which basically requires that all linguistic surface structures (i.e. sentences, or specifically *tensed clauses*, as in the structure described in (1a) several pages ago, “Jürgen read a book”) must have a subject, like “Jürgen” (Carnie (2002): 175). So, when generating surface structures, C_{HL} must generate only tensed clauses that have a subject, since a tensed clause without a subject is ungrammatical in all languages. However, this subject *may* or *may not* be explicitly pronounced by a speaker, depending on the language – subjects must be overtly pronounced in English tensed clauses (e.g. see whether (1a) remains grammatical if “Jürgen” is not pronounced), but they do not have to be in a language like Italian. Moreover, the subject of an English clause normally comes *before* the main verb (compare “Jürgen” and “read” in (1a)),⁴⁹ but it often comes *after* the main verb in a language like Irish. (See Baker (2001): 35-62 and Carnie (2002): 189, for more on these phenomena, and also my footnote 2 above.)

So, we see how a universal principle of grammar, like the one that requires all tensed clauses to have a subject, can be realized in actual linguistic surface structures in one of two ways – i.e. the subject

⁴⁹ In the passive form of (1a), i.e. “The book was read by Jürgen”, “Jürgen” comes *after* “read”. However, the passive form of a sentence is also the result of a transformation like wh-movement, called “noun phrase movement” – which is why it has an unusual word order, just like the interrogative sentence that results from wh-movement. Also, notice that “Jürgen” is no longer a noun phrase (and hence the subject) in the passive form of (1a), but is part of the prepositional phrase “by Jürgen” instead.

must *either* be pronounced *or* unpronounced, or it must *either* come before the main verb of the clause *or* after it. These two options are the two ‘settings’ of a grammatical principle – i.e. the two values of a binary *parameter* that a principle can be ‘fixed’ with, when C_{HL} operates in a given language. So, in the above example we see two parameters of the Extended Projection Principle in action, one having to do with whether the subject required by the principle is pronounced or not, the other being a word order parameter of where the subject appears in the surface structure of a clause. This is why we can think of language as having grammatical *principles* that are *parameterized*. And this is how languages vary too – specific languages are specific parametric settings of certain universal principles of grammar.

Over the last thirty years, the study of generative grammar has primarily been concerned with what the universal principles of grammar are, and how they are parameterized – a framework known as the theory of “Principles and Parameters” or “P&P”. In fact, this is also the theoretical framework within which the Minimalist Program operates. The MP does not have a theory of its own – it is really just a research *program* (hence its name) that has been exploring how the tenets of P&P theory can be understood, and refined, as an internalistic, natural science of language. Which is why there has been so much focus by Minimalists on what kind of study the generative study of language is, given its philosophical basis in methodological and metaphysical naturalism, and this is why Minimalists stress the differences between generative linguistics on the one hand, and the neuroscience or anthropology of language, or any functional approach to language on the other – as we have explored in some depth now.

At this point I can also let you into my dirty little secret – given the connection between P&P theory and the Minimalist Program, this dissertation could easily have been called “Generative Musical Grammar: A P&P approach” too, especially since most of my music-theoretic speculations will borrow from P&P’s technical toolkit. But since I do put some effort into exploring how Minimalist advances within P&P theory might help illuminate certain music-theoretic matters – especially in the entire second half of the dissertation – the title of dissertation ultimately seems apt.

In the next chapter (specifically section 1.2.3), I will get into the technical nitty-gritty of P&P theory, and its Minimalist extension, in greater detail. All that I hope to point out for now is that the P&P approach shows us how linguistic variation can arise from linguistic structure, and also why this makes this structure *underspecified*. This is because our genes do not endow us with a specific parametric specification for C_{HL} . C_{HL} comes underspecified, and is therefore capable of having its parameters set to any natural language, which is how little children are able to acquire their native languages so easily, and even at the young age at which they have not acquired other, more culture-specific forms of knowledge (such as those regarding etiquette). In fact, as the linguist Juan Uriagereka says, someone's native language can be understood as a complete specification of his or her I-language's parametric options – which means that how languages vary across different environments or cultures therefore has less to do with their functions, and more to do with how C_{HL} is parametrically fixed during language acquisition (Uriagereka (1998): 36).

Now if acquiring a language involves fixing the parameters of C_{HL} to that language's parametric settings, then this can only happen through exposure to that language – which is why children can acquire the language of their immediate environment, such as the one spoken at home, so easily.⁵⁰ But the messiness of the linguistic data one normally receives from this environment will make it hard for C_{HL} to figure out what its parametric settings are, unless C_{HL} is very efficiently designed. In fact, sometimes this data is so impoverished that C_{HL} is actually unable to latch on to its parametric settings, so that the language one acquires ends up being a new language altogether. The best examples of this are creoles, which arise from the severely impoverished data of pidgin languages (Uriagereka (1998): 39-42). Pidgins are essentially makeshift languages, which often occur in sociologically tragic conditions (such as among workers forced into slavery on plantations, who do not have a shared language but still need some means

⁵⁰ This does not amount to saying that language is *learned* from one's environment though, in the Behaviorist sense discussed earlier, since even a genetically-inherited biological system often requires exposure to stimulus of some kind for its growth to be triggered. A good example is that of the visual system, which is genetically-determined (since we do not have to be taught how to see), but which still requires exposure to light in the early stages of development, without which it will not develop properly – as David Hubel and Torsten Wiesel demonstrated in their Nobel Prize-winning work of the 1960s (Wiesel and Hubel (1963), Hubel and Wiesel (1970)).

to communicate). So they do not normally display all the standard features of natural language, such as recursion. But the languages acquired by the children of pidgin speakers, i.e. creole languages, *do* display all the properties of a natural language, which means that these children were able to acquire a natural language despite the impoverished data of the pidgins they were exposed to. That is, their computational systems were able to fix certain parametric settings – but just not ones belonging to specific I-languages. So, they end up being new, creole, languages, with some basic, general, parametric settings shared by many I-languages – settings such as subjects must always be pronounced, and must always precede verbs.

In light of the above, we can see that the underspecification of C_{HL} is actually just an aspect of the other crucial property of language mentioned above, viz. its structural economy. Only if the design of C_{HL} is economical can it latch on to whatever parametric setting is remotely available from the data to which it is exposed. To put this in another way, the structure of C_{HL} has to be streamlined, so that it is sensitive only to information that is relevant to language acquisition, especially when the data to which it is exposed is messy or impoverished, as happens in pidgins.

But C_{HL} must have an economical design for other reasons too. Juan Uriagereka discusses this in terms of two forms of economy, viz. *representational economy* and *computational economy* (Uriagereka (1998): 78-85). To understand the first kind of economy, consider the undeniable fact that language is used by people, to communicate with each other. For it to be usable, C_{HL} has to be able to generate surface structures that, at the bare minimum, express our thoughts and that can be perceived and pronounced (or gestured, in the case of sign language). That is, C_{HL} should be able to ‘map’ to the psychological systems involved in thought (i.e. the conceptual-intentional or CI system), and perception and pronunciation (i.e. the sensorimotor or SM system). (The linguistic aspects of the former are often studied under the broad heading of “semantics”, and the latter “phonology”.) To repeat a point that has been made before, it is not the case that the workings of C_{HL} are determined by the semantic or phonological functions of language, just because C_{HL} has to produce usable structures – just as the structure of the heart is not determined by the structure and function of other bodily systems like the lungs. But the heart still needs to interface with the lungs in order for the blood it pumps to be

oxygenated, the absence of which leads not only to a breakdown of the two systems but also *death* – so the language organ, i.e. C_{HL} , needs to interface with, say, the vocal organs (which are part of the sensorimotor system) too, or else the linguistic abilities of our species will appear to be ‘dead’. (Since many non-human animals have semantic and phonological abilities similar to ours, as discussed in the last section, it could be that the reason no other species has human linguistic ability is because their semantic and phonological systems do not interface with whatever computational system for language they might have – which is why they appear not to have such a system, even if they do in reality.)

The basic point is that, given the contingent fact that language is used, the *representations* (i.e. structures, put crudely) generated by C_{HL} must map to other relevant systems – and this must happen efficiently too or the mapping from C_{HL} to these other systems will fail. (This is akin to how the pumping of blood by the heart will be inefficient if its chambers are leaky, which will result in some blood not being oxygenated by the lungs.) This is what is meant by “representational economy” – a kind of harmony between the representational interaction of the linguistic system with other mental systems, particularly the conceptual-intentional and sensorimotor systems.

Linguists have also found that the actual computations through which C_{HL} generates representations have their own internal economy. For example, when certain transformations, including wh-movement, take place they follow the shortest, most optimal path – akin to how phyllotaxis in sunflowers follows an optimal path brought about by a Fibonacci algorithm. (I will look at some of these, rather technical, phenomena, when we look at P&P theory and its Minimalist extension in more detail in the next chapter.) So, C_{HL} seems to have some sort of *computational economy* too.

In light of its underspecified and economical structure, language appears to be a rather different sort of natural system, than the ones normally studied in biology, chemistry or physics. As a point of comparison, traditional biological objects are often *overspecified* – for example, the machinery that

guides the growth of neuronal axons has many redundancies built into it (Travis (1994)).⁵¹ A few examples of (non-linguistic) underspecification do exist in nature, but they are rare. One example is the song structure of adult birds, which young birds have to acquire from an underspecified, but genetically inherited, initial state. (This is one of the reasons for why the connection between bird song and human language has been of interest to so many thinkers, including Charles Darwin.)

There is, however, a complex – and poorly understood – interaction between those aspects of a biological form that are specified in its genes, and those that result from general physical laws and constraints on form. For example, the size that an organism can grow to is partly governed by the physical law that an organism that is too big will collapse under its own weight – so in such a circumstance the organism’s size does not have to be coded into its genes. Another example can be found in the well-known behavior of leaping that is seen in fish, like salmon, who need to negotiate waterfalls and other such obstacles when swimming upstream to spawn. This behavior is presumably coded into a fish’s genes, since it is a species-specific behavior – but the behavior of falling back down *after* leaping is due to gravity, and so does not have to be genetically encoded. So, these two examples would be instances of genetic underspecification in biological phenomena. But nobody knows the truth of the matter here – there are many who believe that growth in size *is* partly controlled by genes, although recent studies also suggest that the human skull and pelvis never stop growing (Morris et al. (2007), Berger et al. (2011)). And it is possible that salmon have a gene that makes them fall back after leaping too, although it is hard to think of a reason for why this should be the case.

The point I am trying to make though is that if the above phenomena are genetically specified, this would be just the normal *overspecification* we see in the biological world, given the physical constraints that already specify these phenomena. Unless, of course, they are not genetically specified,

⁵¹ Language has redundancies too, in the form of structures that are possible within it, but have no obvious use. Examples of this are center-embedded structures, like “the cat the dog the cow the milkman milked kicked chased left” (Uriagereka (1998): 65), or structures that are grammatical but meaningless – and cannot therefore be mapped to the conceptual-intentional system, like the famous “colorless green ideas sleep furiously” (Chomsky (1957): 15). However, the fact that these structures are not usable is a feature of linguistic performance – the aspects of language that show underspecification (rather than redundancies) are precisely those that do *not* concern its use, viz. those involved in linguistic *competence*.

being specified by physical laws instead – or if genetic functioning can be reduced ultimately to physical laws too. In this light, we could reconceive biological systems (including language) as being the result of general physical constraints on form – and language itself as being more akin to a *physical*, rather than a biological, system.

The problem with this, though, is that physics only tells us that certain biological forms might be possible, but not what they *are*. Physics reduces the space of possibilities, say, to the allowable sizes an organism can have, so as not to violate the laws of physics – but it does not tell us what the *specific* form of an organism will be, whose size falls within acceptable limits. In order to specify this form, certain parameters have to be set, just like in language – a “mutation pool” of possibilities for gene expression, as Juan Uriagereka puts it, which determines the specific forms an organism can take when shaped by purely physical forces (Uriagereka (1998): 15-25). Such parameters seem to be responsible for the phenomenon of *convergent evolution*, in which many well-known forms appear again and again in nature, even in organisms that have no discernable evolutionary kinship, because it is assumed that these forms provide optimal solutions to certain problems of growth. We have already discussed an example of this in the context of the (Fibonacci-based) solution to how florets should be optimally distributed on a flower’s face, which leads to the similar spiral shapes we see in organisms as disparate as peacocks and nautilus mollusks – and we also see this in nature’s solution to the optical problem of vision, which leads to the striking similarity of the mammalian eye to the eyes of the octopus and the box jellyfish, despite there being no apparent evolutionary kinship between these organisms either. So the mutation pool that leads to sunflower spirals and octopus eyes could give a clue as to how physics might shape the specific form of these organisms.

However, scientists have barely started exploring such parameters on the physical specification of form – and there is some evidence that phenomena like convergent evolution might be the result of genes after all, i.e. apparently disparate organisms that show similar, convergent forms might actually originate

from a common ancestor (see Quiring et al. (1994) in this regard).⁵² So, we cannot reduce biology to physics just yet. Translated into linguistic terms, this means that any general physical constraints on form that might be able to account for C_{HL} cannot yet account for its *specific* form – i.e. the specific set of parameters it takes for a given I-language. This is why language is not a typical physical system, and linguistics cannot be a branch of (bio-) physics just yet either.⁵³

A similar issue arises when we consider the economical structure of language. It is no doubt true that the economical aspects of language are quite reminiscent of the structural economies we see in a variety of biological forms such as shells, sunflowers, peacock feathers etc., and in non-biological forms like snowflakes. But unlike language, where these economies are an aspect of the internal, *computational* workings of C_{HL} (for example, in movement transformations), it is not clear whether the structural economies we see in other systems are the result of a computational process in those systems too. They might just be aberrations, especially since many such systems often get their ‘computations’ wrong – for example, a sunflower’s spirals often do not correspond exactly to a Fibonacci number as the mathematics of phyllotaxis suggest they should (see footnote 47 in this regard). We know that language has a unique, underspecified and economical, structure, which is governed by some, poorly understood, combination of genes and/or physics. Given the internalist, Cartesian spirit of generative linguistics, we can call this governing entity, simply, “mind”. But unless we want to ascribe minds to sunflowers – or more bizarrely, to snowflakes – it would be rather strange to say that these systems also display economies of the *computational* variety, as language does.

⁵² To put this in slightly more technical terms, the “analogous” forms we see in convergent evolution, might just be “homologous” instead (cf. Boyden (1943)). In this regard, music and language should be homologous too, at least according to the musilanguage hypothesis. Interestingly, psychologist Steven Brown has described actual, structural homologies between music and language (Brown (2001b)).

⁵³ This point is particularly relevant to music, in regards to the form of C_{HM} , given the abiding interest (dating back to at least Pythagoras) in explaining music in physical terms. The relevant thought here being that any general physical properties that C_{HM} might have will not specify the forms of particular I-musics – they will not be able to account for the surface forms of musical expressions in specific musical idioms. So, instead of describing what counts as a grammatical musical expression in one idiom, and how this differs from a grammatical expression in another idiom, such physical theories of music will often reject the grammatical form of musical structures altogether, focusing at best on some vague mathematical properties these structures might have. For examples of this, see Rahn (1983) and Tymoczko (2006).

Which brings us back to the central idea of generative linguistics, which is that language is a *psychological* system, with a unique, hierarchical and recursive structure – and which therefore has to be understood in its own unique terms, rather than being reduced to traditional biology or physics – however adequate these fields of inquiry might be for explaining even the formal aspects of other natural systems, like sunflowers and snowflakes. (Unless of course there is another system that shares aspects of language’s structure – *ex hypothesi*, music – in which case the study of language can be a joint study of language and this other system, rather than a unique study of language on its own.)⁵⁴

Under the assumption of methodological naturalism though, language is still a natural system – and specific to human nature – and the study of this system, i.e. linguistics, is still a natural science, since it still seeks to describe and explain language in the way a natural science would go about describing and explaining any other natural phenomenon, i.e. in terms of general principles of form, especially formal economy. The only problem is that principles of form are often ignored in favor of functionalist explanations in much of contemporary natural science, and to the limited extent that they have been investigated they remain insufficient to explain language, given the latter’s unique properties of

⁵⁴ Much recent work in physics has wrestled with the issue of how natural systems are even possible in the universe, given the laws of thermodynamics. According to the second law of thermodynamics the total amount of disorder (technically, “entropy”) of a closed system should increase – meaning that the total amount of order in a closed system like the universe should increase, mainly by its ordered systems dissipating into disorder. The best example of this can be seen in the cooling and eventual demise of stars, like the sun, as they give off energy. In this light, the increase in order that we see when natural systems are formed should be a thermodynamic violation. This is not true for most physical systems, like the solar system, since these will eventually cool and disappear, as just suggested. This is ultimately not true of biological systems either, since without a source of energy (like food, which helps maintain thermodynamic equilibrium) biological forms also eventually ‘dissipate’, i.e. they die. Therefore, one might attempt to describe the emergence of linguistic form in thermodynamic terms too – particularly in terms of the thermodynamics of complex systems, which is a major focus of research in the subdiscipline of physics known as complexity theory. For example, grammatical complexity might be seen as akin to physical entropy in dynamic systems, which arises out of the initial, underspecified form of an I-language. This is similar to how dissipative physical states (known technically as dissipative “phase transitions”) often arise, by amplifying the smaller deviations of a conservative phase transition, and which is how the celestial bodies are classically believed to have arisen too. (An example of a conservative phase transition is an ice crystal, which occurs in a closed system of thermodynamic near-equilibrium, and is therefore similar to the basic form of language – if we accept the organism-as-crystal metaphor for linguistic form.) But the bigger overall problem with a complexity approach to language is that when a system is too highly dissipative it gets too complex and starts behaving in unpredictable ways – and can only be described probabilistically. So, again, such an approach cannot explain how the specific details of concrete organisms arose. In contrast, the core aspects of the linguistic computational system act in very simple, systematic ways – the linguistic environment might be complex and messy, but the internal aspects of language are not, which is why people never fail to acquire language. This suggests that language does not behave like a dissipative complex system, and is essentially internally-determined (i.e. innate). (For a more thorough discussion of these issues, see Uriagereka (1998): 20-25, 39-42).

underspecification and economy. This does not mean that the search to explain language in terms of such general principles cannot go on though, and this is precisely the agenda of the Minimalist Program.

Earlier in this chapter, I discussed the two levels of adequacy that generative linguistics requires of any formal study of language, viz. that it describe what sentences are grammatically acceptable in a language, according to the innate knowledge of grammar possessed by native speakers of that language, and that it explain why humans minds have, or can acquire, this knowledge. (These are known, respectively, as the levels of “descriptive” and “explanatory” adequacy.) But given the more naturalistic orientation of contemporary generative linguistics, and given that language is similar to, but also different from, other natural systems in so many important ways, the MP adds a third level of adequacy for formal studies of language, which is that they explain how linguistic structure is shaped by more general principles of form and economy, what Noam Chomsky calls “third factor” principles (Chomsky (2005): 6-11).

The attempt to explain reality in economical terms is of course one of the hallmarks of the scientific method anyway, given Occam’s razor – but it is of particular significance to linguistics because of the actual economies that seem to be present within linguistic structure, as discussed above. This is why the Minimalist Program is *minimalist* – as a natural science of an inherently economical system, it aims to give the most streamlined and elegant explanation of why language is the way it is. Chomsky describes this aim best himself, when he says:

“Throughout the modern history of generative grammar, the problem of determining the character of [the human faculty of language] has been approached “from top down”: How much must be attributed to UG [i.e. Universal Grammar] to account for language acquisition? The MP seeks to approach the problem “from bottom up”: How little can be attributed to UG while still accounting for the variety of I-languages attained, relying on third factor principles? The two approaches should, of course, converge, and should interact in the course of pursuing a common goal.” (Chomsky (2007): 4)

So, if the MP is a study of language’s internal form in terms of general principles of form and economy – but principles that cannot be accounted for in terms of current biological or physical descriptions of these issues – what specifically are these principles? One way to answer this question is to think of them in

terms of the three levels of adequacy required by the MP of any internal, formal study of language – i.e. we can think of these principles as being either descriptive, explanatory or “third factor” principles.

Descriptive principles are simply principles that govern whether a surface structure generated by C_{HL} is grammatical or not. So, an example of such a principle would be one that requires that all surface structures, or at least all tensed clauses, have a subject. A list of such principles will therefore give a descriptively adequate account of all I-languages. The theory of Principles and Parameters attempts to provide exactly such a list, which is what makes it a descriptively adequate theory of language too. Descriptive principles are language-specific though, since their job is only to account for what makes linguistic surface structures grammatical – i.e. descriptive principles are not examples of more general principles of form.

Explanatory principles, on the other hand, are principles that govern how C_{HL} generates structures in the first place, which eventually conform to descriptive principles of language – and it is open to debate whether this makes explanatory principles language-specific or not. Given their connection to explanatory adequacy, these principles also constitute, in essence, a person’s innate knowledge of grammar – i.e. it is knowledge of these principles that allow a person to produce and comprehend grammatical structures in their native I-language, after the parameters of that I-language have been set. In the early days of generative linguistics (specifically in Chomsky (1965)), such principles were conceived of as rules for constructing specific kinds of phrases, i.e. noun phrases, verb phrases etc., and were therefore called “phrase structure rules”. (Another set of rules, called transformation rules, were also developed to allow for transformations like wh-movement on the structures generated by phrase structure rules.) So to give a descriptively and explanatorily adequate account of every I-language, a separate set of phrase structure rules for each language had to be proposed. This made these explanatory principles language-specific, which violated the desire for economy inherent in generative theory. So, linguists have searched for simpler ways of accounting for phrase structure over the years, and they have discovered that the work done by all these sets of rules can be accomplished by *one* operation, called “Merge” (Chomsky (1995c): 226). Merge is simply a set-theoretic operation that takes two items (specifically, *lexical* items, i.e. words)

and ‘merges’ them into a set (i.e. a phrase) – and in this manner generates every kind of phrase previously accounted for by phrase structure rules, and across languages too.

Moreover, Merge can also account for the transformational operations performed previously by transformation rules on the “deep” structures generated by phrase structure rules. This means that the distinction between deep and surface structure, which is a distinction I have accepted in my discussion so far, is no longer relevant in the MP. So, Minimalists reject these notions and speak, much more generally, in terms of “D-structure” and “S-structure” when they want to refer to the work previously done by phrase structure and transformational rules. From now on, this is the practice I will adopt as well. (There is another reason for why the notions of “deep” and “surface” structure were replaced with D- and S-structure, which has to do with semantics. I discuss this in section 1.2.3 of the next chapter.)

The fact that Merge is essentially a set-theoretic operation means that it does not have to be specific to language either. The two items that Merge merges do not have to be linguistic items – they just happen to be linguistic (i.e. lexical) items when Merge needs to generate a linguistic structure – which means that, unlike phrase structure rules, Merge is not necessarily language-specific. This means, crucially, that an innate knowledge of Merge could be responsible for how humans generate *musical* phrases too, across musical idioms – a point that I will make much use of in the next chapter. I will also discuss the above aspects of phrase structure rules, versus Merge, more in my detailed exploration of P&P theory in the next chapter. So, all that is worth noting at this point is how clearly the Merge operation streamlines linguistic theory, and pushes it towards its goals of minimalism and explanatory adequacy – which it does by asserting just one explanatory principle (resulting in the singular Merge operation), rather than a large number of rules (which, in turn, explains why humans have knowledge of language).

The above also helps us get rid of the notion of a rule in grammatical theory, along with the attendant baggage of phrase structure and transformation rules. This is important, because the belief that the structure of C_{HL} is rule-based, with these rules being language-specific, also encouraged the belief that these rules could be learned from specific linguistic contexts, implying that human linguistic competence can be learned from one’s environment too (as opposed to its being innate). Generative linguists have

always argued against this belief, because of the impossibility of learning a language by generalizing a learned rule to novel instances. For example, reconsider sentence (1a), which can be turned into the question in (1b), as we have discussed, via wh-fronting:

(1a) Jürgen read a book.

(1b) What did Jürgen read?

On the basis of this, one might say that language can be learned by hearing examples such as (1a) and (1b), and then generalizing the rule inherent in them to new contexts. But consider the sentence “Angela was talking to a man who was reading a book” – applying a ‘wh-fronting rule’ to this leads to the ungrammatical structure “what was Angela talking to a man who was reading?”. This is because the correct form of the above question involves fronting the entire subordinate clause “a man who was reading”, to give “what was the man reading, who Angela was talking to?”. But this is not the correct form of this question according to the wh-fronting rule – which just demonstrates the impossibility of learning a language by generalizing rules. However, children can always generate the correct form of a question, despite this being ‘unlearnable’ from one’s environment (Crain and Nakayama (1986)), which suggests that they must be innate, as generative linguists have always claimed – but in light of the advances made by the MP, where there are no rules – innate or learned – to begin with, this whole issue becomes moot.

This finally brings us to the third factor principles that govern linguistic structure. By definition, these are general principles of form and economy, not meant to be language-specific. But they have to be able to govern language’s specific structure as well, which is why some of the general principles of form and economy we have looked at above (e.g. Fibonacci-based principles of formal optimization) are inadequate for language. This makes figuring out what third factor principles govern language a difficult problem – and this is the focus of much cutting-edge research in contemporary generative linguistics. But given that they are principles of form and economy, it should be clear that these principles, whatever they are,

should attribute as little to C_{HL} as possible, to paraphrase Chomsky's statement from a few pages ago. In other words, given the contingent fact that language is used, these principles should specify the optimally economic way for C_{HL} to create structures for a speaker/hearer to use.

Now for C_{HL} to create usable structures it has to meet certain conditions on the use of such structures, conditions that will necessarily be external to C_{HL} itself, given that the uses to which C_{HL} -generated structures are put have little to do with C_{HL} 's internal form. As discussed earlier, these conditions will presumably be certain external conditions imposed by the sensorimotor and conceptual-intentional systems, since these are systems implicated in the, phonological and semantic, uses of language. So, the conditions these systems impose on C_{HL} will be conditions on how speech sounds or signs are perceived and articulated, and on how linguistic messages are understood. One could even say that the internal properties of the language faculty, as governed by C_{HL} , exist in order to meet such external conditions – i.e. *grammar-internal properties exist only to meet grammar-external conditions on language* (Uriagereka (1998): 90-91). This would be the most economical specification of C_{HL} , because it reduces the form of C_{HL} to only what is necessary for it to create usable structures – attributing anything more to C_{HL} would be conceptually unnecessary, and in violation of third factor principles of language design.

If the internal properties of C_{HL} arise to meet external conditions imposed on it by the SM and CI systems, C_{HL} must have a level of representation each corresponding to these two systems where they are interpreted by these systems – in order to ensure that the language faculty has actually met the external conditions imposed by these systems. To meet this requirement, the MP proposes a level of representation called “Phonetic Form” or “PF” that encodes the phonological information of an S-structure generated by C_{HL} , to create a PF representation of that S-structure (Chomsky (1995a): 2). Along similar lines, the MP also proposes a level of representation called “Logical Form” or “LF” where information about the meaning of an S-structure generated by C_{HL} is encoded, to create an LF representation of that S-structure.

With the help of these two levels, the MP proposes the following form for C_{HL} – it is made up of one operation, Merge, which builds a representation K (i.e. the S-structure) from an array A of lexical

items, and via Σ , Σ being the series of steps K_0 to K_n from which K is built. An operation called “Spell-Out” sends K to PF, where it is converted into a form (i.e. the PF representation π) that can be *interpreted* by the SM system, which means that the SM system can find the correct pronunciation for it. What remains of K is now sent to LF, where it is converted into a form (i.e. the LF representation λ) that can be *semantically* interpreted by the CI system, which allows a hearer to comprehend it. Mathematically, this can be formalized as, $C_{HL}(A) = (\pi, \lambda)$, where $A = (a,b,c,\dots)$. In other words, if π and λ are legitimate PF and LF representations respectively, they converge into the *interpretable* pair (π, λ) for the SM and CI systems to interpret (which is a phenomenon called “Full Interpretation” or “FI”), and C_{HL} ’s generation of the sentence will be successful. If, on the other hand, π and λ are *not* legitimate PF and LF representations, they will fail FI, and will yield an unpronounceable (or ‘unsignable’ in the case of sign language, cf. Perlmutter (1992)) or a meaningless sentence, which will lead C_{HL} ’s generation of the sentence to crash (Uriagereka (1998): 98-103, 147-148).⁵⁵

We see from the above that all that the MP proposes, in order to specify the form of C_{HL} , is the operation Merge, and two levels of representation PF and LF, where an S-structure K , resulting from the workings of Merge, is converted into the PF and LF objects π and λ . (The MP also requires an array A of items, specifically *lexical* items, for Merge to work on – which is a point I will return to in a few pages.) Notice that this is *all* that is conceptually necessary for the MP to propose, for C_{HL} to be able to generate usable (i.e. pronounceable and comprehensible) sentences. Attributing anything else to C_{HL} would be unnecessary – it would unnecessarily add to C_{HL} ’s computational load. However, by proposing just these

⁵⁵ Just because they converge and receive FI, does not mean that π and λ will necessarily be interpreted by the SM and CI systems though, since interpretation by the SM and CI systems is ultimately a matter of linguistic *performance*. This leads to an important separation between grammaticality and interpretability, as a result of which one might try to assign an interpretation to an uninterpretable sentence, just because it is grammatical – as one can do with literally nonsensical, yet grammatical, sentences (like “colorless green ideas sleep furiously”), to arrive at a poetic interpretation of the sentence instead. In contrast, an ungrammatical sentence might also be fully interpretable, as can be seen in the interpretable, but ungrammatical, speech of non-native speakers of a language. Juan Uriagereka argues that this possibly results from a generated structure’s not being the most optimal way of merging its constituents (as in the ungrammatical “there seems a man to be here”, as opposed to “there seems to be a man here”) (Uriagereka (1998): 153).

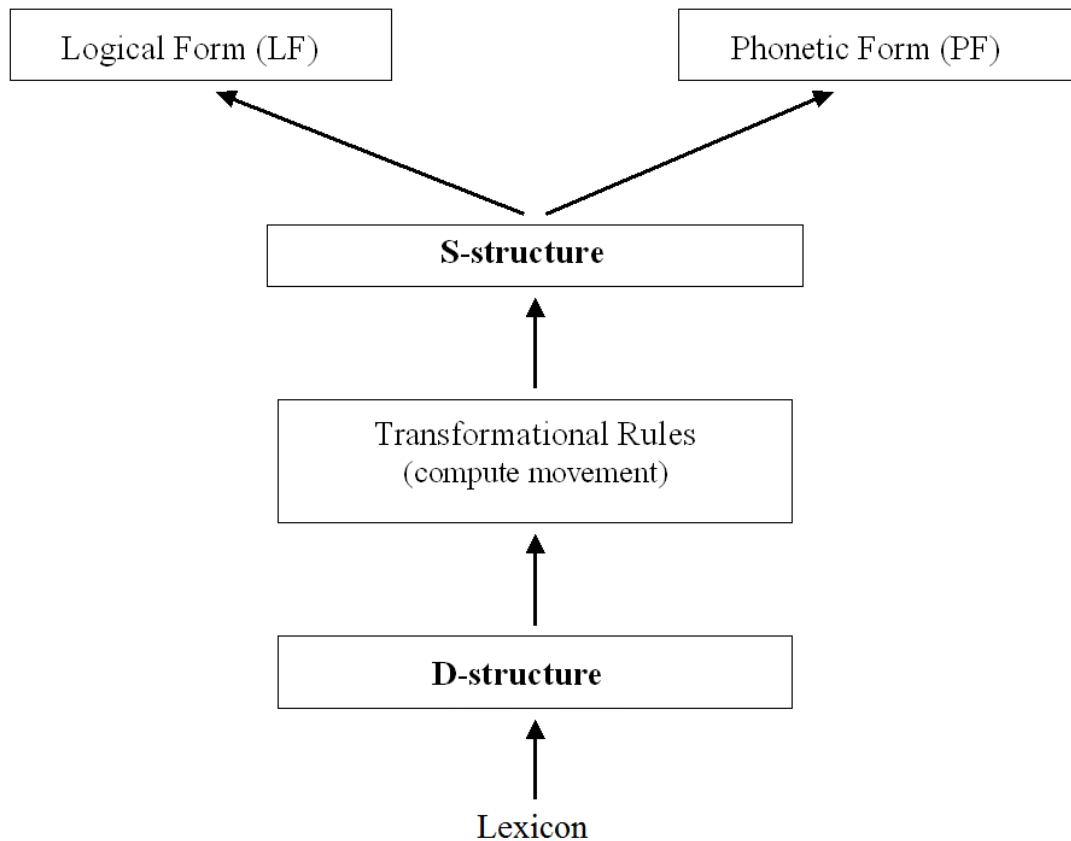
specifications for it, the MP also ends up giving the optimally economical account of C_{HL} possible – which accords with the third factor requirements that the MP makes of any formal study of language.

So, *this* is how the Minimalist Program explains language – this is how it accounts for the form (as opposed to the function) of C_{HL} . Example 1.1-5 gives a sketch of this form. Note that this sketch is a first draft – I will have reason to revise it when we return to the MP’s explanation of C_{HL} in the next chapter, especially to examine some of its more critical, and technical, aspects, insofar as they arise within the framework of P&P theory. The example shows us how the single operation Merge takes items from the lexicon and merges them into D- and S-structures, which are then mapped to the SM and CI systems via the, and only the, two levels of representation called PF and LF. I have noted earlier that the MP actually does away with the distinction between D- and S-structure, but I have depicted them in the example anyway, just to illustrate where the more traditional notions of “deep” and “surface” structure might figure within a Minimalist model of C_{HL} .

Notice that the above view of the form of C_{HL} makes the grammaticality of generated sentences to be of secondary importance to their *interpretability* by the SM and CI systems. So, it will not do for C_{HL} to generate structures that are grammatical, but incomprehensible or unpronounceable (like “colorless green ideas sleep furiously”), because this would defeat the requirement that the properties of C_{HL} exist only to meet grammar-external conditions on language. Therefore, the Minimalist description of C_{HL} ensures that every linguistic structure processed by this system is accessible to the grammar-external SM and CI systems.

Notice also how the above characterization of the human language faculty corresponds with the old Saussurian idea that language is an arbitrary pairing of sound and meaning, for this is exactly what is achieved by the convergence of PF objects (i.e. ‘sound’) and LF objects (i.e. ‘meaning’) under Full Interpretation. This pairing remains arbitrary even in a Minimalist approach to language because the MP

Example 1.1-5. An architecture of C_{HL} according to the Minimalist Program (I)



never attempts to explain how or why this pairing occurs (which one could do only if there was a non-arbitrary reason for this pairing). All that the MP asserts is *that* this pairing occurs, necessarily, for C_{HL} to meet grammar-external conditions on language.

For this reason, the MP is deeply Aristotelian too, as was Saussurian structuralism, particularly in its notion of Full Interpretation. FI maps the matter of language to its form, which accords with the Aristotelian idea that all substances are mappings of matter and form. However, importantly, this mapping of matter to form in language is *not the same* as the arbitrary pairing of sound and meaning in language, even though they both occur during Full Interpretation. (So, Minimalism is Aristotelian in a different way than structuralism was.) FI pairs sound with meaning arbitrarily, in the convergence of PF

and LF representations – but it also maps the meaningful, physical signal that a speaker utters and a hearer perceives to the form C_{HL} gives it, and *this* is where the mapping of matter and form occurs in language. In other words, the MP does not treat LF representations themselves as the matter of language, which are given form through PF representations. Instead, they *both* constitute the matter of language, as ‘informed’ by the workings of C_{HL} .⁵⁶

This makes the workings of C_{HL} , i.e. grammar, conceptually prior to semantics and phonology in the Minimalist description of language – which is actually a rather different way of thinking about grammar than is the case in some other paradigms of language scholarship. In some of these other paradigms, which include Saussurian structuralism, grammar is seen merely as a mediator between semantics and phonology, in which semantics, as the matter of language, is mapped to phonology, which is taken to be the form of language. (So, in these views, the pairing of sound and meaning *would* be equivalent to an Aristotelian mapping of form and matter.)

Such a mapping could happen in two ways – i.e. either by LF representations being generated from PF representations, or by PF representations being generated from LF representations. The former involves a ‘parsing’ view of language, in which grammar merely combines phonological units derived from parsing a sonic structure into meaningful expressions, starting with frequencies, then phonemes, then syllables and finally words and larger lexical units. The latter, on the other hand involves a ‘pragmatic’ view of language, in which grammar is merely a vehicle for communicating efficiently, by mapping semantic units to their appropriate phonological realizations. Both parsing and pragmatic views of language are obviously attractive from a functionalist perspective, because we cannot put language to its various communicative functions if we cannot parse speech sounds, and a language that can be parsed is useless if it cannot be used to connect words with objects and events in the external world. But we have

⁵⁶ There is one significant difference here though, between the MP and traditional Aristotelianism. This is the fact that the mapping of matter to form was *ontologically* necessary for Aristotle, because without this there would be no substances. However, the mapping of matter to form is not ontologically necessary, it is only *conceptually* necessary in the MP, in order to give an optimally economical account of C_{HL} . What C_{HL} really *is*, ontologically, is something generative linguistics has always been agnostic about – which is why even the notion of “grammar”, as mentioned earlier in this chapter, is really just the name of a *theory* proposed by generative linguists to account for the form of C_{HL} . C_{HL} itself is what it is.

already explored in detail the problems inherent in such functionalist approaches to both language and music, which results from neglecting the role of grammar in these systems – which is why the unique focus of generative theory on the grammatical foundations of language (and music) is so striking and suggestive.⁵⁷

There is one last aspect of the MP's description of C_{HL} I would like to discuss, before moving on to an examination of what all of this implies for a formalist, computational theory of music. This has to do with the role of the lexicon in language. As discussed above, Minimalism only proposes Merge, and the two levels of representation called PF and LF, to account for the form of C_{HL} – but Merge needs access to an array of *lexical items* too, without which it cannot even generate the PF and LF structures required for Full Interpretation to occur, and for the generative process to succeed. So, Minimalism requires knowledge of a lexicon too, as a conceptually necessary component of its description of language.

The problem is that this lexicon has to be largely learned – i.e. one has to learn a large list of vocabulary items from one's *linguistic environment* when acquiring a language, which is something young children do quickly and efficiently when acquiring their native languages, often in a stage known (controversially) as a “vocabulary spurt” (Clark (2003): 83-86, Ganger and Brent (2004)). Moreover, it is obvious that lexicons vary from language to language too, i.e. from linguistic environment to linguistic

⁵⁷ Moreover, there is good reason to believe that non-generative approaches to language of the kind suggested above, are not even theoretically feasible, because it is not clear how one can even derive phonology from semantics, or semantics from phonology, as is inherent in these approaches – as opposed to deriving them both from the workings of grammar, as generative theory does. As Juan Uriagereka says, it is much harder to derive phonological structure directly from semantics than from grammar, because the semantic structure of language (specifically its logical structure) is much more messy than its grammatical structure (Uriagereka (1998): 158-162). For example, the famous sentence (which is also the title of a Dean Martin hit) “everybody loves somebody” has at least two interpretations – viz. that every person loves a different person, or every person loves one specific individual – so it is not clear how this ambiguous semantic structure can give rise to the *specific* phonological structure of a sentence by itself (despite the fact that this approach is inherent in some very popular, non-generative approaches to language, e.g. Lakoff (1990)). One could attempt to derive the meaning of a sentence from its phonology instead – but this would not work either because the phonological structure of a sentence is extremely complicated too, given the wide variety of intonational variations, reductions, contractions etc. seen in such structure. For example, consider how hard it would be to derive a consistent LF from the two phonetically-different, but grammatically-identical, sentences “shut your face” and the comic “shaddap you face” (the title of the Joe Dolce song). Despite this, the parsing approach to language (and music, as seems to be the case for the famous theory of musical structure proposed by Lerdahl and Jackendoff (1983)) remains popular as well, possibly for the functionalist reasons discussed above.

environment – even if one ignores famous controversies about to *what* extent they vary, e.g. in the number of words they have for things like snow and reindeer.⁵⁸ All of which implies that there is something language-specific and environment- or culture-specific about *lexical* (as opposed to grammatical) acquisition – which militates against the MP’s desire to explain language in terms of *universal* principles of form and economy.

But it is likely that much of lexical structure has no bearing on the form of C_{HL} . That is, apart from the minimal fact that C_{HL} needs a lexicon to work on, how many and what kind of words a language’s lexicon has, and how this differs from some other language, does not seem to affect how Merge operates on them to generate PF and LF representations. As Juan Uriagereka humorously puts this, there does not seem to be any language “that allows its speakers to use some grammatical process just because it has a word for some variety of lemming” (Uriagereka (1998): 105). Moreover, when talking about lexical structure, an important distinction needs to be made between two different kinds of words – a distinction that thinkers as far back as the 9th century Persian philosopher and scientist Al-Farabi seemed to have been aware of (Uriagereka (1998): 120-122). This is the distinction between the potentially infinite group of garden-variety words, which languages use to talk about snow, reindeer, fire, chairs etc, and a small, finite group of words called “grammatical formatives”, which are often not even pronounced in a sentence, and which includes complementizers like “that”, determiners like “the”, and tense markers like “will”. Garden-variety words vary from language to language and have to be learned from one’s linguistic environment, but they do not seem to affect the workings of C_{HL} , as just mentioned. However, grammatical formatives do affect the workings of C_{HL} in very significant ways, as we will explore extensively in the next chapter – and they seem to be fixed across languages, and are *not* learned from the linguistic environment. This is evident from the fact that they are not normally seen in socially-constructed languages like pidgins (as a result of which pidgins do not normally display the complex

⁵⁸ See in this regard Robson (2013), and the linguist Geoff Pullum’s response in, “Bad science reporting again: the Eskimos are back”, published online January 15, 2013 at <http://languageblog ldc.upenn.edu/nll/?p=4419>. Accessed April 6, 2013.

computations, such as those involving recursion, that these formatives make possible) – however, they still show up in the creoles spoken by the children of pidgin speakers, which also endows these creoles with various, often recursive, grammatical complexities not seen in the pidgins they arise from. Since the children of pidgin speakers are only exposed to the linguistic environment of these pidgins, which lacks grammatical formatives, the fact that these formatives still show up in the resulting creoles spoken by these children suggests that they are innate, or genetically-endowed – and therefore universally present among all members of the species. So, the relevance of grammatical formatives to C_{HL} , and the irrelevance of garden-variety lexical items to C_{HL} (apart from the minimal fact that they need to exist for Merge to merge them) suggests that the relation of the lexicon to C_{HL} is such that it does not threaten the MP’s streamlined, unified, *minimalist* description of human language.⁵⁹

⁵⁹ A final critique worth discussing at this point is that even if the grammatical aspects of garden-variety words, as opposed to formatives, do not determine the workings of C_{HL} , their *semantic* or *phonological* aspects might, because these aspects are required for full-blown grammatical ability as well. This is especially true with regard to how C_{HL} generates semantic (i.e. LF) and phonological (i.e. PF) representations from them, e.g. from different phonetic categories such as [t] and [d]. There has been some research in the philosophy of language that has tried to ascertain what these phonological, and especially semantic, aspects of lexical structure are. For example, Rudolf Carnap discussed features that one would most likely associate with a certain word in terms of “meaning postulates”, which determine the relationships among words (or groups of words) in a sentence, such as between “grown in the tropics” and “coffee” (Carnap (1952)). But as Juan Uriagereka argues, such postulates are unlikely to be *intrinsic* semantic features of a word like “coffee”, i.e. features specified in the lexicon itself, without reference to the numerous socio-cultural, political, and economic conditions that enhance the meaning of a word – since such conditions do not affect any grammatical process in language, as mentioned above (Uriagereka (1998): 134-146). In contrast, features like mass and count, singular and plural, animate and inanimate, masculine and feminine etc. *do* often affect grammatical processes. So they might be considered intrinsic semantic aspects of words – an *innate semantics* for language.

Now, some linguists have argued that the semantic aspects of words are not only more complicated than the simple binary pairings just mentioned, but also affect the grammatical processes of a language in very language-specific ways. For example, in the case of the Australian aboriginal language Dyirbal, George Lakoff has famously described a separate grammatical classifier for women, fireflies, fire, the sun, shields and a whole host of other things, and another classifier for edible fruit and the plants they grow on, cigarettes, honey, wine, and cake (Lakoff (1990): 93). Another example is Burmese, which grammatically distinguishes “long, slender objects” from “vertical, slender objects”, “spherical objects”, and “thin, flat objects” (Burling (1965)). Such arguments lie behind the idea that sentences are generated from semantics – or rather, that PF representations are derived from LF representations with grammar as a mere intermediary – which can be found notably in “cognitive linguistics” approaches to linguistic structure, of which Lakoff’s work is a central representative. But just because various languages show these diverse, language-specific, semantic or phonological features does not mean that simpler, more universal, ways of classifying these features cannot be found. One just has to look for them, which obviously involves losing some preconceptions – of the functionalist, sociocultural kind – about language. Uriagereka specifically suggests a more abstract, general way of classifying semantic features in terms of simple parametric combinations of substance, form and change. This is based on the idea that classifiers for change presuppose substance and form too – there do not seem to be any grammatical morphemes that encode change without form (such as a morpheme for something like an ocean, which changes and also seems to lack an explicit form), and there is no classification for items that have form but no substance (such as numbers). So, the system of semantic features shows an underlying harmony,

In the preceding pages, I have tried to give a summary of the philosophical foundations of the MP, along with a brief review of its technical proposals of Merge, LF and PF. The big question now is what all of this has to do with music. I have explored extensively the idea that music has a unique hierarchical, recursive, computational structure which is difficult to explain in functionalist terms, and which makes music strikingly similar to language too. Therefore, an approach to music and language that focuses on their shared computational features seems to be the best way to demonstrate the resilient connection between music and language noted by so many thinkers over the ages, and which is particularly evident in Henry Wadsworth Longfellow's quotation with which I began this dissertation. And this is the approach that we have explored as the "musilanguage hypothesis".

It is for this reason that the diversity of approaches to understanding music and language explored in this chapter all seem to fail when it comes to describing any deep similarities between music and language, because of their inability to describe and explain the computational aspects of these two systems – and it is precisely for this reason that generative linguistics, with its emphasis on the centrality of C_{HL} in descriptions of language, seem to offer such a provocative and insightful way of thinking about *musical* structure too. But we have seen that in recent Minimalist developments in generative linguistics, C_{HL} is described as having not only the kind of hierarchical, recursive features that might be shared with music, it is also underspecified and economical in its structure. So, a musilinguistic approach to describing any deep similarities between language and music will only succeed if musical structure can be described in terms of Merge-based operations that yield *musical* LF and PF representations – that too in a way that potentially reveals language-like properties of discreteness, underspecification, and structural economy within music as well. And since this seems to be the only approach that might reveal music and language to be shared species-specific traits of humankind, as John Blacking put it – i.e. shared aspects of human *nature* – demonstrating music to have the above features is critical and indispensable, but also a tall order to ask (especially of a limited project such as a doctoral dissertation).

which can be represented in terms of increasing dimensions in a multi-dimensional system of features like substance, form and change. (In this regard, see also Hinzen (2011).)

What is striking though, is that the path to such a Minimalist approach to music seems to have already been paved, albeit implicitly, in the music-theoretic literature, and specifically in the writings of the great early 20th century Austrian music theorist Heinrich Schenker. As I remarked earlier in this chapter, while citing the linguist Derek Bickerton, it is in theoretical approaches to issues in music and language that answers to some of our current questions about musical and linguistic structure may be found, in a way not obtainable from any other disciplinary approach. So, it is perhaps of no surprise that a justification of the musilanguage hypothesis will come from linguistic theory and, as we shall soon see, music theory too. We have already explored the linguistic-theoretical side of this issue in the above examination of linguistic Minimalism, and so now it is time to examine the music-theoretic side of the issue, particularly through the lens of Schenkerian theory.

The rest of this dissertation, therefore, will be devoted to defending the claim that musical structure *can* be described in terms of discrete, underspecified, economical, Merge-based operations that yield musical LF and PF representations – primarily because such a description of musical structure seems to be implicit in Schenkerian theory. The intriguing conclusion that this leads to is that not only are music and language similar, and possibly identical, because of their shared computational structure (as proposed by the musilanguage hypothesis), but that musical and linguistic *theory* are identical too, given the striking similarities between Chomskyan approaches to language and, as I will now describe, Schenkerian approaches to music. This also suggests a remarkable, if hitherto-ignored, parallel history of music and linguistic theory, i.e. in the development of the generative approach to music and language in both traditions. However, there has been little active collaboration between music theorists and linguists to explore the potentially shared structure of music and language. So, the possible identity of musical and linguistic theory seems to be the result of the possible identity of music and language in and of themselves – which is why the generative approach to both these systems has converged historically, in Schenkerian and Chomskyan theory. (This, consequently, provides a different kind of justification for the musilanguage hypothesis too.)

This leads to the two identity theses for music and language from which this chapter gets its title:

Identity Thesis A: Music theory and linguistic theory are identical.

Identity Thesis B: Music and language are identical, as the most plausible explanation for Thesis A.

This dissertation is therefore devoted to defending not only Identity Thesis B, which is just the musilanguage hypothesis we have been exploring, but also Identity Thesis A, since it is in the overlaps between Schenkerian music theory and Chomskyan linguistics that the best case can be made for the musilanguage hypothesis, i.e. for the claim that both music and language have a shared, economical, underspecified, computational structure – which is what makes them identical too, given the centrality of this computational system in defining music and language as a whole.

In the next chapter I will explore the computational structure of musical grammar from a Schenkerian perspective, specifically the possibility that Merge is the only grammatical operation needed to generate musical S-structures. In chapter 1.3, I will further explore whether this Schenkerian, Minimalist description of musical grammar can also describe the generation of S-structures across musical idioms, which will reveal both the underspecified and universal nature of this grammar – and therefore provide a Minimalist account for how variety occurs across musical idioms. Finally, in chapters 2.1 to 2.3, i.e. in the second half of the dissertation, I will explore the possibility that musical structure has LF and PF levels of representation too, as also seems to be implicit in Schenkerian approaches to music.

Now Schenker's ideas have been subject to much debate and scrutiny, and also much controversy, given his status as one of the central figures in Western music scholarship. So, my above claim that a Minimalist approach to musical structure is inherent in Schenkerian theory cannot be stated lightly, and is need of vigorous justification. This is why I shall be defending aspects of this claim step-by-step, over the several hundred pages of this dissertation, through the topics I just mentioned. But there is already a certain historical precedent for interpreting Schenkerian theory in the way I have proposed – at least as a formalist, scientific theory of musical structure, if not a Minimalist one. This happened as part of a broader intellectual approach to music scholarship that arose specifically in North America in the mid-20th century. A brief description of this paradigm might therefore provide an initial justification for

the research program of this dissertation, or at least set the course for my subsequent arguments. So, to this historical discussion I now turn.

1.1.3. Schenker, Humboldt, and the Origins of Generative Music/Linguistic Theory

Around the same time that John Blacking expressed his views on music and language, the famed conductor and composer Leonard Bernstein expressed his thoughts on this topic too – in probably the most famous attempt to explore these issues from a musician’s perspective. Speaking about the musical aspect of this issue, in his 1973 Charles Norton Lectures at Harvard University, titled *The Unanswered Question*, and specifically addressing the crisis of what “music” means with the emergence of atonality in early twentieth century Western music, he highlights the profoundly metaphysical character of the question:

“I’ve always felt he [Charles Ives, in the context of *his* piece from 1908, “The Unanswered Question”] was also asking another question, a purely musical one – “whither music?” – as that question must have been asked by Musical Man entering the twentieth century. Today, with that century sixty-five years older, we are still asking it; only it is not quite the same question as it was then. And so the purpose of these six lectures is not so much to answer the question, as to understand it, to redefine it. Even to guess at the answer to “whither music?” we must first ask Whence music? What music? and Whose music?” (Bernstein (1976): 5)

In this passage Bernstein states that there is a need to characterize the origins, locations, and practitioners of music before one can say what music really is, which clearly reveals his skepticism about musical universals given its cross-idiomatic diversity – a skepticism that he shared with a variety of scholarly traditions including, as we saw at the beginning of the chapter, ethnomusicology. However, he also suggests that analogies between music and language, including Longfellow’s declamation about music being the universal language of humankind, are not necessarily as clichéd and hyperbolic as they might appear at first glance, because there might be a deeper link between our ability to make music, and our innate competence for language, as described in Chomskyan linguistics:

“This philosophical science called linguistics has become our newest key to *self-discovery* [my emphasis]. With each passing year it seems to substantiate ever more convincingly the hypothesis of innate grammatical competence (as Chomsky calls it), a genetically endowed language faculty which is *universal*. It is a human endowment; it proclaims the unique power of the human spirit. Well, so does

music. ... By building analogies between musical and linguistic procedures ... that cliché about the Universal Language [could] be debunked or confirmed, or at least clarified.” (Bernstein (1976): 8-10)

Bernstein went on to build several such analogies in *The Unanswered Question*, which triggered an intense interest in using linguistic models in the study of music that continues to this day (e.g. in this dissertation!), making Bernstein’s text a landmark contribution to the study of the connection between music and language. But his ideas also came under severe attack for making the same crude analogies between language and music that he criticized earlier thinkers like Longfellow for making. For example, in his review of *The Unanswered Question*, music theorist Allan Keiler is particularly harsh when he says “It would be tedious to hold up for scrutiny the inconsistency, faulty argument, and empty terminological morass that characterizes large parts of Bernstein’s lectures” (Keiler (1978b): 198).⁶⁰

Part of Keiler’s critique is that Bernstein, even if just trying to find some innocent equivalences between music and language, does not even make these equivalences consistently, which renders his attempt to relate music and language ineffective (Keiler (1978b): 211-212). For example, Bernstein attempts to describe musical parts of speech in order to equate music with language, which he does by equating musical rhythm with verbs, melodies with nouns, and harmony with adjectives. But he also equates words, which presumably includes nouns, verbs and adjectives, with musical *notes* in other parts of the text – so that a group of notes, i.e. a melody, would be a group of words but also, following his earlier equation, a noun, i.e. a *single* word. If this group of notes constitutes an arpeggio, one can ‘verticalize’ it to get a triad, as a harmonic entity, which Bernstein considers to be an adjective. In other words, he ends up deriving adjectives from nouns – which is clearly counterintuitive to what happens in language.

Criticisms of this sort have made scholars wary of looking for specific analogies between music and language in the years since Bernstein’s Norton Lectures – which is why the *Unanswered Question* has

⁶⁰ In this review, Allan Keiler also cites part of John Blacking’s statement about the species-specificity of music and language referenced at the beginning of this chapter, and he later wrote a paper (Keiler (1989)), whose title is also a play on the title of Blacking’s text. To my mind, this suggests an important intellectual connection between questions about the music/language nexus as pursued by ethnomusicology and by (linguistics-inspired) cognitive music theory. Which is why this chapter makes a point of exploring both!

largely remained unanswered. But even if Bernstein's attempt to relate music and language were ultimately ineffective, his project is important because his basic belief that music and language might have a deeper connection, as aspects of human nature, has resonated with other thinkers.

In particular, Allan Keiler himself was among a small group of scholars who went on to explore connections between language and music. In addition to being a music theorist, Keiler was a trained linguist (he was one of the last doctoral students of the illustrious linguist Roman Jakobson at Harvard), and subsequently became one of the first music theorists to introduce ideas from generative linguistics into music theory, which he did in a series of articles spanning the late 1970s to the mid 1990s (see the bibliography for a more-or-less complete list of these papers). In these papers he explored issues of grammar, hierarchy, creativity, and meaning in music from both historical/philosophical and technical music-theoretic perspectives – and importantly, from within a Schenkerian framework, which makes his work particularly important for this dissertation. (One could argue that this dissertation picks up from where Keiler left off, in his work on Schenkerian theory and musical grammar.)

For the above reasons, much of my discussion of a Minimalist approach to musical structure in the next chapter onwards will be based on Keiler's specific interpretation of Schenkerian theory. But as mentioned earlier, there have been several different 'takes' on Schenker's ideas, of which Keiler's is only one. So, it might be worthwhile at this point to describe some of the foundational and (relatively) uncontroversial aspects of Schenker's ideas – in order to provide a context within which to subsequently understand Keiler's (and my) specific interpretation of these ideas.

Heinrich Schenker (1868-1935) was a composer, pianist, and music critic who lived in Vienna for most of his life. After trying to establish himself primarily as a composer, he became more famous in the second half of his life for his analyses of masterworks by the great composers within the 'Bach to Brahms' common-practice period of Western Classical tonal music, and also for his numerous editions of these masterworks based on *Urtext* editions and other autograph sources. Part of Schenker's interest in the Western common-practice idiom, and specifically the 'Germanic' masters in its canon (including those

that lived and worked in what is now Austria or Hungary, i.e. composers like Mozart and Brahms), stemmed from his conservative, German nationalist political convictions, and his cultural traditionalism. (Which is ironic given that he was a Jew, and that ten years after his death his wife, Jeanette Schenker, died in the Nazi concentration camp of Theresienstadt.) This ideological bent gives parts of Schenker's musical writing an elitist and chauvinist flavor, which results also in its being biting critical of the modernist music of his age. This made Schenker many enemies in the intellectual and artistic world of fin-de-siècle Vienna, which made a wider reception of Schenkerian ideas problematic initially, and which is something that continues to be the case in certain intellectual circles even today (for more on Schenker's politics and polemics, see Cook (1989a)). However, interest in Schenkerian theory surged in the latter half of the twentieth century, particularly in North America, mainly through several of Schenker's students who emigrated to the USA in the post-war years, and who generally distanced themselves from the unsavory aspects of Schenker's politics, deeming them irrelevant to his technical ideas on musical structure – just as the anti-semitic beliefs of Gottlob Frege were deemed to be of little relevance to his revolutionary technical work in logic and the philosophy of language.

Schenker's technical ideas on musical structure arose from his above analyses of common-practice masterworks, which he developed into a more general theory of common-practice tonality. This can be studied through the several articles and books he (often self-) published, the most well-known of which is the three volume *Neue musikalische Theorien und Phantasien* (“New musical theories and fantasies”) consisting of *Harmonielehre* (originally published in 1906, English translation “Harmony” (1973)), *Kontrapunkt* (originally published in two parts in 1910 and 1922, English translation “Counterpoint” (1987)), and *Der freie Satz* (originally published in 1935, English translation “Free Composition” (1979)). Through these works, Schenker attempted to show how general principles of harmony and counterpoint govern tonal music, even in the works of the great master composers. However, given his elitist attitude to music, he believed only the master composers had knowledge of these governing principles, which is why he thought that music students should spend their time

analyzing masterpieces by these composers, rather than trying to directly compose themselves, as a means to gain a true understanding of how tonality works (Schenker (1979): xxii).

This last point makes it evident that Schenker was not trying to formulate a general theory of musical grammar, or a description of the human psychological faculty of music, given his clear lack of interest in how the musical mind works in all but a small number of master composers. This is no surprise given that Schenker was not a scientist or even an academic in the more institutional sense of the term.⁶¹ As primarily a music critic, analyst and performer, he certainly was not inclined to present his ideas as a systematic, scientific theory of music – certainly not in the way “theory” is understood in the sciences. However, his description of the above principles of harmony and counterpoint are abstract enough (albeit supported by the numerous analytical examples he provides from the tonal literature), and are grounded in a particularly rich intellectual tradition, that a case can certainly be made (and has been made by a number of modern theorists) for Schenker’s ideas constituting at least a proto-generative theory of musical grammar.

For instance, consider the concept that is arguably the best known component of Schenkerian theory, viz. the *Ursatz*, which is usually translated as the “fundamental structure” of a tonal piece (Schenker (1979): 6), but might be better translated in our current Minimalist context as the “abstract form” of the piece. The defining features of this abstract form of a piece are an upper, soprano voice (called the *Urfinie*, or “fundamental line”) that descends from either scale degree 3, 5 or 8 (known as the *Kopftone*, or “head tone”) to scale degree 1, and a lower, bass voice (called the *Bassbrechung*, or “bass arpeggiation”) that begins on scale degree 1, proceeds to scale degree 5, and then returns to scale degree 1. The counterpoint between these two voices is what largely governs how the piece is structured. As a result of this, Schenker took the principles of counterpoint, especially strict counterpoint – and particularly the 18th century theorist Johann Joseph Fux’s formulation of species-based strict counterpoint

⁶¹ In fact, the entire title of his aforementioned three-volume *magnum opus* is “New Musical Theories and Fantasies – by an *Artist*” [my emphasis]. As Oswald Jonas – one of Schenker’s preeminent students, and the editor of the best known English translation of the first volume of the set (i.e. *Harmony*) – points out, this inscription suggests that Schenker’s theory was essentially aimed at “finding artistic solutions to artistic problems” (Schenker (1973): v).

– to be especially relevant to understanding tonal structure. However, individual common-practice tonal pieces often cannot be derived from just the principles of strict counterpoint, as we shall explore in more detail later. The principles of harmony, as specified in Schenker’s theory of the scale-step (i.e. *Stufe*) are equally important. Consequently, Schenker often depicts the *Ursatz* as a four-voice structure, with alto and tenor inner voices too – which realizes a root position I – V – I harmonic progression, and therefore shows the *Ursatz*’s dependence on harmonic principles in addition to contrapuntal ones. Such a four-voice *Ursatz* is illustrated in Example 1.1-6. (This example also displays an *Urlinie* that descends specifically from a *Kopfton* of scale degree 3.)

Example 1.1-6. The *Ursatz* (“Fundamental Structure”) in Schenkerian theory

The image shows a musical staff with two systems of staves (treble and bass clefs). Above the treble clef staff, there are three chords represented by whole notes. The first chord is a triad with a scale degree 3 (^3) above it. The second chord is a dyad with a scale degree 2 (^2) above it. The third chord is a triad with a scale degree 1 (^1) above it. Below the bass clef staff, there are three whole notes corresponding to the chords above, with Roman numerals I, V, and I written below them respectively.

Now, the *Ursatz* is the abstract form of a tonal piece for at least a couple of reasons. First, it has no rhythm, specifically no *durational* rhythm – i.e. it has no concrete temporal articulation in the way a real piece of music does (Schenker (1979): 15). (The fact that Example 1.1-6 uses whole notes to depict the *Ursatz* is just a convention for depicting a pitch structure that has no inherent duration – it does not imply that the chords of the *Ursatz* last for, say, a bar each of 4/4 time, as would happen in an *actual* piece of music.) Secondly, and more importantly, the *Ursatz* is not a real piece of music but an entity *from which*

real pieces of music are derived, through certain contrapuntal (or “voice-leading”) operations.⁶² To understand this, consider Example 1.1-7, which shows how the first phrase of the aria “Casta diva” from Vincenzo Bellini’s opera *Norma* is derived. The top stave system of the example, i.e. level A, depicts the Schenkerian *Ursatz*, presented here in F major. As in the previous example, the *Ursatz* here also descends from a scale degree 3 *Kopfton* (represented by the pitch A4 in the soprano), leading to what one might call a “3-line *Ursatz*”.

In level B of the example, the *Kopfton* is held over into the V chord, displacing the V chord’s soprano G4 in the process. This leads to the whole note G4 being held over into the final tonic triad too, which displaces the final soprano F4 to the second half of the final tonic triad. The end result is a chain of suspension figures, created by these displacements in the soprano voice with regard to the bass, which are indicated by the labels “6 - 5” and “9 - 8”. A suspension is a voice-leading operation known as a “rhythmic figuration” (Aldwell and Schachter (2011): 392-410), which normally results from pitches in a voice being displaced in the manner just described (Salzer and Schachter (1969): 78). This leads to a change in the soprano voice from the way it was in level A, which is why suspensions and other such rhythmic figurations are specifically *voice-leading operations*.

The end result of this voice-leading operation is an *actual* musical phrase, i.e. the structure represented in level B – which demonstrates, therefore, how an actual phrase can be derived from an abstract *Ursatz* through one or more voice-leading operations. One of the things that makes level B an actual phrase is that, unlike the *Ursatz*, it has a durational rhythm. This is demonstrated by the fact that the G4 in the soprano voice is held for the *length* of a whole note in it – half supported by the V chord, the other half by the final tonic chord over which it is suspended – so that its whole note representation now

⁶² For this reason the psychologist John Sloboda is wrong when he asserts that the Schenkerian *Ursatz* is not analogous to a linguistic deep structure because, “the *Ursatz* is, in itself, a legal although trivial piece of music ... In contrast ... Chomskyan deep structures are not, in themselves, acceptable sentences” (Sloboda (1985): 15). The *Ursatz* is *not* a legal piece of music because it has no duration – it is merely a description of the most fundamental grammatical relationships between pitches in a musical phrase. It does, however, become a legal piece of music once it is assigned duration and when it is composed out to generate a musical surface via voice-leading operations.

Example 1.1-7. Bellini, “Casta diva” aria from *Norma*: Generation of the first phrase from *Ursatz*

A

I V I

B

6 - 5 9 - 8

C

6 - 5 9 - 8
4 - 3

D

Norma

Ca - - - sta Di - va, etc.

Orchestra

implies an actual duration (equal to two half notes), and is not just a convention for representing abstract, arrhythmic pitches anymore. In other words, the structure in level B can be *performed* in the way an *Ursatz* cannot. However, as an actual phrase, level B is different from the actual phrase that opens the *Casta diva* aria, so further voice-leading operations are required to get us from level B to the actual aria phrase. Level C suggests one such further voice-leading operation, namely a consonant skip to C5 that is introduced between the A4 and G4 of the soprano in level B. This C5 is represented by a quarter note symbol, which is a Schenkerian convention that indicates the later appearance of this pitch in the derivation of a phrase. This also indicates how this pitch belongs more to the actual phrase, than it does to the abstract *Ursatz* from which the phrase is derived, which is why it appears ‘further away’ from the level of the *Ursatz* in the derivation. (A common way of referring to such pitches in a Schenkerian description of tonal passages is to say that it is not a “structural” pitch – which is really just a way of saying that it does not belong to the structure of the *Ursatz* in particular.)

Level D takes this a step further, by introducing a turn figure A4-B-flat4-A4-G4-A4 that fills in the gap between the A4 and the C4 to which the A4 skips in level C. Following the convention already stated, this turn figure is represented with unstemmed note heads, to represent how this figure arises even later in the derivation than the ‘structural’ A4 pitch (i.e. the *Kopftön*) that precedes it in the derivation. The turn figure involves two neighboring voice-leading operations as well, the first an *upper* neighboring motion that takes us from the initial A4 up to B-flat4 and then back to A4, and the second a *lower* neighboring motion that takes us from the middle A4 down to G4, and then back up to A4 again. As a result, the two neighboring motions (and the complete turn figure they give rise to) can be seen as elaborating the A4 pitch, in addition to filling in the gap between this pitch and the subsequent C4. After the C4 is reached in level D, a stepwise, *passing* voice-leading motion takes us down to G4 via a B-flat4 and an A4, which fills in the gap between the C4 and G4 of level C too. By filling in all these gaps, level D realizes a more mellifluous, conjunct melody in the soprano, which is more suitable than level C given that this phrase is specifically a phrase from an aria. For the same reason, other unstemmed note heads are

added to the melody at the right end of level D, which essentially elaborate the last structural pitch F4 of the *Uralinie*.

Finally, level E divides up the notes of level D with the specific rhythmic values of the notes of the aria to derive the exact aria phrase. (It also arpeggiates the inner voices of the I – V – I harmonic progression that undergirds this phrase, which were previously represented as just whole notes.)

The above derivation illustrates several systematic aspects of Schenkerian theory, particularly as it applies to a Minimalist description of musical structure. (There are also aspects of Schenkerian theory that are unsystematic and unscientific – and therefore non-Minimalist in their orientation too, which I shall discuss later.) First of all, the very nature of the *Ursatz* as the abstract form of a tonal piece, reflects an *internalistic* approach to musical structure that is concerned with the form of music, akin to how Minimalist linguistics is concerned with the internal form of language. Moreover, Schenker’s abstract, internalistic description of the *Ursatz* reveals that his interest in tonality’s internal form was also psychological – and specifically epistemological – in nature, not only in his belief that the master composers had *intuitive knowledge* of the *Ursatz* (Schenker (1973): 20, Schenker (1979): 18), but also in his belief that describing the derivation of a tonal phrase from the *Ursatz* requires an *understanding* or *cognition*⁶³ of how the different parts of a phrase are related to the *Ursatz*, presumably in grammatical terms (Schenker (1979): 8, 27, 68, 100).

Even more striking than the above internalistic and cognitive qualities of Schenker’s theory, is his insistence on the *Ursatz* being the *foundation* for a unified description of all other aspects of musical structure (Schenker (1979): 5) – including the semantic and rhythmic aspects of a musical phrase. This is no doubt why the *Ursatz* is often translated as the “fundamental structure” in Schenkerian theory. But this reveals, more importantly in my opinion, a deeply grammar-centric and therefore possibly Minimalist

⁶³ I have used the terms “understanding” and “cognition” here, instead of “perception”, the latter being the term used by Ernst Oster, in his 1979 translation of Schenker’s original German text. My choice of terms here is intentional, given the problematic meaning of “perception” – since this term could be taken as implying a *sensory* process (i.e. hearing) rather than a cognitive, and especially epistemological, process (i.e. thinking or understanding).

conviction about musical structure, i.e. one in which the simple, internal, abstract, grammatical form of music is seen as the basis even for things such as musical meaning – or at least as the basis for a mapping to the conceptual-intentional systems of music, through an LF level of musical structure (as I shall explore in detail in the second half of the dissertation). Which is why Schenker says that understanding musical structure involves understanding “the simpler elements upon which the far-reaching structure is to be based” (Schenker (1979): 18).

In addition to the above, Example 1.1-7 also reveals the *generative* nature of Schenker’s theory, inherent in the way that phrases are derived from an abstract *Ursatz*. This, in turn, implies a hierarchical approach to musical structure, in which parts of a phrase that appear later in the derivation are seen as hierarchically inferior, and hence as elaborating the hierarchically-superior, “structural” parts of the phrase – just as happens in generative descriptions of language. Now, the fact that the hierarchically-inferior parts of a phrase are derived later is also revealed in the several levels of structure (Schenker’s *Schichten*) that arise in the derivation of that phrase (depicted by levels A-E in Example 1.1-7). Schenker spoke specifically of three levels of structure, viz. *Hintergrund* (“background”), *Mittelgrund* (“middleground”), and *Vordergrund* (“foreground”) (Schenker (1979): 3), the foreground arising from voice-leading operations on the middleground, and the middleground from voice-leading operations on the background – with the background therefore corresponding to the ‘deepest’ level of structure, i.e. essentially the *Ursatz* itself, the foreground to the musical ‘surface’, and the middleground a level in between these. My choice of words like “deep” and “surface” here is deliberate – Schenker’s description of “background” and “foreground” are clearly analogous to the generative linguistic notions of D- and S-structure that are involved in the derivation of linguistic phrases. (This analogy comes closer to being an actual point of identity between music and language, when one considers how closely the generation of a temporally-articulated tonal foreground, from an abstract, arrhythmic background, resembles the generation of a surface PF level of representation in language, from an abstract D-structure via Merge-based computational operations.)

The specific way in which a phrase is derived in Schenker's generative approach to musical structure, however, reveals a *computational* view of musical structure too, given the way in which contrapuntal principles *operate* on harmonic structures to generate such phrases. The background-to-foreground *logical* ordering of the derivation of the above Casta diva phrase establishes this quite clearly. (Allan Keiler in particular points out how Schenker was at pains to demonstrate that his system was a background-to-foreground one, rather than the other way around (Keiler (1983-84): 201-207).) Interestingly, the logical ordering of a phrase derivation in Schenker's theory also parallels the chronological order in which Schenker wrote and published the three volumes of his *Neue musikalische Theorien und Phantasien*. That is, the derivation starts with an abstract *harmonic* structure, viz. the I – V – I structure of the *Ursatz*, which is changed through voice-leading operations (i.e. through *counterpoint*). And then, finally, the derivation of the phrase takes us from strict composition to *free composition*. This is because the derivation of level E from level A essentially follows the rules of *strict* counterpoint, but also breaks with these rules as we get closer to deriving the actual phrase – suggesting that the derivation of a tonal phrase involves ‘crossing the bridge’ from strict composition to free composition (Schenker (1987) Book 2: 175).

To understand this last point, consider the fact that in deriving the aria phrase from the *Ursatz* in Example 1.1-7 we invoked a number of voice-leading operations, most of which can be accommodated within a strict model of counterpoint – in particular the species-based model of strict counterpoint, whose specific formulation by Fux, as mentioned before, is the one Schenker seemed to be influenced by the most. Fux proposed his version of strict counterpoint in his 1725 treatise *Gradus ad Parnassum* (partial English translation “Steps to Parnassus” (1943)), in which he describes five types or “species” of counterpoint. Fux's belief was that a student should master these five species of counterpoint, in succession, in order to learn how to compose in a 16th century contrapuntal style, as epitomized by the works of Giovanni Palestrina. But since his treatise was published over a century after Palestrina's death, and at a time when contrapuntal writing based on 16th century practice had taken a backseat to the 18th century practice of Baroque counterpoint, it is slightly anachronistic, and later scholars have sometimes

criticized it for not representing Palestrina's style accurately. (For example, Knud Jeppesen has criticized Fux for allowing dissonances on the third beat of a third species exercise, even when they are preceded and followed by consonances, as being foreign to Palestrina's practice (Jeppesen (1992): 40).)

Moreover, its focus on the Palestrina style, as opposed to the common-practice contrapuntal style of later centuries, means that mastering Fux's five species of counterpoint might enable a student to write music in the style of Palestrina, but not in the style of, say, Bellini – in other words, it would not enable a student to write the *Casta diva* passage discussed in Example 1.1-7. However, Fux's five species introduce the student to the contrapuntal phenomena of passing, neighboring, and suspended motion in a systematic and rigorous way – and given how these phenomena undergird the voice-leading operations involved in the generation of the *Casta diva* passage, mastering them through a study of species counterpoint can improve a student's understanding of how voice leading works even in common-practice tonal music.

A brief summary of Fux's five species of counterpoint might make the above point clearer. So, in a first species exercise, a contrapuntal voice in whole notes is written above or below a *cantus firmus* – the latter also written in whole notes, so that that the resulting contrapuntal structure involves a note-against-note model of counterpoint. (In other words, any sense of rhythmic motion between the voices is lacking in the first species.) The intervals between the voices of a first species exercise must be either perfect or imperfect consonances as well (i.e. perfect unisons, major or minor thirds, perfect fifths, major or minor sixths, perfect octaves, and their compound forms) – with the exception that the dissonant (vertical) interval of a perfect fourth or augmented fourth/diminished fifth can also appear in a first species exercise, as long as it only appears between the upper voices of a first species exercise with three or more voices (i.e. as long as the dissonance does not involve the lowest voice). In other words, a first species exercise with three or more voices can only contain root position and first inversion major or minor triads, and first inversion diminished triads.⁶⁴

⁶⁴ This leads to the interesting problem of why the 6/5 sonority is not allowed in three-voice, first species counterpoint – and more importantly why the 6/5/3 sonority is not allowed in four-voice, first species counterpoint.

After the purely consonant texture of first species, dissonances are introduced into the counterpoint in second species onwards, but only in a very specific rule-based way – hence the description of species counterpoint as a form of “strict composition”. So, passing tones are introduced in the contrapuntal voice in second species, neighboring tones in third species, and suspensions in fourth species – all of which create dissonances with the cantus firmus, but all of which must therefore be treated in a very careful, rule-based way. Specifically, passing tones and neighboring tones must be approached by stepwise motion from a consonant preceding tone, and left by stepwise motion to a consonant

Four-voice counterpoint allows the introduction of seventh chords into the counterpoint exercise, since these chords are made up of four pitches. But they are not normally allowed in four-voice, *first species* counterpoint because the lowest voice tends to be involved in a dissonance in such chords – which is not allowed in first species counterpoint. For example, consider the minor seventh chord built on D, which has the notes D, F, A, and C. If this chord were used in a four-voice counterpoint exercise with D in the lowest voice, i.e. in *root position*, the lowest voice would make a dissonant seventh with the C above it. If the chord were in *second inversion*, with A in the lowest voice, the lowest voice would make a dissonant fourth with the D above it, and if it were in *third inversion*, with C in the lowest voice, it would make a dissonant fourth with the F, and a dissonant second with the D, above it. But an exception seems to arise if the chord is in *first inversion*, with F in the lowest voice – i.e. when it has the figured bass signature 6/5/3 – since the remaining voices all form consonances with the lowest voice. (F-D is a major sixth, F-C is a perfect fifth, and F-A is a major third). Since first inversion minor seventh chords do not have dissonances involving the lowest voice, they should be permissible in four-voice first species counterpoint, just as first inversion diminished triads are. (A similar argument can be made for other 6/5/3 sonorities, such as the first inversion major and half-diminished seventh chords, but not for major-minor (i.e. “dominant”) seventh or fully-diminished seventh chords.) However, 6/5/3 sonorities are normally not allowed in first species counterpoint, since the seventh of these chords is still treated as a dissonant entity, as a result of which 6/5/3 sonorities are normally introduced only in *fourth species* counterpoint, where the seventh of the chord is normally prepared as a *suspension*. Why such sonorities are treated exceptionally in first species counterpoint does not seem to have a satisfactory explanation in the music-theoretic literature. For example, Robert Gauldin just states that they have been traditionally treated according to rules regarding suspensions, without explaining why this is the case (Gauldin (1995): 92-93). (This might be the reason why Schenker at times thought about the origin of seventh chords in fourth species terms too – or more specifically as a sonority that arises from combining second and fourth species counterpoints in a single passage (Schenker (1987) Book 2: 210-222).) Felix Salzer and Carl Schachter attempt to explain the exceptionality of 6/5/3 sonorities by saying that they involve the dissonance of a second (or a seventh) between the upper voices represented by the “6” and the “5” of the 6/5/3 figured bass signature, which is too strong of a dissonance even when it does not involve (and is therefore offset by) the lowest voice (Salzer and Schachter (1969): 28). But to accept this, one has to accept that the dissonance of a seventh between the upper voices of a 6/5/3 sonority is stronger than the dissonance of a tritone (i.e. the “devil’s” interval) between the upper voices of a first inversion diminished triad – which is why the latter, unlike the former, is allowed in a first species exercise. This is a rather counter-intuitive statement, which Salzer and Schachter do not justify any further. John Rothgeb seems to agree with this evaluation, since he rejects Salzer and Schachter’s explanation for the argument that the intervals of a 6th and 5th above the lowest voice in the 6/5/3 sonority give conflicting cues about the root of the sonority – and this is what makes the sonority unacceptable as a ‘consonance’, which is what first species sonorities are required to be (Rothgeb (1975): 282). But the invocation of the concept of a “root” here seems to explain this contrapuntal phenomenon in *harmonic* terms, and it is not clear why harmony should have any role to play in explaining 16th century contrapuntal practice. (On a related note, Salzer and Schachter also reject augmented triads, in any inversion, in first species exercises, on the grounds that such sonorities do not exist in the diatonic system. But, again, it is not clear why diatonicism – something one usually associates with common-practice tonal harmonic composition – should have any role to play in explaining 16th century, modal, contrapuntal practice. And note that 6/5/3 sonorities *do* exist in the diatonic system – meaning that on this ground they *should* be acceptable in first species counterpoint exercises.)

following tone, and they must both be metrically unaccented relative to these consonant tones, so that their dissonance is not metrically emphasized as well. Suspensions on the other hand must be metrically accented relative to the consonant tones that precede and follow them – which is why they normally appear on downbeats, and which is why they are also a form of rhythmic figuration as we saw earlier. But they too must resolve by step (specifically *down* by step) to the following tone, as was the case with passing and neighboring tones, and they must be held over from the preceding beat as well, as we saw was the case with the soprano G4 in level A of Example 1.1-7, when it was held over on top of the final tonic triad to create a 9-8 suspension in level B.

In the fifth and final species of strict counterpoint, the three forms of dissonant voice leading introduced in the earlier species are all mixed together in the contrapuntal voice, to create the semblance of an actual melody. So, here we already see the progression from abstract voice-leading models in the earlier species to a more free contrapuntal texture, with its mixed note values, in fifth species counterpoint. Schenker noted this too, and proposed quite insightfully that the species-based model of counterpoint could be taken as a model for how actual tonal phrases are generated from an abstract *Ursatz*. In fact, this is what happens in Example 1.1-7, where the different levels of the image parallel the progression from first to fifth species counterpoint. We know that the *Ursatz* is an arrhythmic structure, represented in whole notes, and made up consonant root position triads (specifically I and V chords) – just as one would expect of the sonorities in a first species counterpoint exercise. In subsequent levels, more notes were added to this *Ursatz* form, and which specifically realized passing, neighboring and suspension voice-leading operations, which is exactly what happens in the second through fourth species models of counterpoint. And since these different motions are all mixed together in the actual aria phrase derived in level E, that level resembles a fifth, mixed species contrapuntal model too.

In this manner, we see how understanding the five species of counterpoint can help us understand further how even common-practice tonal passages, like the one described in Example 1.1-7, can be derived from an abstract *Ursatz* form. There are, however, important reasons for why understanding the five species of

counterpoint is not *sufficient* for understanding the structure of common-practice tonal pieces. Part of this has to do with the fact that Fux conceived of the species model as a pedagogical tool for teaching Palestrina-style contrapuntal composition, which means that a first-species counterpoint exercise is an actual, model composition, in which the whole notes represent actual durations. The *Ursatz*, on the other hand, is not an actual composition, as we have seen, which also accords with Schenker's conviction that the study of the five species of counterpoint should be aimed at helping the student understand how the works of the masters are structured (as we attempted to do in Example 1.1-7 too), rather than helping the student compose directly. This is a conviction that made Schenker critical of Fux, despite his admiration for the latter's model of species counterpoint (Schenker (1987) Book 2: 2, see also Federhofer (1982)).

More importantly, and as mentioned before, the principles of *harmony* also play a role in how tonal pieces are generated – and which can be seen in the very I – V – I harmonic structure of the *Ursatz* from which the *Casta diva* passage was generated in Example 1.1-7. This harmonic foundation allows the contrapuntal aspects of tonal pieces to *break* with strict contrapuntal norms in crucial ways, meaning that Fux's species-based proposals about strict counterpoint – being based on the norms of 16th century contrapuntal practice, an idiom that is not generally considered as having a basis in functional harmonic progressions like I – V – I – is often unable to account for how the specific structures of tonal pieces arise. Schenker was critical of Fux for this reason too, and spent much of his later writings in exploring how tonal composition crosses the 'bridge' from the strict principles of species-based composition to the more relaxed principles of 'free' composition – hence the title of his famous text "*Der freie Satz*". Beginning with the final part of the final volume of his *Kontrapunkt* text, titled "Bridges to Free Composition", and concluding with *Der freie Satz* itself, much of this involved Schenker developing Fux's ideas into a theory of *combined* species counterpoint, i.e. a contrapuntal phenomenon in which two or more different species of counterpoint (such as a second species counterpoint in one voice, and a fourth species counterpoint in another voice) are set simultaneously against a *cantus firmus*. In order to make such combined species structures workable, some of the rules of strict counterpoint need to be relaxed – and

this allows some of the contrapuntal peculiarities of common-practice tonal-harmonic music, i.e. of free composition, to emerge.

It should be noted that the precise description of how the bridge to free composition is crossed, through the more relaxed principles of combined species counterpoint, is arguably the most complicated part of Schenkerian theory, and certainly the most technically challenging. For this reason, it continues to be the focus of intense scrutiny, explication, and debate within the music theory community – some going so far as to say that this part of Schenkerian theory fails to account for certain aspects of tonal structure (as Eytan Agmon (1997) does, concerning the origin of seventh chords in tonal pieces). However, Schenker’s basic idea that a more sophisticated approach to counterpoint, and its interaction with harmony, can explain the intricacies of free composition, still remains one of his more revolutionary contributions to the understanding of tonal structure.

We can see this in how a Schenkerian approach can explain some aspects of the *Casta diva* passage in Example 1.1-7 that a more traditional Fuxian contrapuntal approach cannot, because of the way this passage violates some of the principles of strict counterpoint. For example, we have already discussed how the soprano A4 *Kopfton* in level A of Example 1.1-7 is held over on top of the following V chord in the form of a suspension in level B. As a suspension, the A4 must then resolve down by step to G4, which it does in level B too, to create the “6-5” suspension figure. But we also saw that from level C onwards, notes are introduced in between the A4 and the G4, so that the A4 cannot directly resolve down to the G4 anymore. Therefore, the elaboration of A4 by these intermediate notes (specifically the turn figure we discussed earlier) *breaks* with the rules of strict species counterpoint – and this is what Schenker refers to as an example of “free composition”. Importantly though, the A4 *does* resolve to G4 in the deeper levels of the phrase’s structure, i.e. in levels A and B – so, the free treatment of the A4 in the higher levels of C through E is still dependent on the A4’s strict treatment at the deeper levels of structure from which this phrase is derived. So, even free composition is governed by the grammatical structure of the *Ursatz* – in other words, what happens at the free musical surface is governed by what happens in the strict deep structure of a piece. Which again reveals Schenker’s hierarchical, generative view of tonal

structure – and also his computational view of this structure, given *how* the musical surface is governed by deep structure, i.e. through voice-leading operations that essentially ‘rewrite’ the pitch structure of a deeper level to generate the pitch structure of a shallower one.

All of the above points clearly suggest that Schenker developed a theoretical system that is generative in orientation, and which focuses on the abstract computational form of music (i.e. what we have been calling C_{HM}) – and which also has a strikingly Minimalist flavor – despite Schenker’s not being explicitly interested in developing such a theory. And the story does not end here, since there are at least two more Schenkerian proposals that seem to justify this Minimalist interpretation of his theory. I will briefly review these, before moving on to a discussion of some of the problems inherent in such an interpretation.

The first proposal has to do with Schenker’s notion of *prolongation*. There seem to be at least two definitions of this term. A more common, albeit possibly inaccurate, usage of the term takes prolongation to be the elaboration of a harmonic structure through voice leading – which is what the suspension figures in Example 1.1-7 do to the initial tonic and dominant harmonies of the *Ursatz* in level A of the example. Schenker seems to have referred to such an elaboration of harmony as an “*Auskomponierung*” of that harmony (“composing out”, Schenker (1979): 11-12), but this is what prolongation has come to mean in common, particularly North American, usage.⁶⁵ (For this reason, I shall use prolongation in this sense from time to time too.) Prolongation in this sense is hierarchical, since it involves the elaboration of a hierarchically-superior pitch structure with a hierarchically-inferior one. And as we saw in the case of the passage from Mozart’s *Sinfonia Concertante* in Examples 1.1-2 and 1.1-3, such prolongation can be *recursive* as well. (In those examples, a cadential 6-4 sonority prolongs (i.e. composes-out) a dominant triad through passing motion, i.e. second species-based voice leading – which is why I said there that the cadential 6-4 is *embedded* in the larger harmonic structure it prolongs, in the way in which “that Jürgen

⁶⁵ For example, see William Drabkin, “Prolongation”. In *Grove Music Online*. Oxford University Press. Published online August 2, 2011 at <http://www.oxfordmusiconline.com/subscriber/article/grove/music/22408>. Accessed March 6, 2013.

read a book” is embedded in the larger structure “Kwame said that Jürgen read a book”, which the embedded clause elaborates.)

The other sense of “prolongation”, which is arguably closer to what Schenker meant by the term, takes this term to mean the application of the principles that govern the background of a musical structure to the middle- and foreground levels of that structure too (Schenker (1987) Book 1: 241). This is essentially a fancy way of saying that the (strict) principles of voice leading that govern the background govern even the (free) surface of a musical passage, in the manner we just discussed in the *Casta diva* phrase. This sense of prolongation is explicitly recursive, since it implies that, for example, the same second-species based passing voice-leading operation that composes out part of the background can be seen as what governs the composing out of some part of the foreground too. In other words, the same background passing motion involving scale degrees 3 – 2 – 1, which we see in the structure of a 3-line *Urlinie*, can be the basis for a more foreground passing motion *within* this background passing motion, like so: (3 – 2 – 1) – 2 – 1. (In this instance, the initial scale degree 3 is *itself* elaborated by a 3 – 2 – 1 foreground voice-leading operation, represented by the parenthetical (3 – 2 – 1), which is clearly a recursive phenomenon.)

Which definition of “prolongation” is correct does not matter here – what matters is that *both* definitions involve a hierarchical, and specifically recursive, understanding of tonal structure. This again reveals Schenker’s essentially computational perspective on tonality – but more importantly, this reveals Schenker’s (implicit) concern with what we now know to be the *species-specific* computational form of music, given the uniqueness of the phenomenon of recursion to the human mind, and by extension to *human nature*.

The second Schenkerian proposal that seems to justify a Minimalist interpretation of his theory has to do with the very origins of the *Ursatz*. We know that the *Ursatz* is of fundamental importance in Schenker’s view of the form of music. But where does the *Ursatz* itself come from? The answer to this is where we see Schenker at his most naturalistic – again despite his primary, explicit goal not necessarily being one of developing a natural science of music. For Schenker, the *Ursatz* arises by composing out

Der Naturklang (“the chord of nature”, Schenker (1979): 10-11, 25). The chord of nature is nothing but the harmonic series of pitches that exists above a given fundamental pitch. If we take the first five of these pitches, we get the major triad – and so Schenker believed that major-minor tonality, i.e. the language of the Western Classical common-practice, could be grounded in natural laws of acoustics (Schenker (1973): 20-30). Schenker belonged to a long tradition of music theorists who believed this proposition, going back to at least the 18th century French theorist Jean-Philippe Rameau in his *Traité de l'harmonie* (Rameau (1971)), and possibly even the 16th century Italian theorist Gioseffo Zarlino (e.g. see Zarlino (1968): 6-10). The *Ursatz* in particular seemed to enact such a law according to Schenker in the way its *Urlinie* composes out one of the intervals between the root of a triad and its other members – viz. the root and the third of the triad (i.e. scale degree 3 – 1, as in the 3-line *Urlinie*), the root and the fifth of the triad (i.e. scale degree 5 – 1, as in the 5-line *Urlinie*), and the root and the root itself, an octave higher (i.e. scale degree 8 – 1, as in the 8-line *Urlinie*). Schenker believed this composing out of an acoustically-derived triad to happen in the bass of the *Ursatz* too, in the way it arpeggiates the important interval of the fifth of a triad, in the I – V – I bass motion of the *Ursatz* (hence the name “bass arpeggiation” for the bass voice of the *Ursatz* too).

Despite its naturalistic orientation, a description of the *Ursatz* in acoustic terms is seriously flawed. For one, it does not explain the origin of the minor triad, since the minor third of such a triad (e.g. the F in a D-minor triad) is not among the first five pitches of the harmonic series above a fundamental. Consequently, and rather arbitrarily, Schenker was forced to relegate minor mode tonality to a secondary status – since he could not explain it in “chord of nature” terms. Also, Schenker did not even engage with the problem of why the interval of a perfect fourth, which has a simpler frequency ratio than a major third (i.e. 4:3 vs 5:4), is still considered dissonant relative to the major third. This has led some scientifically-inclined modern theorists to reject the “chord of nature” explanation for tonality, in favor of one in which the *Ursatz* is just taken to be a primitive in an axiomatic tonal system.⁶⁶

⁶⁶ The best example of this can be found in the work of Michael Kassler (1967, 1977). The Schenkerian theorist Matthew Brown has also defended Kassler’s axiomatic model (Brown and Dempster (1989): 88, Brown (2005):

In my opinion, however, even this solution does not work, leaving the naturalistic status of the *Ursatz* still up for grabs – and which brings us to the discussion of some of the problems inherent in a scientific, and specifically Minimalist, interpretation of Schenkerian theory. First of all, consider that the *Ursatz* is meant to be the abstract form of a finite set of tonal structures – i.e. those present in the works of a small group of master composers. This means that it is clearly not the abstract form of a general computational system of music, which gives rise to a variety of actual musical phrases *across* idioms, in the way C_{HL} does for language.⁶⁷ However, this is not a problem in and of itself, since the *Ursatz* could just be a primitive in an axiomatic description of *Western Classical tonal music*, i.e. the first step in a broader description of tonality *per se* that could emerge in years to come. (Of interest in this regard is the fact that before he presented his ideas as a universal generative grammar of language, Noam Chomsky’s earliest writings were just about the grammatical structure of sentences in English (and upto a certain extent Hebrew) – i.e. the languages he knew best, as a native speaker (or near-native in the case of Hebrew), and could therefore theorize about most convincingly, just as Schenker theorized, quite appositely, about the music he knew best, as a ‘native speaker’ of the Western Classical tonal idiom. The moral being that even a general scientific theory has to start somewhere, usually somewhere specific and familiar to the theorist.)

The problem arises though, when one considers Western Classical tonality to be a closed system, and the study of this system to be an end in itself, because this runs the risk of incorrectly treating a limited, culturally-circumscribed system as a natural system instead. In turn, this can lead to the *Ursatz* being considered a natural object – a primitive in an axiomatic system – when it is really just a cultural artifact. A case in point here is an assertion made by the music theorist Matthew Brown, who is well-known for being one of the few Schenkerians to take on the task of interpreting Schenkerian theory in scientific terms, which is a goal shared by this dissertation. However, Brown says, rather problematically,

211-214), although he argues that the empirical foundations of this model need to be clarified. Brown’s own approach has been to suggest that structures like the *Ursatz* summarize certain general laws of tonal motion, which connects them empirically with a body of data (personal communication).

⁶⁷ Put in terms of some linguistic concepts we have already explored, one could say that the *Ursatz* allows Schenkerian theory to have *observational adequacy* in describing the common-practice masterworks they deal with, but not *descriptive* (let alone *explanatory*) adequacy as a theory of musical grammar. See DeBellis (2010): 111-112 in this regard.

that “if we treat tonality as a property of some specific culture and time period, then we have taken the first step down the psychological route mentioned earlier” (Brown (2005): 214-215).

But this cannot be the case – if we treat tonality as a property of some specific culture, then we will end up going down a specifically cultural, and not psychological, route. If we are to travel down a psychological route, and a naturalistic one at that, this will require treating tonality as a *species-specific* property, not a culture-specific one – as I have been arguing since the beginning of this dissertation. Relevant to all of this is also the status of the *Ursatz* as a primitive in an axiomatic system. If the axiomatic system under consideration is a culture-specific one, then the *Ursatz* could be a primitive in that system.

However, this works if such a system is really a stepping stone to a more general, *cross-cultural* axiomatic system, as I just said – and there is no reason why the *Ursatz* should be a primitive in such a system. Remember that such a system would be essentially a universal, generative grammar of music, and to that extent, as per Minimalist criteria, it should only propose components for the system that are conceptually necessary. The *Ursatz* might be conceptually necessary for a generative theory of Western tonal structure (although I will contest even this in a bit), but it is certainly not conceptually necessary for *all* music – which means that in a generative theory of music, the *Ursatz* should be the *product* of simpler, conceptually necessary entities and procedures. Going by Minimalist proposals in this regard, these conceptually necessary components would be some sort of lexicon, and a Merge-based procedure for combining lexical items into more complex structures. In this light, the *Ursatz* should not be a primitive, but rather the product of *Merge combining musical ‘lexical’ items*.

If we take chords to be like lexical items for music, which is a proposal I will defend in the next chapter, then we can understand the *Ursatz* as itself being the product of certain chords being merged – specifically the tonic and dominant harmonies that we know constitute it. Allan Keiler in fact makes such a proposal, when he conceives of the *Ursatz* as essentially a binary-branching set, made by merging the initial and final tonic triads of the *Ursatz*. I will have reason to explore Keiler’s proposal in more depth in

the next chapter, when I defend my above proposal about chords constituting the musical lexicon. (This also demonstrates why Keiler's ideas are such an important influence on this dissertation.)

The above suggests why the *Ursatz* cannot be the primitive in an axiomatic tonal system, but is really just an artifact in a cultural system. This point becomes even more evident when we consider the fact that in his mature theory, Schenker considered the *Ursatz* to be the abstract form for entire pieces of music, such as symphonic movements several hundred measures long. This makes it plausible that Schenker did not even conceive of the *Ursatz* as the fundamental structure from which surface *grammatical* structures are generated, in a psychological system, but conceived it instead as the structure from which artworks are generated, in a *poetic* system. In fact, this (i.e. the psychological implausibility of Schenker's mature *Ursatz*) is one of the reasons why Lerdahl and Jackendoff decided to ultimately distance themselves from Schenkerian ideas in their work in musical grammar (Lerdahl (2009): 187-188).

There are other issues that problematize a Minimalist interpretation of Schenkerian theory. For example, I described the ordering of the levels in the Schenkerian derivation of a phrase to be logical, in that they express the relationships between the hierarchically-superior and inferior constituents of a derived musical phrase. This is similar to what happens in the derivation of a linguistic grammatical structure in generative linguistics – the derivation expresses the relationships between the constituents of the structure, rather than the series of steps (in an algorithm) for how the structure is to be generated in real time. However, the ordering of the levels of a Schenkerian derivation could be the steps in a generative *process* through which a musical surface is derived in real time – i.e. the ordering of the levels could be chronological rather than logical. Some Schenkerians seem to take such a process-oriented approach to Schenkerian theory (e.g. Beach (1985): 294), and this certainly seems to be the position Schenker took himself earlier in his life (e.g., see Keiler (1989): 288) – but there is no reason to believe that this was the position Schenker subscribed to in his mature *Neue musikalische Theorien und Phantasien* years too.⁶⁸

⁶⁸ In fact, there is reason to believe that Schenker subscribed to the logical, as opposed to the chronological, perspective being advocated here, when he says, "I would not presume to say how inspiration comes upon the

Another problem with a Minimalist interpretation of Schenkerian theory lies in the transition from strict to free composition in the generation of a musical surface. It is quite evident that Schenker believed that free composition continues the prolongation of the *Ursatz* at the musical surface that was begun strictly, at deeper levels of structure. However, Kofi Agawu says:

“By declining to specify the full range of historically specific stylistic resources that enable the generation of a given composition from a generalized background to a unique foreground, or by consigning such tasks to a less urgent category in the hierarchy of theoretical concerns, he evaded one of the challenging issues in understanding musical style.” (Agawu (2008b): 111)

Finally, the internalism of Schenker’s position renders problematic the role of those musical structures within Schenkerian theory that are not directly subsumed under (pitch-) grammar, e.g. those pertaining to meter and rhythm. As Fred Lerdahl says with regard to his attempt to model musical grammar:

“The non-rhythmic character of the *Ursatz* presented a formal and musical problem. How was rhythm to be introduced into the derivation, and why should it have inferior status? ... [Therefore] We could not build a rule system to assign a hierarchy of events without first developing a theory of rhythm.” (Lerdahl (2009): 188-189)

I believe that all of these objections have straightforward answers, as a result of which I will continue to defend a Minimalist interpretation of Schenkerian theory. For example, the reason for considering the steps of a derivation, linguistic or musical, to be logical rather than chronological lies in the fact that we do not yet have a clear understanding of how the mind generates grammatical structures in real time, and specifically how such processes are implemented in the hardware of the brain. Furthermore, if the steps of a derivation represent an actual generative process, then they must be the process by which composers create actual pieces of music, and listeners uncover the structure of those pieces – assuming that the generative process is a psychological one. But how do we know this? How do we know what was going on in Beethoven’s mind when he came up with the first movement of the Eroica symphony, or in Bach’s mind when he came up with the Chaconne in the D-minor Partita for solo violin? The inability to answer

genius, to declare with any certainty which part of the middleground or foreground first presents itself to his imagination: the ultimate secrets will always remain inaccessible to us” (Schenker (1979): 9).

questions like this renders the chronological approach to musical derivation hopelessly speculative, and subject to committing the intentional fallacy.

Kofi Agawu's point about the inability of Schenkerian theory to specify particular, historically-situated musical surfaces has more teeth. However, one can ask why this should even be the task of a generative theory of musical structure? Generative theories specify the knowledge required for musical or linguistic *competence* – i.e. the knowledge that allows the mind to generate surfaces that can be interpreted by the conceptual-intentional and sensorimotor systems. But these systems then interpret such surfaces in manifold ways, not governed by the grammatical system itself since the grammatical system only generates structures – it does not interpret them too. The way in which the external systems of the mind interpret a given surface could involve a consideration of how the surface happens to realize a certain stylistic feature, which in turn depends on our knowledge of a whole network of pragmatic considerations that determine the elements of style, i.e. that determine how music is *used* in certain historical or cultural contexts. But none of this is part of grammatical knowledge, which only generates structures for the historically- and culturally-situated parts of the mind to interpret – i.e. all of this lies in the realm of musical *performance*, and beyond the scope of a generative theory of musical structure. Of course, if the difference between two styles of music is parametric (in the Principles & Parameters sense of the term) – i.e. if the two styles are actually different musical *idioms* (or I-musics) – then the specification of how an *idiomatic* surface is generated from the musical background should be part of a generative theory, within its specification of the grammatical parameters of those idioms. But anything having to do with the discursive or communicative functions of music is not what a generative theory of musical structure should have to deal with.

Finally, Fred Lerdahl's points about the arrhythmic aspects of Schenkerian theory make sense too, but really only from an externalist position, i.e. one that asserts the independent role of pitch and rhythm in musical phrase generation. This is opposed to the internalist position in which rhythm is mapped to a purely pitch-based generative procedure, at something akin to a PF level of musical surface structure. (Implying that musical rhythm is part of a musical *phonology*, in the way speech rhythm is part

of linguistic phonological structure – and should therefore be understood in the context of a mapping between grammar and phonology, as generative linguistics does.)

The externalist position subscribed to by Lerdahl, however, seems to be more widely accepted in music scholarship. For example, Simha Arom (2001) talks about the cognitive models behind certain Central African polyrhythmic structures, which are not necessarily tied to any understanding of the pitch aspects of these structures. Jeff Pressing (1983) talks about similar cognitive foundations that he believes underlie, and are shared between, rhythmic structures in certain West African and Balkan idioms. Some scholars have also referenced the cultural functions of rhythmic structures, such as in dancing, in helping shape these structures, which is clearly a pitch-grammar external approach to these structures. Perhaps not surprisingly, given his generally functionalist orientation, this is something John Blacking does in his discussion of African polyrhythm (Blacking (1973): 74-75), and this is a position Kofi Agawu reaffirms in his discussion of West and Central African time lines, when he says that “the key to understanding the structure of a given *topos* [i.e. a paradigmatic, culturally-articulated rhythmic figure, such as a time line] is the dance or choreography upon which it is based” (Agawu (2003): 73-74).

The above externalist attitude in ethnomusicology can also be seen in the worlds of music psychology and neuroscience. For example, some scholars have proposed separate mental modules, with disjunct associated brain areas, for pitch and rhythmic information processing (Peretz and Coltheart (2003)), while others have asserted that how we judge melodic phrase structure depends on *independent* pitch and rhythmic information processing (Palmer and Krumhansl (1987)). Carolyn Drake and Daisy Bertrand have identified universals in rhythmic information processing that do not necessarily depend on pitch structure (Drake and Bertrand (2003)), and Tecumseh Fitch argues that the evolution of rhythm in human music might share something with ‘drumming’ behaviors in certain great apes, such as chest-beating in gorillas, drumming on tree buttresses by chimpanzees etc. (Fitch (2006): 194-195) – even though these apes might not have (the specifically recursive) pitch-processing abilities of humans. Approaching the issue from a slightly different perspective, Aniruddh Patel and Joseph Daniele have

argued that rhythmic structure in music is influenced by the structure of *speech* rhythm in certain idioms, irrespective of the role of pitch in all of this (Patel and Daniele (2003)).

Now rhythm clearly constitutes a separate system in itself, for no other reason than the fact that it deals with a different kind of information than pitch does, viz. durational information. Moreover, this separation is clearly evident in musical idioms that are percussive in nature, such as the various African drumming traditions explored by ethnomusicologists like Simha Arom and John Blacking. But the relevant issue here is not whether rhythm is a separate system, but whether the generation of rhythmically-articulated musical surfaces happens independently of pitch processing (as an externalist might argue), or whether it is somehow governed by, or mapped to, pitch processing too (which is the internalist position on this matter). Fred Lerdahl seems to be arguing for the externalist position when he rejects the possibility of describing the hierarchical structure of a musical surface in pitch-based terms *without* an explicit theory of rhythm to inform this.

As I will argue in chapter 2.2, I believe that pitch structure does play an important role in the generation of rhythmic surfaces, which I believe is implicit in Schenkerian internalistic approaches to this issue. Moreover, ignoring pitch in the generation of rhythmically-articulated surfaces, even in percussive musical idioms, runs the risk of misunderstanding these idioms in crucial ways. Kofi Agawu has specifically warned against this in the context of various African drumming traditions, where ignoring pitch often amounts to ‘inventing’ a false view of African music as being intrinsically rhythmic in nature – and therefore inherently different from other, more explicitly pitch-based, idioms, like Western common-practice tonal music (Agawu (2003): 65-66). Finally, there seems to be evidence that suggests a deeper role for pitch in rhythmic information processing than the various above approaches from within the cognitive, evolutionary, and neurosciences seem to acknowledge, possibly because of the resiliently externalist attitude many of these approaches have taken to the study of rhythm. For example, it is widely accepted that the ability to perceive rhythmic structure depends on our ability to *entrain* to a given beat. As Stephen Brown and his colleagues have said, “what is special about humans is not only their capacity to move rhythmically but their ability to *entrain* their movements to an external timekeeper, such as a

beating drum” (Brown, Merker and Wallin (2001): 12). However, Adena Schachner and her colleagues have recently discovered that the ability to entrain to a beat seems to occur only in species that are capable of *vocal mimicry*, i.e. those species that engage in specifically the kind of pitch-based behavior that we know lies at the basis of music and language – which suggests that the ability to entrain to a beat (and therefore comprehend rhythmic structure) is a *by-product* of a more basic ability to communicate vocally (Schachner et al. (2009)).

In sum, I believe that the above considerations justify an internalist, Minimalist interpretation of Schenkerian theory, and the use of this paradigm to develop a Minimalist model of generative musical grammar. However, as I said a little while ago, the Minimalist aspects of Schenkerian theory seem to be *implicit* in much of what Schenker and several later Schenker-influenced theorists have said – they are by no means explicit, at least not all of the time, in Schenker’s writings. Which means that I am happy to concede that the Minimalist interpretation of Schenkerian theory is exactly that – an *interpretation* of Schenkerian theory (albeit a fair one, in my opinion), as opposed to a historically-accurate *description* of Schenker’s ideas. This does not imply that Schenkerian theory has *no* scientific basis as a theory of musical grammar – just that the implicitly scientific aspects of it have to be made explicit, and then supplemented with other ideas if necessary, to yield a genuine grammatical theory. This dissertation aims to do that, but also concedes that an interpretation of Schenkerian theory as more of a poetic system, meant to describe the culturally-circumscribed structure of a finite set of artworks, is equally legitimate – despite the various methodological and philosophical problems this might lead to, some of which I have discussed above.

I also believe that there are some aspects of Schenker’s theory that might never have a scientific explanation, and will always remain on the unsystematic side of things. Due to this, this dissertation does not try to reconstruct a genuine scientific theory from Schenker’s writings as he wrote them, as some scientifically-inclined Schenkerians have attempted to do (e.g. Brown (2005)), and is happy to accept this current project as a *neo-Schenkerian* one, if that is more plausible historically. To this extent, I agree in

part with Leslie Blasius, who, in describing the North American appropriation of Schenkerian theory in the latter half of the twentieth century, writes:

“While we find it easy to grant at a distance Schenker’s experiential claims (and indeed much has been made of the resemblance between Schenker’s stratification and the mental models of language proposed by Chomsky), his articulation of these claims in *Free Composition* is so opaque, so arrogantly transcendental as to discourage us from discerning any trace of a psychological argument.” (Blasius (1996): 33, see also his footnote 39)

But I also think that this statement is too strong. I have already suggested how there seems to be more than just a “trace of a psychological argument” in Schenker’s ideas, and I have suggested that there is a rich intellectual tradition that justifies these ideas, and within which these ideas should be understood.

For example, Schenker’s conviction that musical understanding is a matter of intuition, rather than overt instruction, already shows a Rationalist intellectual orientation consistent with an internalist approach to the musical mind. “Intuition” is how the German word *Anschauung* is often translated, a word that is, of course, of singular importance in Kantian epistemology. Kant used it to describe the *a priori* conditions of space and time without which sensory experience would be impossible – we can only experience, say, a certain physical object as *that* object if we recognize it as existing in time, and as having a certain spatial form. But more importantly, “intuition” can be taken as referring to a kind of knowledge that is acquired without experience, or more importantly, *prior* to experience because it is what *informs* experience to begin with. So, by describing the musical knowledge of the master composers as “intuitive”, Schenker was asserting that one has to have knowledge of a musical surface’s structure prior to experiencing it (by being a genius in Schenker’s opinion), because one cannot derive this knowledge *from* the experience of a surface, i.e. by listening to music in the way ordinary people do – especially since it is this knowledge that provides the conditions for ‘parsing’ that surface to begin with.

Once we ignore Schenker’s elitist application of the above ideas to just the great composers, it becomes clear that the above is just a fancy way of saying that an understanding of the (particularly recursive) grammatical structure of a tonal piece is not something we can derive from listening to it, but is something we must know innately instead. That is, we cannot know that a surface was generated from a

certain abstract *Ursatz* just by hearing the surface, i.e. the final, generated product, but must already know, intuitively, that the surface has a specific form accorded to it by the *Ursatz* from which it was generated, since without this knowledge we cannot even make sense of a perceived surface – i.e. our innate knowledge of the *Ursatz* and the grammatical system is what informs our perception of a surface to begin with. This is of course just the generative approach to knowledge of language too. So, in invoking intuitions in his description of musical understanding, Schenker is asserting a philosophical attitude that is shared with generative theory, but which also falls in line with Rationalist, and particularly Kantian, attitudes towards knowledge and experience – a connection that Kevin Korsyn has previously affirmed, through his systematic exploration of the topic (in Korsyn (1988)).

Moreover, Schenker's use of terms like *Ursatz* and *Umlinie* clearly suggest an awareness of the Rationalist philosophical interest in organic form, epitomized by Goethe's notion of the *Urphänomen* or "abstract phenomenon". In the last section, I talked about Minimalism being a research program in biolinguistics, i.e. a research program that attempts to describe and explain the biological foundations of the human language faculty. But we also saw then that Minimalism's interest in language's biological foundations involves particularly an interest in its biological form – i.e. its *morphology* – as opposed to its biological functions, for example in natural selection. In this goal, Minimalists have been influenced by thinkers who focused on issues of morphology, such as D'Arcy Thompson – and it is here that Goethe's influence on Minimalism can be seen too.

Goethe was particularly interested in organic phenomena, especially in the fact that they grow and metamorphose, because he felt that the prevalent Newtonian science, with its fascination with the material world, was unable to account for them. Now Newton did not have an inherently materialistic worldview, as we have discussed before, since he attempted to explain how the universe works in terms of rather ethereal forces instead. But these attempts certainly did not pay as much attention to the *organicism* of phenomena compared to their inorganic properties – after all, Newton was more interested in why the apple falls to the ground (as opposed to moving up, out into space) than he was in how and why the apple is born from a seed, grows into maturity, decays, and then dies. This latter issue was

therefore what caught Goethe's interest. With his concept of the *Urphänomen*, he tried to explain organic phenomena in terms of abstract forms, in the sense that there is an "abstract phenomenon" that governs the actual forms that organic phenomena end up having – so that there is an "abstract plant" (i.e. the *Urpflanze*) to which the form of actual plants 'conform', there is an "abstract animal" (the *Urtier*) to which the forms of actual animals conform and so on. In other words, Goethe's proposals were an attempt to account for organic substances in terms of their form rather than their function – which is similar to D'Arcy Thompson's proposals about sunflower morphogenesis.

However, unlike Thompson, Goethe's proposals were not mathematical or even computational, since his description of these abstract forms to which actual substances conform was not in terms of certain geometrical shapes or mathematical growth functions that are fulfilled during development. Instead, his description of the *Urphänomen* was much more intellectual, in keeping with his Rationalist orientation – i.e., the *Urphänomen* was something that had to be *intuited* (Goethe (1891): 121).⁶⁹ Even more strikingly, Goethe realized that this abstract form gave rise to a process of growth through which actual organic forms are *generated*, so that the locus of growth, say, of a plant, is not to be found in its root or stem, but in its *node* ("Knoten"), which Goethe realized, ahead of his time, allows one to conceive of plant species in generative terms (Goethe (1891): 225).⁷⁰

The idea that an actual organic substance can result from a generative process conditioned by an abstract form, which we see as being the essence of Goethe's theory of biological (and particularly plant)

⁶⁹ Goethe's original German text reads: "Wie sie sich nun unter einen Begriff sammeln lassen, so wurde mir nach und nach klar und klärer, daß die *Anschauung* [my emphasis] noch auf eine höhere Weise belebt werden könnte: eine Forderung, die mir damals unter der sinnlichen Form einer übersinnlichen Urpflanze vorschwebte. Ich ging allen Gestalten, wie sie mir vorkamen, in ihren Veränderungen nach, und so leuchtete mir am letzten Ziel meiner Reise, in Sicilien, die ursprüngliche Identität aller Pflanzentheile vollkommen ein, und ich suchte diese nunmehr überall zu verfolgen und wieder gewahr zu werden."

⁷⁰ Again, from the original German: "Die höhern Organe der Pflanzen darf er nicht von Wurzel und Stengel, sondern einzig und allein aus dem Knoten ableiten, aus dem auch Wurzel und Stengel erst geworden. Die ganze Pflanze darf er nicht als Object der Anschauung so gerade zu für ein Individuum nehmen, sondern nachforschen, wie dieselbe durch allmähliche Reihung eines Knoten an den andern, deren jeder das Vermögen hat unter Umständen selbstständig zu vegetiren, zu der Gesamtförmung gelangte. Daraus geht dann *ein bestimmter genetischer Begriff der Species im Pflanzenreich* [my emphasis], welchen viele beinahe aufgegeben, weil sie ihn auf anderm Wege vergebens gesucht, gleichsam von selbst hervor; und die Kritik der in unserer Zeit so oft behaupteten und bestrittenen Verwandlungen einer Pflanze in die andere, welche der Naturforscher, ohne aller Gewißheit zu entsagen, nicht einräumen darf, gewinnt wieder einen festen Boden."

morphogenesis, is clearly present in Schenker's notion of an actual musical structure being generated from an abstract *Ursatz*. This makes Schenker's theory clearly an *organicist* theory of musical structure – i.e. a theory of musical structure within a larger Rationalist science of organic form. But “organicism” is a term that is not only well known in academic musical circles, particularly as a term with strong Schenkerian associations, it also appears as a heavily fraught term in academic musical discourse. In general, “organicism” is equated with the notion of *unity*, and particularly with Schenker's idea that tonal masterworks by the great composers show a unity in their structure, as a result of their generation from a simpler *Ursatz* (e.g. see Solie (1980): 148). So far this is an unproblematic characterization of Schenkerian thought, but things get complicated when these terms are used in a hermeneutic context, i.e. as ways of interpreting musical artworks. In this context, “organicism” can be (and is often) taken as a prescription for how one *should* hear (in an interpretive sense, that is) a piece of music – with their being an associated moral that one who cannot hear a tonal masterwork as unified is somehow musically incompetent. (Or at least critics of Schenkerian theory often allege that this is what Schenkerians believe, and what Schenkerian organicism implies, e.g. see Russ (1993).) Worse still, this hermeneutic use of the term can be taken to imply that a piece of music that is not unified in some way is deficient. Given his elitism and traditionalism, Schenker certainly believed that musical pieces not written by the common-practice masters *are* deficient, and deficient precisely because they do not realize the organic possibilities afforded by the *Ursatz*. (For example, see his criticism of Wagner and Stravinsky in this regard, in Schenker (1979): 106, and Schenker (1996): 18.) But as I said earlier, Schenker's value judgments can be divorced from his profound insights into musical structure, and this is certainly what most of his followers have done too.

To this extent, the hermeneutic, as opposed to scientific, use of ideas associated with Schenkerian theory, such as organicism, is quite unfair, since it gives critics of Schenkerianism ammunition with which to attack the system as being elitist, dogmatic, and parochial, even though these attacks often ignore the scientific and technical foundations and implications of Schenkerian theory – which is where Schenker's ideas have arguably their greatest merit. But the hermeneutic treatment of Schenkerian theory

is popular and influential. Only a handful of scholars who have situated Schenker's ideas in their (in my opinion, proper) scientific context, which is a context that justifies their Minimalist interpretation too. For example, Jamie Kassler has described Schenker's organicism as exemplifying a theory of how creativity arises and evolves – in a biological sense, i.e. as an alternative to, say, Darwinian theories of evolution (Kassler (1983)). This is an important insight, given how Minimalism is a research program that explores the origin and development of creativity too, in the Cartesian sense of the term.

Even more important has been William Pastille's role in grounding Schenkerian theory in Goethe's ideas about organic form (e.g. in Pastille (1985)). In fact, Pastille has argued that Schenker explicitly believed in the Goethean idea that the musical work, akin to a biological organism, "is based on an inner model, that governs its external, individual characteristics" (Pastille (1990): 35), which is an idea that we know also underscores the biolinguistic aspects of the Minimalist Program. Pastille points out how at times Schenker even appropriated Goethe's terminology, e.g. when he describes in his *Kontrapunkt* text how a certain model of strict counterpoint represents the "*Urform* of all possible forms of dissonance in free composition which occur on the strong beat" (Pastille (1990): 36). Finally, the Goethean (and Rationalist scientific) connection can be seen in the famous inscription that Schenker provides at the beginning of *Der freie Satz* that not only cites a passage from a text by Goethe, but specifically one from Goethe's *Theory of Color* (Schenker (1979): 3). All of which leads Pastille to conclude that "Goethe would be pleased to know that through his scientific ideas he had had a hand in the creation of the first and most influential morphology of music" (Pastille (1990): 44).

But probably the scholar who has made the most forceful argument in favor of contextualizing Schenker's thoughts in the Romantic Rationalist tradition of Goethe, Kant and others is Allan Keiler, perhaps unsurprisingly, given Keiler's generative linguistics-inspired interpretation of Schenkerian theory. This can be seen from a debate that arose within the music theory community in the light of some of Schenker's earlier thoughts about organic form. In a famous essay written when he was 27, called "Der Geist der musikalischen Technik" (Schenker (1988)), Schenker makes some comments that led William Pastille to conclude that Schenker was *denying* the organicism of tonal music in his early years (Pastille

(1984): 31). Pastille argues that Schenker was initially disinclined to consider tonality as being organic because (1) he had not yet realized the role of harmony and counterpoint in generating unified surfaces from a common abstract background (which is something that would only begin with the *Harmony* treatise of 1906), and (2) Schenker still thought that each musical work was the result of the idiosyncratic will of its composer, rather than general, organic forces of tonal organization, the latter being a view he would come to only later when he realized that the greatest composers create music that reflects the “will of the tone” instead, of which they (and only they) have intuitive knowledge. (This latter point was crystallized into the title of one of Schenker’s more mature, organicist writings on music, the two volume *Der Tonwille* of 1921-24 (Schenker (2004, 2005)).)

But in an article published in 1989 (and whose title involves a play on the words of the title of John Blacking’s *magnum opus* “How Musical is Man”), Allan Keiler argues that Schenker’s apparent anti-organicism in the “Geist” essay was partly polemical, as a response to the ‘dry’ formalism of Eduard Hanslick (Keiler (1989): 289). Hanslick did see the nature of the musical work in formalist and organic terms, but in a way that was divorced from the compositional process – which is what Schenker was reacting to, since this was still the period in his life when Schenker was pursuing a career as a composer. This is why the younger Schenker might have considered a musical piece to be organic only if its organicism was conceived of by the composer of the piece, as opposed to being determined by natural forces of tonal organization. But as Keiler points out on the basis of his other writings in this period, Schenker clearly believed that:

“There is indeed music that is coherent, or that sounds coherent. He never explains how you can recognize such coherence from the music, but it is normal nonetheless, he says, to characterize such coherence as having arisen in a certain way. One then describes the music as having a logical beginning and end, a continuous sense of development, and so on, ideas that he claims are not inherent to the music but borrowed from logic and rhetoric. Schenker’s argument, therefore, throws away the very evidence for which it was created. There is no mistaking coherent from non-coherent music, at least for Schenker [i.e. even in this early phase of his musical thinking]. The terminology of rhetoric is simply used to distinguish the one from the other. It is, in other words, metalanguage, not musical (i.e. analytical) language, and as such it happens to come from rhetoric and logic. ...Perhaps [Schenker] felt more comfortable with this metalanguage the more he came to understand its causes, that is, the more he was able to make explicit the nature of musical content. Indeed, in his most mature work, once the specific musical content was worked out in the usual form of a series of analytic levels leading from the *Ursatz* to the surface,

Schenker would often paraphrase and elaborate the musical content in this very metalanguage.” (Keiler (1989): 290)

Based on this, Keiler goes on to say that:

“I think that any view that characterizes Schenker, during any part of his intellectual development, as fundamentally opposed to essential attributes of organic thought would have to appear questionable, if not downright odd. The evidence from the totality of his work is that he accepted unequivocally the German idealist tradition of his earliest education and background and knew in a fairly intimate way the works of Goethe, Kant, Hegel and Schiller and, of course, many others. Certainly it would be foolish to argue that, because it is only during his middle and later periods of work where the names of the great German masters come to be mentioned and quoted, it was only then that he came to know them and understand and acknowledge their determining influence. The fact is that he always knew them, and he could have always quoted them. What changed significantly during the course of Schenker’s work is that he came to have reasons to refer to them, that is, he could summon them up to provide support and understanding for his musical discoveries. Indeed, they could very well have helped him to see more clearly their implications ... It would not be out of place, in fact, to characterize the relationship of these views, a universal musical competence of individual faculties and the possibility of a potentially infinite variety of musical styles and cultures, as the relationship between a more limited and constraining background and the ever evolving foreground of musical styles. And although Schenker could not have used these concepts of background and foreground at the time, they underlie what would come to be seen as an essential strategy of Schenker’s thinking ... Schenker [was not] able, during this period, to tolerate a completely synchronic and formalist Organicism [only because] his attention was not yet turned to the problems of harmony and counterpoint, where purely musical discoveries would eventually lead him to find a naturally synchronic context for Organicist method, but [also] because the dominance of his thoroughgoing belief in the primacy of melody and of the fantasy and will of the individual composer kept the purely formal sphere of musical content to a large degree sealed off from participating in Schenker’s dominant Organicist impulses.” (Keiler (1989): 291-292)

We see from the above that Keiler clearly thinks that Schenker’s thinking was inherently organicist and firmly within the Romantic Rationalist tradition of Goethe, Kant and others from the outset.⁷¹ This makes it evident to my mind that it is this broader intellectual context in which Schenker’s ideas should be understood – which then repudiates Leslie Blasius’s earlier statement that there is no “trace of a psychological argument” in Schenker’s writings. I believe it is evident that Schenker’s ideas are deeply

⁷¹ Kevin Korsyn has attempted to defend William Pastille from Allan Keiler’s critique, thus re-affirming the idea that Schenker was indeed “anti-organicist” in his early thought. However, Korsyn grounds his argument in a problematic hermeneutic use of organicism, which he actually *distinguishes* from its scientific use. He says, for example, that “organicism is not a scientific doctrine, despite the proliferation of biological metaphors in organicist thought. The comparison of a work of art to a biological organism is not a reduction to a physical explanation; in the organicist appeal to nature, nature is not an impersonal mechanism as it is for modern science” (Korsyn (1993)). This shows a misunderstanding of science though, since not all science is physical science. Comparing a work of art to a biological organism is not a reduction to physical explanation, but it is to a *psychological* explanation instead, since physical explanation is inadequate when it comes to organic forms, if one accepts the Minimalist position on this. This, however, still affirms organicism as a scientific doctrine, unless one believes that psychological explanation is not scientific either.

psychological, and were influenced quite strongly (even if implicitly) by a long tradition of (Romantic Rationalist) reasoning about human thought, knowledge, and creativity.

Which brings me to one individual in the Romantic Rationalist tradition who I have not discussed much yet, but who is of immense importance to this dissertation – since he provides the strongest link between the generative paradigms of Schenkerian music theory and Chomskyan Minimalist linguistics, and is therefore the strongest piece of evidence for the identity of musical and linguistic theory.

Friedrich Wilhelm von Humboldt (1767-1835) was a Prussian philosopher and diplomat, and founder of the University of Berlin (in 1810, which, in 1949, was renamed the Humboldt University of Berlin in joint honor of him and his illustrious younger brother, Alexander von Humboldt). The older Humboldt is perhaps better remembered as a linguist though, given his particularly influential philology of the Basque language, and his study of the ancient Javanese language of Kawi, which resulted in his *magnum opus*, the *Über die Verschiedenheit des menschlichen Sprachbaus und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts* of 1836 (English translation, Humboldt (1999)).

Now, Wilhelm von Humboldt is not quite a household name in the musical community, just as Heinrich Schenker is virtually unknown to the wider world outside of musical scholarship. (In fact, the only references to Humboldt I have found in the music-theoretic literature, in the context of Schenkerian or generative music theory, are two, brief, citations, both of them in articles by – no surprises here – Allan Keiler (i.e. Keiler (1978a): 175-176 and Keiler (1989): 274).) The fact that Humboldt's ideas are so little known within the community of music scholars is a shame though, because his ideas belong strongly within the Romantic Rationalist tradition too, being particularly influenced by Goethe – and constitute a specifically *organicist theory of linguistic structure*. In fact, it would not be too far-fetched to suggest that Humboldt essentially developed a Schenkerian theory of language (or that Schenker developed a Humboldtian theory of music, given that his work followed Humboldt's by almost a century).

What makes Humboldt's work an organicist theory of language is, again, its attention to the organic form of the language system and the sentences it generates. This has been explicitly recognized by Noam Chomsky, who writes:

“The Cartesian emphasis on the creative aspect of language use, as the essential and defining characteristic of human language, finds its most forceful expression in Humboldt's attempt to develop a comprehensive theory of general linguistics. Humboldt's characterization of language as *energeia* (“activity” [Thätigkeit]) rather than *ergon* (“product” [Werk]), as “a generative activity [*eine Erzeugung*]” rather than “a lifeless product” [*ein todes Erzeugtes*] extends and elaborates – often, in almost the same words – the formulations typical of Cartesian linguistics and romantic philosophy of language and aesthetic theory. For Humboldt, the only true definition of language is “a productive activity” [*ein genetisches*] ... There is a constant and uniform factor underlying this [productive activity]; it is this which Humboldt calls the “Form” of language ... The concept of Form includes the “rules of speech articulation” [*Redefügung*] as well as the rules of “word formation” [*Wortbildung*] and the rules of formation of concepts that determine the class of “root words” [*Grundwörter*]. In contrast, the substance [*Stoff*] of language is unarticulated sound and “the totality of sense-impressions and spontaneous mental activities that precede the creation of the concept with the aid of language”. The Form of language is a systematic structure. It contains no individual elements as isolated components but incorporates them only in so far as “a method of language formation” can be discovered in them.” (Chomsky (1966): 69-70)

The above idea that the form of language is “a systematic structure” that contains no “isolated components”, clearly reveals Humboldt's organicist view of language, i.e. as a system in which the substance of language gives rise to products that are generated from a “constant and uniform” underlying factor. This is of course exactly the description of the computational form of music we have in Schenkerian theory, i.e. a system in which products (i.e. musical surfaces, made up of the ‘substance of music’, i.e. sound and meaning) are generated from a constant and uniform underlying factor, viz. the *Ursatz* and the rules, not of “speech articulation” or “word formation”, but of chord and musical phrase formation, given in the ‘constant and uniform’ rules of harmony and counterpoint.

The similarity of Humboldt and Schenker's ideas does not even remotely end here. I will focus on two more salient points of identity in this discussion, so as not to take us too far off course from our current Minimalist concerns – but it should be said that there are so many fascinating parallels between Humboldtian linguistics and Schenkerian music theory, that only a full discussion of this topic would do

it justice.⁷² The two points I will focus on are the points Noam Chomsky focuses on in his description of Humboldtian linguistics. The first point has to do with the issue of creativity – what Chomsky refers to as the “the essential and defining characteristic of human language” inherent in the Cartesian aspect of generative linguistics. We have already explored the proposal that both music and language exemplify human creativity, in the way they allow a potentially infinite number of surface structures to be generated from a finite set of lexical items and generative procedures. This is where the hierarchical and recursive emphasis of both Chomskyan and Schenkerian theory make their presence felt, because it is through recursive embedding that the potential infinity of surface structures can be generated, as I discussed in section 1.1.1. Within Schenkerian theory in particular, the above proposal has relevance not only as a description of the generative system of musical grammar, but also as a description of creative genius. This is because, as Schenker certainly believed, it is only the master composer who is able to create the kind of musical surface that reveals the unifying force of the *Ursatz* from which it was generated, while *also* revealing a strikingly imaginative and original skill at free composition, demonstrated by the composer’s masterful, intuitive knowledge of harmony and counterpoint.

What is of immense importance here though, is that this idea – of creativity being revealed through the infinite, free, and imaginative use of a finite set of structures and procedures – is a *Humboldtian* idea. It is Humboldt who is responsible for the maxim that language makes “infinite use of finite means” – it is Humboldt who said that “the domain of language is infinite and boundless ... [making] the fundamental property of a language... its capacity to use its finitely specifiable mechanisms for an unbounded and unpredictable set of contingencies” (Chomsky (1966): 70). So, when Schenker declares, e.g. when speaking of Beethoven’s “visionary” use of a passing tone high up in the foreground of a simple tonic prolongation in *Fidelio*, “to a genius, a simple prolongation can create who knows what unseen opportunity!” (Schenker (1979): 64) – he is just repeating *Humboldt’s* idea that it is the creative

⁷² See my forthcoming paper “Schenker, Humboldt, and the Origins of Generative Music/Linguistic Theory” for this fuller comparison of Schenkerian and Humboldtian theory.

aspect of music/language that allows us to use a “finitely specifiable mechanism” to *create* an unpredictable (or “unseen”, as Schenker puts it) contingency.

The second point of identity between Schenkerian and Humboldtian theory I would like to discuss here involves the notion of *freedom*. We have now explored both Schenker and Humboldt’s shared belief that music and language have an internal, organic form, in which myriad surfaces are generated from a finite, abstract background via “consistent and uniform” rules of some sort. This, in addition, reveals the creative nature of both music and language too. But for Schenker, the generation of structures from an abstract background (via the consistent and uniform rules of counterpoint) only leads to a bridge, which must then be crossed – i.e. to *free* composition, which is where the true creativity of the musical mind shines through.

So, “freedom” assumes some significance in Schenkerian theory. Its main significance lies in the fact that musical creativity results from *freewill*, as opposed to being *caused* by some external, physical ‘force’ (in the way a falling apple is subject to gravity, or in the way Descartes thought birds ‘spoke’). This is why we can open our mouths and sing whenever we want to, and which is why we do not *have* to say “ouch” when we are hurt, although we can if we *want* to. This already reflects a certain Romantic disposition towards human nature, as opposed to a mechanistic, Empiricist one, given the importance of freewill in a number of Romantic philosophies, like those of Schopenhauer and Nietzsche.

But recall the discussion from a few pages ago, in which we explored Schenker’s belief that composition is free only when this manifests itself in a specific way, i.e. when composers compose not according to their own, idiosyncratic will, but according to the *will of the tone*. In other words, truly *free* composition manifests itself when it is still governed by the internal organic form of music (i.e. the *Ursatz*), so that the ‘visionary’ treatment of pitch in a truly free musical surface still originates from the *strict* prolongation (or composing-out) of the *Ursatz*, which the greatest composers know how to do intuitively (Schenker (1979): 61). Put in modern parlance, this amounts to saying that we can still sing or speak whenever we want to, free from any external, physical causes, but this ability to sing or speak freely will only be useful if we know *what* to say and *how* to say it – i.e. if we are musical or linguistically

competent. Which, of course, can happen only if we have innate ('intuitive') knowledge of musical or linguistic grammar. So, freedom in Schenkerian theory is still situated in a generative theory of musical grammar, i.e. within a theory of music as an *aspect of human nature*, to repeat a point made by John Blacking (and Leonard Bernstein) that has been resonating since the beginning of this chapter.

Crucially, the notion of freedom is foundational for Humboldt's view of human nature too, and therefore for his organicist view of human language. In fact, it was even more critical for Humboldt than for Schenker, whose views on human freedom were pessimistic (and elitist):

"Art can bring together as many as two or three thousand people. But to assemble and entertain 50,000 people – this can be accomplished only by bullfights, cock fights, massacres, pogroms: in short, a brutal ranting and raving, a demented and chaotic outcry. Art is incapable of uniting such large numbers. It is the same in art as in politics. Just as "freedom" for all is no longer true freedom – it is merely a Utopian dream to "reconcile the ideal form of the liberalism, which really wanted only a new selection of elite in place of the obsolete feudal order, with the great experience of society and its great metamorphoses" (Coudenhove-Kalergi) – so "art for everyone is not art" (E.J. Schindler, the painter, in his diary)." (Schenker (1979): 159)

In contrast, Humboldt's political convictions were those of a liberal, and he made important contributions to moral philosophy too – so for him, freedom was a much more indispensable notion. Which is why it plays an equally, if not more, important role in his view of human nature, than it did for Schenker, especially with regards to his belief that humans have an innate desire to express themselves freely. Citing Humboldt's assertion that if a man acts in a purely mechanical way, "we may admire what he does, but we despise what he is", Chomsky says, "it is clear, then, that Humboldt's emphasis on the *spontaneous* [my emphasis] and creative aspects of language use derives from a much more general concept of "human nature", a concept which he did not originate but which he developed and elaborated in original and important ways" (Chomsky (1966): 74). So, despite being an organicist, i.e. a believer in the doctrine that the human ability for linguistic expression is governed by the internal, organic form of language, he strongly believed in the *freedom of expression* this internal form enables humans to have too – the freedom of expression that humans would not even have if it were not for the creative properties of the language faculty, and the potential for infinite creative expression it endows us with. In other words, the very fact of human linguistic competence gives humans a right to free expression, as this competence is

an aspect of human nature. So, the spontaneous generation of a (linguistic) expression is still situated in a generative theory of grammar for Humboldt, just as it was for Schenker, and still requires an intuitive knowledge of the form of language, as it did for Schenker in the case of music – although *who* is able to generate such expressions, and *how*, is a point of disagreement between the two thinkers, possibly because of their political differences (with Schenker reserving the ability to express freely only for the Germanic master composers, given his belief that only they had intuitive knowledge of music’s internal form). If we remove the two thinkers’ political value judgments from their quite profound, and rather technical, contributions to music/linguistic theory though, two remarkably congruent views of human nature and musical/linguistic expression emerge.

There is one last point of identity between Schenker and Humboldt that deserves a brief mention. This has to do with Chomsky’s appraisal of Humboldt’s ideas in the light of their being the basis for modern generative linguistics, as Chomsky has acknowledged himself. Now, we have already explored the debate about whether Schenkerian music theory can be defended as a genuine grammatical theory of tonal music. Notice what Fred Lerdahl and Ray Jackendoff have to say about this:

“Schenker can be construed (especially in *Der freie Satz*) as having developed a proto-generative theory of tonal music – that is, as having postulated a limited set of principles capable of recursively generating a potentially infinite set of tonal pieces. But, remarkable and precursory though his achievement was, he did not develop a formal grammar in the sense one would expect nowadays of a generative theory.” (Lerdahl and Jackendoff (1983): 337)

Now compare this to what Noam Chomsky has to say about Humboldt:

“For all his concern with the creative aspect of language use and with form as generative process, Humboldt does not go on to face the substantive question: what is the precise character of “organic form” in language. He does not, so far as I can see, attempt to construct particular generative grammars or to determine the general character of any such system.” (Chomsky (1966): 75)

The similarity of both critiques is truly intriguing – which just adds to the evidence for the identity of Schenkerian and Humboldtian theory, albeit from the paradoxical perspective that they both *failed* to do something in the same way. However, there is an interesting historical moral here, which is that Chomsky went on to develop his famous program in generative linguistics, leading to the current Minimalist

Program, based on Humboldt's ideas. But apart from Allan Keiler's work, no such attempt has been made to develop a similar universal generative grammar of music based purely on Schenker's ideas, despite its connections to the, successful, research program in Humboldtian generative linguistics. Which is why a Minimalist approach to generative musical grammar seems all the more important.

The above discussion demonstrates just how convergent the generative study of language, and that of music, have been in the history of ideas, and just how similar their origins are too. This is why I asserted the first Identity Thesis proposed earlier, viz. that music theory and linguistic theory are identical.

But at this point a caveat made earlier needs to be repeated too. This is the point about how Schenker was primarily a musician, not an academic (his sole non-musical, academic credential was that of a student of law earlier in his life; see Alpern (1999) in this regard), and that his ideas can be interpreted, equally legitimately, as constituting either a scientific theory or an artistic one. Since I situated his ideas in the broader intellectual tradition described above partly in defense of a Minimalist interpretation of Schenkerian theory, an objection can be certainly raised concerning the extent to which Schenker was genuinely influenced by this broader tradition – i.e. to what extent do any of the aforementioned authors (especially Wilhelm von Humboldt) have any genuine connection to what Schenker really believed? (Nicholas Cook has even raised the possibility in this regard, that all of the above intellectual influences on Schenker's thought are grossly exaggerated, given his limited academic affiliations and interests (Cook (2007): 44-48), but see Korsyn (2010) for a rebuttal of this position.)

The important point here though is that I am not even asserting any *explicit* link between Schenker, and the other thinkers cited above, including Humboldt. There is no evidence, as far as I can tell, that Schenker knew Humboldt's work, and Humboldt died before Schenker was even born – and so could not have been influenced by the latter's ideas. So, the links proposed above between the two thinkers are, crucially, *implicit*, which is demonstrated in the way their theories ended up being so similar, *independently* of the other's influence. (Humboldt was a close friend of Goethe though, so the latter's influence on the former's linguistic theories is explicit and well-documented.) Yet, the fact that

Schenker's and Humboldt's theories did end up being so similar just reinforces the two identity theses proposed in this chapter, viz. that music and language are identical, and that *this* identity is what compels scholars to theorize about them in similar ways too.⁷³

In general, the historical and technical links between the ideas of Humboldt and the generative tradition in linguistics, on the one hand, and Schenker and the generative tradition in music on the other, seem pretty remarkable – and which is what drives the Minimalist Program for language *and* music proposed by this dissertation. But what is even more remarkable is how both traditions have been explicitly or implicitly critiqued in the *same* way by corresponding anti-generative, anti-Rationalist, and even anti-scientific intellectual traditions in both music and linguistic scholarship. In the next section, the final section of this chapter, I will describe some of these traditions.

The history of modern generative or scientific music/linguistic theory, and the anti-generative responses to it, are complex, and a thorough treatment of these well beyond the scope of this current project.⁷⁴ Consequently, my subsequent discussion will necessarily be brief – and therefore my treatment of these anti-generative traditions perhaps unfair. My primary goal here, however, is less to critique these traditions, but rather to show how their widespread influence in music scholarship has been the main obstacle in developing a joint Minimalist Program for language and music. Also, I hope to show just how similar (and common) the attacks on generative music and linguistic theory have been, thus demonstrating not only the centrality of the ideas of Schenker and Chomsky in their respective disciplines

⁷³ There is one point of similarity between Schenker and Humboldt that results from actually separating the latter from the generative linguistic tradition. Michael Losonsky, in his introduction to the Cambridge University Press English edition of Humboldt's 1836 masterpiece, says that one thing that the Chomskyan tradition has ignored in Humboldt's work is the latter's interest in the *aesthetics* of language (Humboldt (1999): xxxi). That is, Humboldt was deeply interested in connecting rhythm and other aesthetic features of the sound of language to inner mental activity, in a kind of 'grammar of thought'. The generative tradition has taken an interest in aesthetic matters from time to time, but this, as Losonsky correctly asserts, has certainly not been the primary focus of most Chomskyan linguists. Losonsky goes on to say that Humboldt might have been "onto something when he explained that the emergence of diverse sound-forms is in part a function of the influence of *inner musical forms* [my emphasis]." We have explored the possibility of Schenkerian theory being an aesthetic, rather than generative grammatical, theory – in which case Humboldt's own aesthetic (and musical) musings might serve as evidence for an even deeper connection between Schenker and Humboldt, although, strikingly, not from a generative perspective.

⁷⁴ A more comprehensive treatment of this complex history is in preparation in my monograph *The Princeton Grammarians and the Foundations of Music Theory*.

but also the uniformity of the debates in these disciplines – which if anything, acts as robust evidence for the identity of musical and linguistic theory, and therefore for the identity of music and language.

1.1.4. The “Princeton Schenker Project” and GTTM revisited

But first, some more evidence for Identity Thesis A. We have already explored the fascinating overlaps between generative musical and linguistic theory in the 19th, leading into the early 20th, century in the works of Wilhelm von Humboldt and Heinrich Schenker. We have also seen how under the guidance of Noam Chomsky, the modern science of generative linguistics picked up on this earlier generative project specifically in language scholarship, and created a paradigm that has seen much revolutionary research over the last 50 years, leading to the current Minimalist Program in linguistics. This has been the recent history of the generative project, much of which has been created in the Department of Linguistics at the Massachusetts Institute of Technology, where Chomsky, and other leading linguists like Morris Halle, Kenneth Hale, and many of their illustrious colleagues and students, have maintained their intellectual residence over the last fifty years.

But what about the recent history of the generative project in music? Well, we have already explored a bit of it in Leonard Bernstein’s famous *Unanswered Question* lectures in the 1970s, and through the brief glimpses we have had of Allan Keiler’s work, beginning in the same time period. But there is a wider intellectual context for this, and like the role of MIT in generative linguistics, Princeton (and subsequently Yale) in the 1950s to the 1980s has been the locus for much of this activity, i.e. this activity in generative, and specifically Schenkerian, *music theory*. The recent history of the generative project in music can therefore be referred to as the “Princeton Schenker Project” or “PSP”.⁷⁵ Unlike the MIT project in linguistics, which remains active and trendsetting even today, the Princeton Schenker Project reached its zenith in the 1980s though, with two major contributions, viz. Allan Keiler’s aforementioned work in Schenkerian theory; and Fred Lerdahl and Ray Jackendoff’s celebrated *A*

⁷⁵ This name was suggested to me by Joseph Straus.

Generative Theory of Tonal Music (the *GTTM* in this section's title), the latter coming out of the PSP, but which then pushed generative music theory in a more anti-Schenkerian direction.

After this the PSP all but disappeared from the music-theoretic scene, partly due to the success of *GTTM*'s anti-Schenkerian proposals, and partly because music theorists moved on to other areas of interest. So, in this final part of chapter 1.1, I will give a brief history of the PSP, after which I will conclude with a description of some of the anti-Schenkerian trends in music scholarship that led to the demise of the PSP in the 1980s, especially Lerdahl and Jackendoff's *GTTM*.

So, now we travel back to Princeton in the 1950s. The Princeton Department of Music was home then (and continued to be, until his death in 2011) to Milton Babbitt, whose role in founding the contemporary discipline of music theory is not dissimilar to that of Noam Chomsky in contemporary linguistics. Babbitt was one of the leading American composers of the 20th century, known particularly for his cerebral and systematic approach to modernist idioms like serialism and dodecaphony, and later, electronic music. Due to his influence, and that of the other eminent composers on the Princeton faculty (including his teacher Roger Sessions, and colleagues Edward Cone and Earl Kim), Princeton became a preeminent center for contemporary composition.

But Princeton in the 1950s and 60s was also home to Alonzo Church, with whom Alan Turing had spent time working about a decade earlier, and the Princeton mathematical scene was also home in those years to John Forbes Nash (of game theory fame) and John Tukey (who co-invented the popular Cooley–Tukey version of the Fast Fourier Transform, which was significant for the nascent computer music scene) – and Kurt Gödel, John von Neumann, and even Albert Einstein were in residence down the road at the Institute for Advanced Study, which itself was directed by the famed nuclear physicist Robert Oppenheimer. (Noam Chomsky was in residence at the Institute during 1958-59 too.)

So, Princeton became a legendary hub for science, especially the mathematical and computational sciences, in the middle of the 20th century. Much of this had to do with Cold War politics, and Department

of Defense sponsored military projects that required mathematical innovation.⁷⁶ But this was also the environment that fomented both the computer revolution and the cognitive revolution, one result of which was Chomsky's development of a more computationally- and cognitively-oriented study of language at MIT.

These developments had a significant impact on the music composition scene at Princeton. Computers, mathematical models, and a generally technological approach to writing music soon came to influence a number of composers, in both the Princeton faculty and student bodies. (A significant moment here was the founding of the Columbia-Princeton Electronic Music Center in 1959, to further both electronic composition, and the scholarly study of computer music.) But the computer revolution did not just influence Princeton composers practically, i.e. in helping them find new instruments with which to express their musical thoughts. The cerebral nature of the music they were writing, and the radical new listening experiences this engendered, made many of them self-conscious of their music, which inspired some of them to engage with deeper, philosophical questions about the nature of music, the validity of different musical experiences, the deeper structure of musical systems, and so on. As Princeton composer Paul Lansky says:

“Let me flash back now to the fall of 1966 when I entered the graduate program at Princeton. These were very heady times in the musical world (pun intended). The paroxysms of postwar music had come to a boil and the world was full of institutions staking claims to hegemonic superiority, with Princeton perhaps leading the pack in America. Stravinsky had become a card-carrying 12-tone composer and my first week at Princeton coincided with a visit by him for the premiere of his *Requiem Canticles* at McCarter Theater. The work was commissioned by Stanley Seeger, a Princeton alumnus, in memory of his mother. We all felt a kind of glee and sense of superiority: the future was ours and the rest of the world would come to its senses eventually and jump aboard. Even Aaron Copland was writing 12-tone music. (A well-known performer of new music was reportedly raising his children listening to nothing but 12-tone music.) It is hard to exaggerate the influence and brilliance of Milton Babbitt at that point. He was just 50, had hit his stride, and gave wonderful seminars on the theoretical and mathematical aspects of the 12-tone system, and was writing scintillating pieces. Required reading was Nelson Goodman, Rudolf Carnap, Quine and others. The famous *Princeton Seminars in Advanced Musical Studies* had taken place in 1959 and 1960 (that led to the Musical Quarterly issue and book appropriately entitled, *Problems of Modern Music*), and *Perspectives of New Music* had just been launched in 1964 at Princeton University Press, supported by Paul Fromm.” (Lansky (2009))

⁷⁶ Robert Oppenheimer was of course the director of the Manhattan Project at Los Alamos, which is a position he held prior to his taking up the directorship of the Institute for Advanced Study. Also, Milton Babbitt was involved with certain projects in Washington D.C. during the Second World War that still remain classified.

So, an interdisciplinary program of composition and scholarship came to represent the trade of the Princeton composer – long before interdisciplinarity became fashionable in the humanities. This is what led to the institutionalization of music theory, as a professional – and particularly, formal – discipline in the academy, a process that was consolidated by the creation of a doctoral program in music theory (the first of its kind) at Princeton in 1961 (the same year the MIT doctoral program in linguistics was inaugurated!).

Following in Princeton's footsteps, Yale created a doctoral program in music theory too, in 1965. Yale's situation was different from Princeton's though because Yale already had a School of Music, where theory had been taught as a practical discipline, both for School of Music undergraduates and for Yale College students (i.e. 'liberal arts' majors) – most notably by Paul Hindemith. With Hindemith's departure from Yale, and the phasing out of Hindemith's theory program at the School of Music, Yale needed a way to provide instruction in music theory for its students, and this is what led to the establishment of the music theory program in the *Graduate School* at Yale, where it was separated from the compositional activities occurring in the School of Music – unlike at Princeton, where they were both performed by the Princeton "composer-theorists" (see Dubiel (1999) for an examination of this term). This also led Allen Forte, who directed this new program at Yale, to become essentially the first 'professional' music theorist, since his duties were not divided between composing and theorizing, but were devoted solely to the latter instead. This orientation also allowed Yale to focus more on the teaching of theory, which can be seen in the numerous doctoral dissertations that were advised by Forte (compared with the rather few, specifically theoretical, dissertations advised by the Princeton composer-theorists), and also in the textbooks written by the Yale faculty, including Forte's important pedagogical treatise on Schenkerian theory, co-written with Steven Gilbert (i.e. Forte and Gilbert (1982)).

Like Princeton's *Perspectives of New Music*, Yale also had a music-theoretic journal of record, viz. *The Journal of Music Theory*, which was established in 1957. These two publications remained the flagship, peer-reviewed journals of the field until they were joined in their ranks by *Music Theory Spectrum* in 1979, *Spectrum* being the journal of the recently formed (in 1977) Society for Music Theory.

The SMT was the first national academic institution devoted to music theory – so, with its formation, the institutionalization of music theory, as a nationally-recognized professional discipline, was complete.

Now, one could easily, and inappropriately, over-emphasize Milton Babbitt's role at Princeton (or Allen Forte's at Yale) in the establishment of academic composition and music theory as professional disciplines, as Aaron Girard has pointed out (Girard (2007)). There were many individuals who were involved in this 'movement', including scholars who were not practicing composers or theorists (such as Arthur Mendel at Princeton, and Claude Palisca at Yale) – with the reasons proposed by these individuals for why composition and theory should be institutionalized being numerous and diverse. Often the reason was pragmatic – e.g. composers with degrees had an easier time getting jobs. And for some, particularly older scholars, who came of age before the computer and cognitive revolutions, the role of music in the academy was simpler and more practical, i.e. it was meant to train good musicians and good audiences, as Roger Sessions – the seniormost member of the Princeton composer-theorist community – seemed to believe, and which was arguably Paul Hindemith's mission at Yale too. However, it is clear that Milton Babbitt's own attitude towards specifically music *theory* was that of a scientist:

“Musical theory is today being transformed from a collection of dubiously derived and inaccurately stated prescriptives and imperatives into a subject that draws, as it must, upon the methods and results of the formal and empirical sciences: logic, the philosophy of science, analytical philosophy, physics, electronics, mathematics, experimental psychology, structural linguistics, and computer methods. Such investigations can be undertaken only in a university, and we wish to encourage them and see them take place at Princeton. We are enthusiastic particularly about their interdisciplinary character, and their already evident contributions to music and to the teaching of music at the university level.” (quoted in Girard (2007): 216)

And given that the institutionalization of music theory (as opposed to academic composition) in his home department was more of an individual effort on Milton Babbitt's part – and given his eminence in the academic musical community – the scientific origins of the modern discipline of music theory (again, as opposed to composition) cannot be doubted.

In other words, we see that a generative project in music theory is consonant with the events that conspired in the Princeton and Yale music departments in the middle of the 20th century, given their

shared scientific orientation. This means that the modern discipline of music theory, as it emerged at Princeton, might have striking parallels to the modern discipline of linguistics, as it emerged at MIT – which creates the possibility that part of the recent history of music theory is essentially the recent version of the generative project in music theory, begun by Schenker around the turn of the century. This is what we might call the Princeton Schenker Project.

And Schenker's ideas did attract the attention of several theorists at Princeton, and later Yale, too. However, given their commitment to music theory as a science, especially in Babbitt's case, this attention was explicitly scientific in its disposition (although what *kind* of science they understood Schenkerian theory to be is something we will have to explore a bit). This means that the Princeton Schenker Project took a stand in the Schenker-as-science versus Schenker-as-art debate, in favor of the former. Given the uniquely American origins of this stand – in light of the American origins of the modern, scientific discipline of music theory – William Rothstein has referred to this as the “Americanization” of Schenkerian theory (Rothstein (1990b)). Given the institutionalization of this paradigm in the (Ivy League) academy as well, away from music conservatories, as particularly happened at Yale, this led to a split between the newer practice of Schenkerian theory by the Princeton and Yale Schenkerians (and also the more systematically-inclined, if not explicitly scientifically-oriented, followers of Schenker's ideas) and the less systematic, interpretive practice of Schenkerian analysis, often by individuals who, like Schenker, were performing musicians and music analysts/critics. This leads to a distinction between “university Schenker” and “conservatory Schenker” too, as Rothstein (2002) describes it, even though the distinction is not watertight, given that many Schenkerians are happy to work within either paradigm depending on the kind of project they happen to be pursuing at the time.

In this light, the notion of a Princeton Schenker Project is a bit of an abstraction, since Schenkerians often wear both “university” and “conservatory” hats. In fact, I will suggest in a moment that some of the most interesting work within the Princeton Schenker Project has been done by Schenkerians who were never affiliated with either Princeton or Yale, and who spent most of their careers in the conservatory – a good case in point being the great Schenkerian theorist Felix Salzer, and his

eminent student Carl Schachter. Therefore, the clearest case of an Americanized, university Schenkerian is really just Milton Babbitt himself, along with a handful of his students. This ‘hard line’ version of the Princeton Schenker Project emerges from a couple of interesting features in Babbitt’s appropriation of Schenkerian theory. First of all, Babbitt was politically conservative, like Schenker, and a cultural elitist to boot (e.g. see Babbitt and Grimes (1986)) – despite being a Jew (again like Schenker), and having experienced anti-Semitism at the hands of politically conservative and elitist individuals and institutions as well.⁷⁷ But where he and Schenker differ are in his rejection of Schenker’s commitment to common-practice tonality, as a natural system – which is understandable given Babbitt’s interest in musical modernism. So, Babbitt’s interest in Schenkerian theory lay in its *systematic* rather than *naturalistic* features.

Some comments that Babbitt makes in his review of Felix Salzer’s *Structural Hearing* provides evidence for this. For example, he concedes that the origins of Schenker’s ideas are probably empirical, rather than logical, but still (cautiously) advocates interpreting Schenker’s ideas axiomatically instead:

“Schenker’s analysis originated in aural experience, and the *Urlinie* is, at least indirectly, of empirical origins. On the other hand, it is (and this is merely an additional merit) completely acceptable as an axiomatic statement (not necessarily the axiomatic statement) of the dynamic nature of structural tonality. Stated in such terms, it becomes the assertion that the triadic principle must be realized linearly as well as vertically; that the points of structural origin and eventuation must be stabilized by a form of, or a representation of, the sole element of both structural and functional stability: the tonic triad.” (Babbitt (1952): 260)

Later, he explicitly rejects Schenker’s criticisms of modern music as being irrelevant to the latter’s theory:

“Schenker’s contribution has often been subjected to criticism for its presumable inapplicability to music written prior to Bach and after Brahms. Schenker himself is responsible for his apparent vulnerability on this point, but, in fact, his ill-tempered and often inconsistent attacks on contemporary music, his dedicatory description of Brahms as “the last master of German composition” are as irrelevant to the core of his theory as his many and unfortunate excursions into the realm of the political, social, and mystical.” (Babbitt (1952): 264)

⁷⁷ For example, there is a well-known story of how Babbitt’s appointment to the Princeton faculty was delayed by a year because the department chair did not want to appoint a Jew in the first year of the department’s existence. See Robert Hilferty and Laura Karpman, “Milton Babbitt: Portrait of a Serial Composer”. Published online January 13, 2011 at <http://www.npr.org/event/music/144763523/milton-babbittportrait-of-a-serial-composer>. Accessed January 14, 2012.

Elsewhere, Babbitt continues to advocate the axiomatic approach, especially because of the way this reveals the systematicity of the theory, as akin to that of generative linguistics:

“The Schenkerian theory of tonal music, in its structure of nested transformations so strikingly similar to transformational grammars in linguistics, provides rules of transformation in proceeding synthetically through the levels of a composition from “kernel” to the foreground of the composition, or analytically, in reverse. Since many of the transformational rules are level invariant, parallelism of transformation often plays an explanatory role in the context of the theory (and, apparently, an implicitly normative one in Schenker’s own writing). The formulation of this theory in relatively uninterpreted terms (as Kassler is doing), as a partially formalized theory, serves to reveal not only its essential structure but its points of incompleteness, vagueness, and redundancy, and the means for correcting such flaws. The laying bare of the structure of an interpreted theory, in a manner such as this is an efficient and powerful way also of detecting false analogies, be they between systems (for example, the “tonal” and the “twelve-tone”), between compositional dimensions (for example, that of pitch and that of timbre), or between compositions (with a composition regarded as an interpreted theory).” (Babbitt (1965): 60)

So, Babbitt seems to be saying that the best way to make the “tonal” versus “atonal” distinction is through *formalizing* Schenkerian theory, rather than on the empirical and polemical grounds on which Schenker himself bases this distinction. This makes sense within Babbitt’s worldview too, because this validates twelve-tone music, as long as one can develop a formalized, axiomatic way of describing *that* system. In this sense, Babbitt’s elitism is tempered, since he is willing to accept the legitimacy of any musical system that can be formalized in the above manner (a point Aaron Girard makes too, in Girard (2007): 238-239) – but formalize one must, because:

“A composer who asserts something such as: “I don’t compose by system, but by ear” thereby convicts himself of, at least, an *argumentum ad populum* by equating ignorance with freedom, that is, by equating ignorance of the constraints under which he creates with freedom from constraints. In other words, musical theory must provide not only the examination of the structure of musical systems – familiar and unfamiliar by informal conditioning – as a connected theory derived from statements of significant properties of individual works, a formulation of the constraints of such systems in a “creative” form (in that, as a language grammar does for sentences, it can supply the basis for unprecedented musical utterances which, nevertheless, are coherent and comprehensible), but – necessarily prior to these – an adequately reconstructed terminology to make possible and to provide a model for determinate and testable statements about musical compositions.” (Babbitt (1965): 49)

As we explored earlier, a generative grammar is a theory of the internal form of a system (such as language or music), but not necessarily a *formalist* one. (Here I am using “formalist” technically, to refer to the kind of systematic phenomenon that can be given a logical analysis.) That is, if one accepts the Minimalist line on this, a generative theory is a *naturalistic* theory instead; so it is validated or invalidated

not on logical grounds, but on empirical grounds, e.g. on the basis of whether or not its hypotheses accord with the intuitions of competent observers (such as the native speakers of a language).

Babbitt's lack of interest in empirical validation, as evident in the way he underprivileges Schenker's empirical judgments, is understandable – because such empirical validation is not available for atonal music.⁷⁸ But in this lies the major difference between the Princeton Schenker Project, with its emphasis on formalism, and the MIT project in generative linguistics, with *its* emphasis on naturalism. And this is also why I believe the PSP did not succeed in the end, i.e. because of its inability to conceive of the generative study of music as a natural science of music – based on Schenkerian organicism – in the way MIT linguistics conceived of the generative study of language as a natural science of language, based on Humboldtian organicism.

A naturalistic reworking of the PSP would have been more in alignment with what was happening in MIT linguistics, and this would have implied a greater emphasis on the psychological study of musical grammar, i.e. a psychological study of the internal, organic form of music. But the PSP evolved in two rather different ways, and the task of naturalizing Schenkerian theory was only taken up by scholars who were outside of the 'hard line' of the Project.

The first of the two directions the PSP took in the late 1960s, going into the 1970s and 80s, was a continuation of the original formalist focus, i.e. the project of formalizing Schenkerian theory, as an axiomatic system. This happened most prominently in the work of Babbitt's protégé Godfrey Winham, and particularly in the work of their mutual protégé Michael Kassler, whose computer models made important strides in generating musical structures based on explicit rule systems (Kassler (1967, 1977), also see Blasius (1997)). This led to the emergence of computational music theory as an important paradigm, especially within the university Schenker fold, even though some of the scholars who

⁷⁸ Babbitt achieved notoriety for an article in *High Fidelity*, whose title, "Who Cares If You Listen?", seemed to confirm the stereotype that high-modernist composers are elites, who do not care about how their compositions are received by a wider audience – even though he actually titled the article "The Composer as Specialist", this being changed to the more scandalous title without his permission. But there is some basis to this notoriety though, given the foundation of his art in logic rather than in perception or intuition – as the above discussion implies.

developed generative computer models were not directly affiliated with Princeton, such as Terry Winograd (1968), and Stephen Smoliar (1980). The first doctoral dissertation supervised by Allen Forte at Yale was that of the eminent Schenkerian John Rothgeb, who wrote a computer program to realize the upper voices of an unfigured bass line digitally (Rothgeb (1968, see also 1980)). (Although much of Rothgeb's later Schenkerian work would not be in a specifically PSP vein.) Rothgeb's student James Snell also made important contributions to the computer modelling of musical phrase structure (e.g. Snell (1979, 1983)). Finally, the formalization of Schenkerian theory remains an area of some interest today (e.g. see Mavromatis and Brown (2004), Marsden (2005)), and parallels, in this respect, the attempt to formalize Chomskyan theory by computational linguists who maintain an interest in generative grammar, such as Edward Stabler (see e.g. Stabler (2009, 2011), Collins and Stabler (2012)).

The second way in which the PSP evolved into the 1970s and beyond did show an engagement with the psychological attributes of music, but in a way that was characteristic of the Princeton music department. For the Princeton composer-theorist, theory was of course not just an instrument for musical composition and analysis, it was also an instrument for self-reflection, and with which to reflect on the creations of others. This led to the emergence of a characteristically phenomenological attitude in some corners of the Princeton program, part of which led to a phenomenological, and also metatheoretical, exploration of issues that had been at the heart of the PSP. Benjamin Boretz's *Meta-Variations* probably symbolizes this side of Princeton theory best to most people, especially those parts of his text that deal with the issue of analysis (i.e. Boretz (1972, 1973)) – whose consideration had been arguably overshadowed by the focus on composing and theorizing elsewhere. But PSP-related issues, such as an examination of the notion of musical grammar, also came under Boretz's metatheoretical gaze (e.g. in Boretz (1970, 1971)), as they did in the work of other Princeton theorists, like Joseph Dubiel (e.g. in Dubiel (1990)).

What I find interesting though, is that while the various projects in formalization, phenomenology and metatheory were developing within and out of the PSP, many Schenkerian theorists not affiliated with

Princeton were developing Schenkerian projects that were closer in spirit to a naturalization of Schenkerian theory than maybe they were aware of themselves. A case in point is Carl Schachter's celebrated series of articles on rhythm from a Schenkerian perspective (Schachter (1999a, b, c)), which I shall engage with in detail in chapter 2.2. As a theorist working in conservatory environments like the Mannes College, the Aaron Copland School at Queens College, and later Juilliard, Schachter might appear to be the consummate conservatory Schenkerian to some. (Schachter taught a popular course at Mannes for many years devoted to analysis for performers, which resulted in Schachter (2005) – which is something that would have been anathema in Forte's Yale, and probably Babbitt's Princeton too.) However, Schachter's negotiation of the thorny issue of rhythm in tonal theory was so profound, so lucid – and so *systematic* – that it became a model *theory* of rhythm. To the extent that Lerdahl and Jackendoff, who had distanced themselves from Schenkerian theory for its lack of explicit theorizing about rhythm, refer to Schachter's ideas as coming closest to their – naturalistic, linguistics-inspired – generative theory of rhythm (Lerdahl and Jackendoff (1983): 335).

Like Schachter, there have been other theorists who have made important contributions to the PSP, especially in the naturalization of Schenker, some of who were officially affiliated with Princeton, and to varying degrees with the paradigms of Schenkerian formalization or phenomenology too – e.g. Arthur Komar (1971), Peter Westergaard (1975), and David Epstein (1979). However, none of these projects yielded a generative grammar of music, in the way that was being developed for language at MIT, and the diversity of individuals and projects here lack the singular vision that, say, the formalization of Schenker project has had since the inception of the PSP. For this reason, the PSP seems to have reached its peak with the two explicitly-Chomsky inspired projects aimed at naturalizing Schenker that were proposed in the 1980s, viz. those of Allan Keiler, and Fred Lerdahl and Ray Jackendoff.

I have already alluded to Keiler's ideas a little, and have talked about the influence they have had on this dissertation. Essentially, Keiler's project came the closest to being a genuinely musilinguistic

project, given his loyalty to both Schenkerian and Chomskyan points of view.⁷⁹ Since I will be revisiting Keiler's ideas many more times in this dissertation, I will not discuss them any further right now.

However, it is certainly time for us to discuss Lerdahl and Jackendoff's highly influential work on musical grammar. Some might wonder why I have even delayed a discussion of their important work so far into this dissertation, given that to many Lerdahl and Jackendoff's *A Generative Theory of Tonal Music* (henceforth "GTTM") is the very epitome of a musilinguistic research project. But GTTM actually departs from both Schenkerian and Chomskyan perspectives in quite important, but often, subtle ways, which, given the importance of both of these paradigms for the musilanguage hypothesis, makes GTTM's musilinguistic credentials problematic – but which can only be critiqued once GTTM has been properly contextualized within the Princeton Schenker Project. Hence my postponement of this critique until now.

That GTTM emerges from within the PSP is undeniable. Fred Lerdahl was a student of Milton Babbitt, Edward Cone, Roger Sessions, and Earl Kim at Princeton in the 1960s, and continues to be an important composer-theorist in the academic musical community. In addition, his interest in (cognitive) scientific investigations of musical structure was fully consonant with the similar orientation of the PSP, although Lerdahl was more explicitly naturalistic in this regard. This, combined with his unwavering commitment to writing and thinking about music that was tonal, made Lerdahl somewhat of an outlier at Princeton, given the high-modernist trends of 1960s-70s Princeton:

“Near the beginning of my composing career, around 1970, I underwent a crisis of belief. Modern music had splintered into mutually incompatible styles, each with its own aesthetic, and any coherent sense of the historical trajectory of art music was gone. Contemporary compositional methods were often highly rationalized but inaccessible to listeners except by conscious study. I sought instead to establish my music on a foundation free of the labyrinthine history of twentieth-century music and its often *perceptually obscure* [my emphasis] techniques. A reading of Noam Chomsky's *Language and Mind* opened new vistas. If it was possible to study the language capacity that lies beneath the variety of human languages, was it not also possible to study the musical capacity? I wanted to base my musical development not on

⁷⁹ Though he was never affiliated officially with Princeton as a full-time student or faculty member, Allan Keiler's affiliation with the Princeton Schenker Project is clear. Scott Burnham, who was a doctoral advisee of Keiler's, reports (in a personal communication) that “Milton Babbitt once told me that he thought that Keiler was doing the most intelligent work of anyone on Schenker, and he (Milton) always asks me about this same book [i.e. Keiler's unfinished book project from Harvard University Press, titled “Schenker and Tonal Theory”].”

history but on nature. Such a quest has a long history in many guises, and mine was nothing if not utopian. But a young composer worth anything at all must have big dreams.”⁸⁰

Therefore, Lerdahl’s contributions to the PSP actually came after he left Princeton, particularly after he joined forces with the linguist Ray Jackendoff, who had been a student of Noam Chomsky at MIT. Jackendoff had already made important contributions to the study of linguistic generative grammar in the 1970s, particularly in the field of X-bar theory, but as a professional clarinetist he had an abiding interest in music too. This led to Lerdahl and Jackendoff’s collaboration in the late 1970s, which culminated in their first set of proposals about musical generative grammar (in Lerdahl and Jackendoff (1977)), which described tonal structure in terms of Chomskyan tree diagrams, and which also introduced their “preference rule”-based approach to modeling tonal grammar. This early text later gave rise to the more mature theory that Lerdahl and Jackendoff are best known for, i.e. GTTM (Lerdahl and Jackendoff (1983), and more recently, Jackendoff and Lerdahl (2006)).

Now the GTTM project started within an explicitly Schenkerian view of tonal structure, as Lerdahl and Jackendoff have acknowledged themselves (e.g. see Lerdahl and Jackendoff (1983): 337-338). But soon after its inception, it departed from Schenker’s vision in significant ways too (Lerdahl and Jackendoff (1983): 5 and 112, Lerdahl and Jackendoff (1985): 148, Lerdahl (2009): 188). We have already investigated a couple of reasons for this departure: first, some of the problems associated with Schenker’s conception of the *Ursatz* deterred Lerdahl and Jackendoff from basing their model on it in the way Schenker did, and secondly, their externalist interest in adding an independent theory of rhythmic structure to their model – and their belief that such a theory was not available within Schenkerian theory – prevented them from maintaining a Schenkerian focus in their model too.

⁸⁰ Fred Lerdahl, “What role has theory played in your compositions and how important is it for people to know the theory behind the music in order to appreciate it?”. Published online April 1, 2003 at <http://www.newmusicbox.org/articles/What-role-has-theory-played-in-your-compositions-and-how-important-is-it-for-people-to-know-the-theory-behind-the-music-in-order-to-appreciate-it-Fred-Lerdahl>. Accessed February 17, 2012. As a result of these commitments on Lerdahl’s part, Richard Taruskin has referred to him, somewhat bizarrely, as being “postmodernist” (Taruskin (2009): 445).

But there is another, and more important reason for why Lerdahl and Jackendoff parted ways with a more traditionally Schenkerian approach in GTTM, and this has to do with their focus on *perception*. The late 1970s and early 1980s was the period when the experimental study of music, particularly within psychology departments, was beginning to emerge, with important contributions to the field by scholars like Roger Shepard, Carol Krumhansl, Walter Dowling, Diana Deutsch etc. being published during this time (e.g. Krumhansl and Shepard (1979), Krumhansl (1979), Deutsch (1982), and Dowling and Harwood (1986)).⁸¹ This reflects a serious emerging interest in the idea that proposals about musical structure should be empirically verifiable, and particularly in the controlled environment of a laboratory. In such an environment, the examination of musical perception then takes centerstage, because this is something that can be tested relatively easily, e.g. through experiments that investigate how people hear a given musical surface.

Lerdahl and Jackendoff wanted GTTM to be an empirically testable theory (Lerdahl and Jackendoff (1983): 5). So, rather than conceiving of their model as one that begins with the description of the abstract form from which surfaces are generated (as Schenker did), they began with the surface itself, in order to give a formal, yet empirically-verifiable, account of “the structure that the experienced listener infers in his *hearing* [my emphasis] of the piece” (Lerdahl and Jackendoff (1983): 6). This led to the development of their famous “preference rule” system, which is a set of rules that govern how musically-competent listeners arrive at a preferred structural description of a musical surface, on the basis of their hearing that surface. (This also explains their emphasis on an independent theory of rhythm within a generative theory too, since rhythmic, and especially metrical cues, clearly play a role in how listeners parse a musical surface, as GTTM’s Metrical Preference Rule system attempts to reveal – which is a

⁸¹ *Music Perception*, the flagship journal of the field, was introduced in the fall of 1983 too. The very next issue was devoted exclusively to three music *theory* articles: two contributions in the generative tradition, viz. one by Lerdahl and Jackendoff, summarizing the main claims of the recently-published GTTM (Lerdahl and Jackendoff (1983-84)), and an article by Allan Keiler that examined some aspects of Schenkerian theory within a more contemporary generative framework (Keiler (1983-84)). The third article was an anti-generative one by Eugene Narmour, which challenged the notion that tonal musical structure is hierarchical (Narmour (1983-84)).

function that meter and rhythm do not necessarily have in the way those same surfaces are generated from an abstract background.)

GTTM proposes four preference rule systems – viz. for grouping structure, metrical structure, “time-span” structure (which basically combines the first two), and prolongational structure (which is basically a hierarchical description of pitch structure of the kind found in Schenkerian analyses of musical surfaces). Given that Schenkerian-style pitch analyses of a surface come last in its structural description of that surface, GTTM essentially reverses the manner in which a Schenkerian description of tonal structure arises. In this way, GTTM becomes a fundamentally anti-Schenkerian description of tonal structure.

However, its preference-rule based system makes GTTM a fundamentally anti-Chomskyan theory of grammar too. For example, the emphasis on perception, and the picture of tonal structure this gives rise to, reveals GTTM’s closer affinity to a general theory of perception, exemplified by certain theories of vision, than to generative linguistics, as Lerdahl and Jackendoff say themselves (Lerdahl and Jackendoff (1983): 302-307). GTTM’s anti-Chomskyan orientation can also be seen in its constraint-satisfaction approach to tonal grammar, in which multiple (preference rule-based) constraints are in operation, and must be satisfied, *simultaneously* when listeners hear a musical passage. (Although some of these constraints are weighted more heavily than others.)⁸² This approach is similar to Optimality Theory’s proposals regarding *linguistic* grammar, which is a relatively recent branch of linguistic theory that arose not from generative grammatical theory, but from phonology (i.e. the study of speech *sounds*). Lerdahl and Jackendoff acknowledge the overlaps between GTTM and Optimality Theory too (see Lerdahl and Jackendoff (1983): xiv), and Ray Jackendoff has explored parallel constraint-based

⁸² For example, GTTM weights more heavily the preference that suspensions should appear on strong beats (Metrical Preference Rule 8), than the preference that the strongest beat in a group of pitches should appear early in the group (Metrical Preference Rule 2) (Lerdahl and Jackendoff (1983): 347-348). This means that in a group of two pitches, if the second pitch creates a suspension above the bass, then the preference will be to accord it a higher status in the metrical structure of the group (meaning that the beat that it occurs on will be considered stronger than the preceding beat), even though this means that the stronger beat appears later in the group (which is a violation of Metrical Preference Rule 2).

architectures in his own, essentially anti-Chomskyan, descriptions of linguistic structure (Jackendoff (1999)).

Despite its origins in both Schenkerian and Chomskyan theory, GTTM's ultimately perceptual orientation led Lerdahl and Jackendoff to make a striking claim about the relationship of music and language, and particularly about projects that attempt to connect the two (like this one):

“Many previous applications of linguistic methodology to music have foundered because they attempt a literal translation of some aspect of linguistic theory into musical terms – for instance, by looking for musical “parts of speech,” deep structures, transformations, or semantics. But pointing out superficial analogies between music and language, with or without the help of generative grammar, is an old and largely futile game. One should not approach music with any preconceptions that the substance of music theory will look at all like linguistic theory. For example, whatever music may “mean,” it is in no sense comparable to linguistic meaning; there are no musical phenomena comparable to sense and reference in language, or to such semantic judgments as synonymy, analyticity, and entailment. Likewise there are no substantive parallels between elements of musical structure and such syntactic categories as noun, verb, adjective, preposition, noun phrase, and verb phrase. Finally, one should not be misled by the fact that both music and language deal with sound structure. There are no musical counterparts of such phonological parameters as voicing, nasality, tongue height, and lip rounding.” (Lerdahl and Jackendoff (1983): 5)

This is, of course, openly militates against the Identity Theses for music and language being advocated by this dissertation. Therefore, this passage requires a careful and detailed response if a joint Minimalist Program for music and language is to hold any water at all.

I will argue, however, that essentially every point made by Lerdahl and Jackendoff in the above passage can be invalidated, and specifically from the joint Schenkerian/Chomskyan view that they reject. (So, in many ways one could conceive of this dissertation as being a Minimalist, Schenkerian (and specifically Keiler-ian) rebuttal to the above quotation.) The next chapter will directly challenge Lerdahl and Jackendoff's claim that music has no parts of speech, deep structures, and especially transformations, which is a discussion that will continue into the third chapter, and therefore the entire first half of the dissertation. The second half of the dissertation will take up the issue of musical meaning and sound, in comparison to linguistic meaning and sound, and will specifically challenge the notion that musical semantics is in “no sense comparable to linguistic meaning”. I will argue instead that Lerdahl and

Jackendoff's interpretation of "meaning" is suspect, mainly on the grounds that there is no more reason to believe that linguistic meaning includes notions of sense and reference or analyticity than musical meaning does – as Minimalist linguistics asserts – implying that musical meaning can be compared to linguistic meaning, as long as we can defend some notion of a musical LF (which is the level of grammatical representation where the issue of linguistic meaning becomes relevant in generative theory). Along similar grounds, I will argue that though Lerdahl and Jackendoff are correct in asserting that there are no musical counterparts to phonological phenomena like voicing etc., the study of these phenomena properly belongs to a study of the grammar-*external* conditions imposed by phonology on grammar at a PF level of representation – meaning that these phenomena are not aspects of either language or music's internal, organic, computational form. This suggests that Lerdahl and Jackendoff's assertions have no bearing on the identity of the form of music and language, which is, after all, the locus of the musilanguage hypothesis.

What is interesting for our present purposes though, is that GTTM is a theory that developed out of both Schenkerian music theory and Chomskyan linguistics, but then grew to reject both paradigms, and for similar reasons (i.e. given its emphasis on perceptual and constraint-based architectures of grammar). This demonstrates how anti-Schenkerian and anti-Chomskyan proposals often reject both paradigms on similar grounds, while also demonstrating the centrality of both paradigms to their musical or linguistic proposals. This presents even more evidence for Identity Thesis A, i.e. music theory and linguistic theory are identical, not only because of the overlaps between Schenkerian and Chomskyan theory, but because of the centrality both of these paradigms have in their respective disciplines, and the similar ways in which both anti-Schenkerian and anti-Chomskyan proposals reject these paradigms.

I believe one can isolate four different attitudes within both music and linguistic scholarship that reflects the above phenomenon – and which have therefore stood in the way of a further rapprochement between Schenkerian and Chomskyan scholarship. In other words, these are the four attitudes that eventually led to the demise of the Princeton Schenker Project in the late 1980s. We might refer to these

four attitudes as the four Ps of anti-Schenkerianism: (1) perception, (2) physicalism, (3) pedagogy, and (4) poetics. I will end this chapter with a brief examination of these four categories.

(1) Perception: We have already seen the role of perception-oriented theories in problematizing the Schenkerian/Chomskyan approach to issues in musical and linguistic structure, since GTTM is itself the best example of such a theory. But there is more to be said on the matter. For example, consider Fred Lerdahl's assertion that GTTM is essentially a "listening grammar" of how people comprehend heard musical surfaces (Lerdahl (1992): 102). Listening grammars can be contrasted with "compositional grammars" that govern how musical passages are constructed, and which can often be idiosyncratic and subjective depending on what kind of musical passage a composer wants to create. Moreover, compositional grammars can be artificially created and have no basis in human nature or how the mind works. Lerdahl believes that such artificial compositional grammars underlie much of the avant-garde Western art music of 20th century that many people have a hard time comprehending. Therefore, such grammars should not be the focus of a scientific, psychological study of how people hear and comprehend music. This is what motivated Lerdahl and Jackendoff to pursue a listening grammar approach in their theory of tonal structure instead, since such a grammar can model the psychology of human musical perception, and can be tested in scientific experiments too (e.g. Lerdahl and Krumhansl (2007)).

Now, Schenkerian theorists have often described their system as a theory of perception, viz. one that reveals *audible* hierarchical relationships in tonal passages. (Witness the title of the major text by the eminent Schenkerian theorist, and Schenker disciple, Felix Salzer in this regard, viz. *Structural Hearing* (Salzer (1962)).) However, if the way I have been describing Schenkerian theory so far is correct, then a Schenkerian analysis of a Western tonal passage is really a description of the hierarchical structure that a listener *unconsciously* recovers when they hear the passage, but which they might not be able to describe explicitly in the way a theorist can. This is similar to a linguist's description of the hierarchical organization of a sentence, into complementizer phrases, tense phrases, adjuncts and so on, that a person unconsciously recovers when they hear someone else utter that sentence, even though neither speaker nor

hearer might be consciously aware of what a complementizer phrase is, without being a linguist him or herself.

In this sense Schenkerian theory is not a theory of musical perception, but a theory of the *knowledge of musical structure that underlies our perception* of such structures. Now, we might be able to devise an experiment that assesses how people perceive musical passages on the basis of this knowledge that Schenkerian theory ascribes to them (or that Schenker himself ascribed to only the great composers) – but this would ultimately be peripheral to the core claims of the theory. In fact, this is also why the notion of *reduction*, which is often associated with Schenkerian theory, is of only secondary importance to that of *generation* within this theory. Reduction is closely related to perception – when an analyst reduces a musical passage, s/he reveals explicitly the same hierarchical organization of that passage that a listener recovers unconsciously through the act of perception, as I just suggested. And this is what happens much of the time in Schenkerian scholarship too – Schenkerian theorists usually focus their energies on creating reductions of interesting musical passages to reveal their underlying harmonic and contrapuntal structure. But it is not as if the end result of these reductions is unknown – a reduction is normally expected to reveal the I-V-I *Ursatz* of the passage. So, in this sense, the reduction of a musical passage is secondary to, and predicated on, the generation of that passage in the first place, from its intuitively known *Ursatz*.⁸³ This in fact why, as Allan Keiler has noted, Schenker labeled his analyses of passages starting from the *Ursatz* and moving up, rather than down from the surface (Keiler (1983-84): 201-207) – which just reinforces the generative, as opposed to reductive and/or perceptual, orientation of the theory.

This is also why Schenkerian theory is so different from more reduction-oriented theories of musical structure, where a description of structure normally starts with the surface and then locates deeper patterns and principles of organization from the surface. And this is why a Schenkerian approach to

⁸³ This accords more generally with the Rationalist and nativist aspects of generative theory, which believes that humans are equipped with innate knowledge of how their native languages and musical idioms work, which allows even little children to acquire competence in them. We could extend this to Schenkerian theory, and say that humans are innately equipped with knowledge of the grammatical principles that generate musical surfaces from underlying *Ursatz* representations – so that the act of reduction merely reveals what we already know *intuitively*.

comparing musical and linguistic structure is so different from that of GTTM. Given the latter's emphasis on notions like "time-span *reduction*" and "prolongational *reduction*", the notion of *generation* almost never figures in Lerdahl and Jackendoff's approach – despite the word being part of the title of their classic text.⁸⁴ This point was noted by the Schenkerian music theorist David Beach in the early days of the publication of Lerdahl and Jackendoff's text, when he said that "Jackendoff's and Lerdahl's [*sic*] theory is reductive, taking the musical surface as their point of departure rather than generating the surface from the background" (Beach (1985): 294), and Lerdahl and Jackendoff seem to have accepted this themselves (e.g. in Lerdahl and Jackendoff (1985): 147-148), even though they still insisted on describing their work as constituting a generative theory, because its ultimate goal – i.e. to give structural descriptions of musical passages – was shared with Chomskyan generative theory.

However, another Schenkerian, Matthew Brown, has criticized basing a theory of musical grammar on a perceptual foundation, which he states in terms of Lerdahl's own duality of composing-versus-listening:

"Whereas Schenker saw his goal in compositional terms as creating a production system for generating an infinite number of tonal pieces, [Lerdahl and Jackendoff] see their goal in psychological terms as a means for describing how people represent tonal relationships when they listen. There are good reasons, however, why we should not separate these activities too sharply. For one thing, we have no reason to suppose that when expert composers listen to music, they process their knowledge in different ways than when they compose. On the contrary, the evidence suggests that expert listening requires similar mental representations to expert composition." (Brown (2005): 217-218)

The psychologists Johan Sundberg and Björn Lindblom have also criticized Lerdahl and Jackendoff's lack of emphasis on composition, as preventing GTTM from being able to describe specific compositional idioms, and which musical structures are acceptable within these idioms (Sundberg and Lindblom (1991): 260), which is the requirement of descriptive adequacy that a generative theory should fulfill. Building on this critique, the linguists Jonah Katz and David Pesetsky have suggested more recently that GTTM is not even a theory of generative musical grammar but rather a "generative parser ... explicitly designed to produce a well-formed parse (or set of parses) for any sequence of sounds" (Katz

⁸⁴ Interestingly, the initial, working title of GTTM was *A Formal Theory of Tonal Music* – identical to the final title in all but the use of the word "generative" (see Lerdahl and Jackendoff (1977): 170, note 21).

and Pesetsky (2011)). This implies that GTTM's preference rules can assign a structure to essentially any musical surface, in any idiom – and therefore does not distinguish what is grammatical in one idiom from another. This is particular problematic when it comes to constructed, 'poetic' languages like the post-tonality of the Second Viennese School, which should not be amenable to a generative description – any more than Elvish, the poetic language of the elves in Tolkien's *Lord of the Rings*, should be.

Most significantly, Allan Keiler has even compared reductive systems like GTTM more generally to a set of discovery procedures, of the kind seen in the older, pre-generative, structuralist tradition in linguistics, especially of the kind associated with the work of Leonard Bloomfield (Keiler (1978a): 181; see also the debate between Keiler, and Lerdahl and Jackendoff (Keiler (1979-80), Jackendoff and Lerdahl (1980)), in which Keiler makes similar points).⁸⁵ The structuralist tradition, unlike the Rationalist, generative tradition that followed it, based itself on the belief that the underlying grammatical structure of sentences in a language must be discovered empirically, rather than as something we innately know. Therefore, such a discovery should start from a 'neutral' surface sentence structure, about which no prior assumptions have been (or can be) made, to reveal its underlying organization by means of the aforementioned discovery procedures.⁸⁶ In this sense, discovery procedures are not like the rules of generative grammar (such as the Phrase Structure Rules discussed in Chomsky (1965)), from which one can generate a description of the hierarchical tree structure of a sentence that native speakers of a language *ex hypothesi* already know.

⁸⁵ It is worth noting that this debate took place not in a music cognition journal, but in *Perspectives of New Music*, which was the journal of record of mid-century Princeton theory, i.e. the journal famous for publishing, for example, Benjamin Boretz's *Meta-Variations*, and other papers in that Princeton tradition. This re-affirms the origins of Keiler, and Lerdahl and Jackendoff's, projects in the Princeton Schenker Project. Incidentally, Keiler's points about discovery principles have been explored by another Schenkerian theorist, Kofi Agawu, as well, in Agawu (1989).

⁸⁶ Perhaps it is not surprising in this regard that many of the attempts to apply ideas from structuralist linguistics to music have focused on idioms that are obscure, such as avant-garde Western art music (exemplified by Jean-Jacques Nattiez's analysis of Edgar Varese's *Density 21.5* (Nattiez (1982)), or non-Western idioms whose underlying structure is assumed to be unknown or unfamiliar to the foreign, usually Western, music scholar (such as Simha Arom's noted study of Central African 'pygmy' music (Arom (1991)), and Nattiez's studies of Canadian Inuit music (Nattiez (1983a, b, 1999)). This is in stark contrast to the exclusive focus of generative theorists like Schenker, and Lerdahl and Jackendoff, on their own native musical idiom of Western Classical tonality – comparable to early work in generative linguistics, which focused just on Chomsky's native language of English, and up to a certain extent Hebrew.

Noam Chomsky points out that one of the problems with structuralism was that it had no way of validating one set of discovery procedures, based on whether it accounts for facts that native speakers innately know to be true better than another set of procedures, because it does not assume such innate linguistic knowledge on the part of native speakers and hearers, or even linguists for that matter, to begin with (Chomsky (1957): 13, 50-55). In Chomsky's opinion, this weakness led ultimately to the downfall of the structuralist tradition, and its replacement by the generative approach instead – which is a 'paradigm shift' seen within the work of those music scholars who initially took a structuralist approach too, only to switch to a generative one later (for example, Ruwet (1987); see also Allan Keiler's critique of the structuralist tradition in music, in Keiler (1981): 139-151).

However, Allan Keiler argues, based on an earlier suggestion of Chomsky's, that discovery procedures always had a psychological flavor to them, even though they did not make the kinds of Rationalist assumptions about the mind and its innate knowledge of language that generative theorists did later (Keiler (1979-80): 511-514, Keiler (1981): 140-142). This is seen, for example, in the general structuralist assumption that speakers comprehend the grammatical structure of sentences based on facts of speech perception. In this sense, analyzing the structure of a linguistic sentence through discovery procedures is similar to the analysis of a musical passage through a reductive procedure that is founded in facts about musical perception – which is exactly the kind of approach seen in Lerdahl and Jackendoff's work, hence Keiler's above comparison of these two approaches. The problem here though, again, is that without any assumptions about how innate knowledge of how either music or language guides our perception of musical or linguistic structures, attempts to analyze linguistic or musical passages by working down from their surfaces will not be justifiable. This is why an approach to musical or linguistic structure that focuses on reduction over generation, and on perception over *cognition* (i.e. musical knowledge – especially unconscious intuitive knowledge, of the *Ursatz* for example, and the grammatical principles from which it is generated) is problematic. As Keiler says:

“The danger is for them [i.e. musical analysis and perception] to be necessarily related in the wrong way, especially when premises about the very nature of, for example, musical (or linguistic) perception (and

not intuitions, which are another matter) become the basis for constraining the notion of grammar in the first place.” (Keiler (1979-80): 514)

This is precisely why reduction takes second place to generation in Schenkerian theory. And this is also why any attempt to evaluate Schenkerian theory as a theory of perception, or of reduction, or as a listening grammar of music (for example, as the music theorist David Temperley does in a recent critical evaluation of Schenkerian theory (Temperley (2011): 157-163) is unfair – it attempts to evaluate Schenkerian theory for something it does not attempt to be in the first place, and probably should not be anyway, given the above problems with a purely reductive/perceptual approach to music.

Now, it is understandable why a scientifically-oriented music scholar would be interested in empirically-testable hypotheses about musical structure, and why, as a result, s/he would choose to ground a scientific research project on musical structure in testable facts about human perception. So, it is quite understandable why Lerdahl and Jackendoff distanced themselves from a more Rationalist Schenkerian approach to music in GTTM. But it is also not the case that such a Rationalist approach is totally divorced from ‘the facts’. In fact, one could say that the wealth of actual music that music theorists, and especially Schenkerian music theorists spend their time analyzing is itself a treasure chest of empirical facts about how humans make music – which is as large and complex a dataset as any scientist could hope for. In the last chapter of this dissertation, I even make the case for musical analysis as a form of experiment. This is the exact argument that generative linguists subscribe to, in defending their approach to the study of language, which is almost entirely based on similar *linguistic* analyses of sentences across languages. (This is opposed to basing a study of language on facts about human speech perception and so on that forms the basis for many empirical research projects pursued by cognitive psychologists, evolutionary biologists, neuroscientists and the like.)

Finally, a scientific approach that does not get to the bottom of what it is trying to study ceases to be of any critical importance, even if it does pursue empirically admirable goals of rigor and objectivity. So, an empirical music-theoretic project is not necessarily a better one to follow, compared to the

Rationalist scientific projects inherent in Chomskyan theory, and (the Princeton Schenker Project's interpretation of) Schenkerian theory. Noam Chomsky's words on this topic are prescient:

“One whose concern is for insight and understanding (rather than for objectivity as a goal in itself) must ask whether or to what extent a wider range and more exact description of phenomena is relevant to solving the problems that he faces. In linguistics, it seems to me that sharpening of the data by more objective tests is a matter of small importance for the problems at hand. One who disagrees with this estimate of the present situation in linguistics can justify his belief in the current importance of more objective operational tests by showing how they can lead to new and deeper understanding of linguistic structure. Perhaps the day will come when the kinds of data that we can obtain in abundance will be insufficient to resolve deeper questions concerning the structure of language. However, many questions that can be realistically and significantly formulated today do not demand evidence of a kind that is unavailable or unattainable without significant improvements in objectivity of experimental technique.” (Chomsky (1965): 21)

(2) Physicalism: Another resilient paradigm in both music and language scholarship has been that of physicalism. Physicalism takes musical structures to be physical objects that can be studied using the language of physics, i.e. mathematical formulas and equations. In music, this physicalist paradigm has a specific connection to Princeton theory. Much of the music that interested the Princeton composer-theorists was high modernist of course, for which the methods and concepts of traditional tonal theory were often inadequate – and which, therefore, merited the development of a new set of music-analytical tools, methods and theories. Given the highly systematic nature of this music (particularly in its total serialist manifestation), and given the formalist nature of much of the music-theoretic discourse at Princeton during that time, it is perhaps unsurprising that some of these new methods and theories ended up being borrowed from mathematics – first in the development of musical pitch-class set theory (e.g. in Forte (1973)), and then in the appropriation of group theory into musical transformational theory (e.g. in Lewin (1987)).

Now, we have seen, in the case of language, that Minimalism takes the study of linguistic structure to be a biological science, i.e. biolinguistics. This makes the reduction of the study of language to a *physical* science problematic, given Minimalism's argument that contemporary physics cannot adequately describe the underspecified and economical aspects of linguistic structure. This is why generative linguistics has focused so much on the *psychological* aspects of linguistic structure instead,

such as the foundation of generative grammar in innate knowledge, or in the connection of grammar to meaning. If we accept the Schenkerian idea that music has an internal, abstract, psychological form too, then the reduction of the study of musical structure to a physical science becomes problematic as well.

Some mathematically inclined music theorists, however, have persisted with purely mathematics-driven descriptions of musical structure, including ones based on concepts used specifically in contemporary physics (e.g. the use of the topological concept of an “orbifold” by Tymoczko (2006), which is a concept that is used commonly in string theory). But there have been other mathematical theorists who have been more conscious of a need to ground their descriptions of musical structure in human psychology, although many of their proposals have a particularly Princetonian attitude, perhaps given the origins of many of these theories in the Princeton tradition of music theory. For example, one of the great Princeton composer-theorists, and the originator of musical transformational theory, David Lewin, introduced a characteristically *phenomenological* attitude to mathematical music theory that was reminiscent of Princeton theory (e.g. see Lewin (1986)).

But the grounding of such theories, however formal and rigorous, outside of a psychological *science* of music (as opposed to a phenomenological critique of music), has sometimes created rather strange bedfellows between physicalist music theories and other perspectives on music that are often quite unscientific. A case in point is a contemporary branch of musical transformational theory that evolved in the 1990s out of David Lewin’s ideas, based heavily on the writings of Richard Cohn (Cohn (1996, 1997, 1998)) that has come to be known as *neo-Riemannian* theory (although this is a term that Cohn has disputed himself (Cohn (2012): xiii). Neo-Riemannian theory has engaged with music from the Western tonal common-practice much more frequently than earlier transformational music theories did (e.g. see Cohn (1999)), but from a perspective that is quite distinct from Schenkerian approaches to this repertoire. Indeed, the emphasis of much of Cohn’s writing does seem to be on tonal grammar (e.g. Cohn (2012): 13-

16), but from a perspective that problematizes Schenkerian readings of certain tonal passages (e.g. Cohn (2012): 45, (1999): 220, (1992b)).⁸⁷

However, this emphasis also requires negotiating the rich, expressive possibilities of tonal music too – in fact, this is what often gives rise to the very “purple patches” (to use Donald Francis Tovey’s term) in certain tonal pieces that invite such divergent music-theoretic interpretations, as those coming from Schenkerian and neo-Riemannian theory. Given its organicist foundation, Schenkerian analyses of such passages normally understand their expressive aspects in internalist terms too, i.e. as emerging from the conditions of possibility laid out by the *Ursatz* itself. In other words, Schenkerians normally describe how the expressive interest of such passages arise in terms of how they are generated from the *Ursatz* via the principles of harmony and counterpoint – which suggests a mapping between tonal grammar and tonal meaning (an idea I will explore more in chapter 2.1, as suggesting a musical equivalent to the LF level of structure proposed by Minimalist linguistics).

But lacking such an organicist foundation – or any broader consideration of the psychological aspects of music – physicalist theories of music have often had to describe the meaning of certain problematic musical passages in terms that do not emerge, organically, from within their chosen theoretical paradigm itself, and which, therefore appropriate ideas from other paradigms of scholarship. And this has led to some surprising borrowings. For example, despite the mathematical rigors of neo-Riemannian theory, Richard Cohn has described musical signification in psychoanalytical terms (in Cohn (2004)), which is an approach more commonly associated with postmodernist deconstruction (e.g. see Derrida (1987)). Henry Klumpenhouwer, one of the major figures in transformational theory (and arguably David Lewin’s most well-known protégé) has even observed psychoanalytical themes in

⁸⁷ It might be argued though, that neo-Riemannian theory’s aim is not to model tonal grammar specifically, but rather to model general musical features that might be found even in hypothetical musical idioms – as opposed to the ‘natural’ musical systems found across the world’s cultures (see Cohn (1997): 23). In this regard, the systems being modeled by neo-Riemannian theory might be considered “compositional grammars”, in Fred Lerdahl’s sense of the term, i.e. musical systems that are idiosyncratic, and represent the individual will of a composer, rather than general properties of tonality.

Lewin's phenomenology, in what he considers the latter's attempt to distance himself from Cartesian cognitivism of the kind found in generative theory (see Klumpenhouwer (2006)).

However one evaluates physicalist music theories (including their descriptions of musical meaning), what is interesting in terms of our current musilinguistic concerns is how much these approaches parallel similar approaches in *language* scholarship. I have already suggested how physicalist music theories like neo-Riemannian theory (or at least large parts of it) are explicitly anti-generative, and specifically anti-Schenkerian, mainly based on sentiments expressed by Richard Cohn himself, and other thinkers who have been influenced by his ideas, or maintain a physicalist attitude in their music-theorizing (e.g. see Cohn (1992a), or Tymoczko (2011): 19).⁸⁸ But there has *also* been a consistent strain in language

⁸⁸ Many of Dmitri Tymoczko's ideas in this regard seem to be based, however, on a misunderstanding, and/or a mischaracterization of, the generative project in music. For example, he claims that Fred Lerdahl's conception of musical structure, based partly on Schenker's ideas as we saw above, is based on "a kind of lossless listening in which ordinary people typically recover *all* of the details in a musical score" (Tymoczko (2011): 424). However, Lerdahl never asserts anything of the sort, and it would be strange to assert something like this too, since there are all kinds of details in a musical score – stylistic, idiosyncratic, performance-practice related, figurative – that no one could possibly perceive in a score, certainly not without extensive practice with it. All that Lerdahl asserts is that the competent listener has an innate knowledge of certain grammatical conditions that allows him/her to assign a structure to (a hearing of) the score, and which only *then* creates the possibility that the listener could find other details of interest in the score – presumably based on multiple, culturally-situated re-hearings of it. In other words, Lerdahl is only asserting what the listener who is musically *competent* knows (although he and Jackendoff disagree with Schenker on what this knowledge is, since they include, for example, a metrical component in it). Lerdahl never asserts that the competent listener must also be able to recover all the *performance* details in a score (i.e. as opposed to "competence" in the Chomskyan sense), since that would be too much to expect of even the most experienced listener, and impossible to demonstrate empirically.

Tymoczko's critique of Schenkerian theory seems to be based on an incorrect understanding of the latter's ideas as well. He attempts to describe (and discredit) aspects of Schenkerian theory based on a false dichotomy between *harmonic* explanation and *contrapuntal* explanation, with the further evaluation that contrapuntally-inclined Schenkerians are theoretically inept because of their inability to explain certain phenomena that clearly require a harmonic explanation – which Tymoczko is biased towards himself (see Tymoczko (2011): 258-267). But there is no such divide between harmonic and contrapuntal explanation to begin with, since they are flip sides of the same coin (e.g. a diminished triad can be thought of as a self-standing harmonic entity, and simultaneously as a contrapuntal entity in which a tritone is added above a bass through voice leading) – and Schenkerian theorists happily invoke both forms of explanation depending on the situation. For example, a contrapuntal explanation is clearly preferable in describing, diachronically, certain harmonic phenomena that arise out of the 'non-harmonic' practices of the Renaissance (such as the ii^{6/5}-V-I Baroque cadential formula, see Aldwell and Schachter (2011): 209-210, Gauldin (1995): 138-139), which can be seen in Schenker's advocacy of species counterpoint too. Tymoczko discredits this as a problematic "monism", which rejects harmonic explanations for such phenomena. However, harmonic explanation also clearly *does* play a role in Schenkerian theory, e.g. in describing the origins of the *Ursatz* as a I-V-I structure, and in describing how other harmonies interact with the I and V of the *Ursatz* to create certain harmonic progressions – a topic that concerned Schenker enough that he fleshed it out in book-length form in the *Harmonielehre* of 1906. Moreover, harmony and counterpoint *are* separated in Schenkerian theory, for expository simplicity, in describing how harmonic events are composed-out *via* voice leading. But the same

scholarship that has taken a more mathematical approach to issues of linguistic structure than the psychological one inherent in generative linguistics – and much of which has been devoted to challenging Noam Chomsky’s specific proposals about language. An example is the linguistic subdiscipline devoted to the study of linguistic meaning called “formal semantics” (e.g. see Partee (1996)). It might be unfair of me to characterize this paradigm as physicalist and anti-Chomskyan – but formal semantics arose out of the logician Richard Montague’s work in the 1960s, which attempted to describe how a language like English can be described as a formal (i.e. logical) system, and which was influenced by his earlier work in set theory (particularly as a doctoral student of the great mathematician and logician Alfred Tarski) – which suggests that formal semantics is at least closer in spirit to the “formalizing Schenker” paradigm in music theory than it is to the “naturalizing Schenker” one advocated by this dissertation. Also, Montague’s work was not conceived as a response to Chomskyan linguistics, and the formal semanticist Barbara Partee has even described her early interests in linguistics as being sympathetic to, and influenced by, Chomskyan work in grammar (Partee (2004): 2-5). However, Chomsky has criticized formal semantics as being too focused on a logical reconstruction of language rather than on the natural form of language (Chomsky (1975, 1980b)) – which is exactly the criticism I made of the physicalist approach to musical structure above – and this has motivated responses to Chomsky by scholars like Partee (e.g. in Partee (1975), Bach and Partee (1980)), who even says that:

(harmonic) entity that is composed out via voice leading at one level of structure (e.g. a V that is composed out by a cadential 6-4), can itself compose out, as a *voice-leading* entity, a different harmony, *at a deeper level of structure* (such as the final tonic triad of the *Ursatz*). So to maintain the harmony/counterpoint distinction too strongly involves over-simplifying things. (Allan Keiler even criticizes Schenker for being inconsistent in this regard, i.e. in failing to explain sonorities at shallower levels of structure in harmonic terms, in the way he does with sonorities deeper down, even though they can *all* be explained in both harmonic and contrapuntal terms (Keiler (1977): 12).) For this reason, many Schenkerians find harmony and counterpoint to be inseparable – which Tymoczko rejects as an untenable “holism”, preferring the “pluralism” of the previous point, where harmony and counterpoint are separated. But as should hopefully be clear now, Schenkerians can subscribe to all three positions of monism, holism, and pluralism simultaneously, without any contradiction, depending on whether they are interested in the diachronic explanation of how a certain tonal phenomenon evolved versus the synchronic explanation of that same phenomenon within Schenker’s prolongational, *Ursatz*-based model of grammar, or depending on whether they are interested in explaining a background phenomenon versus a foreground one. But by creating false distinctions between these categories – which should be understood in their proper context, i.e. within the hierarchical description of tonal phrase structure (which he rejects too) – Tymoczko clearly misunderstands the nuances of generative, and particularly Schenkerian, music theory, i.e. why Schenkerians sometimes insist on contrapuntal explanations for phenomena when they can easily provide a harmonic explanation for these phenomena if they want to as well.

“Non-Chomskyan approaches [to semantics] are more often seen among most (though not all) computational linguists. Although I started out as a syntactician, I haven’t been able to call myself a working syntactician for some time... Since then I have tended to declare myself agnostic about syntax, largely in order to co-exist compatibly with my mainly Chomskyan colleagues in the department and to work with students whose syntax has almost always been Chomskyan... But I have sometimes regretted that we don’t have non-transformational syntactic theories represented in our department... I myself feel more attracted to non-transformational approaches, and I’ve been sorry not to be part of a community of colleagues working with theories of that kind.”⁸⁹

The physicalist aspects of formal semantics can be seen even more explicitly in some of the formalisms that emerged later, and were partly influenced by it, such as the Generalized Phrase Structure Grammar developed by Gerald Gazdar and his colleagues (Pullum and Gazdar (1982), Gazdar et al. (1985)), which later gave rise to Carl Pollard and Ivan Sag’s Head-Driven Phrase Structure Grammar (Pollard and Sag (1994)). For example, Gazdar and his colleagues clearly reject the psychological basis of their theory, claiming that mathematical precision is their motivation instead:

“We make no claims, naturally enough, that our grammar is *eo ipso* a psychological theory. Our grammar of English is not a theory of how speakers think up things to say and put them into words. Our general linguistic theory is not a theory of how a child abstracts from the surrounding hubbub of linguistic and nonlinguistic noises enough evidence to gain a mental grasp of the structure of a natural language. Nor is it a biological theory of the structure of an as-yet-unidentified mental organ... Thus we feel it possible, and arguably proper, for a linguist (qua linguist) to ignore matters of psychology... If linguistics is truly a branch of psychology (or even biology), as is often unilaterally asserted by linguists, it is so far the branch with the greatest pretensions and the fewest reliable results. The most useful course of action in this circumstance, is probably not to engage in further programmatic posturing and self-congratulatory rhetoric of the sort that has characterized much linguistic work in recent years, but rather to attempt to fulfill some of the commitments made by generative grammar in respect of the provision of fully specified and precise theories of the nature of languages.” (Gazdar et al. (1985): 5)

Their mathematical commitments can be seen further in their statement that:

“The body of chapters 2, 3 and 4 is relatively informal, having been kept as free as possible of mathematical notation, but each chapter has a final section in which the crucial concepts are formalized with much greater precision.” (Gazdar et al. (1985): 11)

⁸⁹ In Barbara Partee, “Reflections of a Formal Semanticist as of Feb 2005”. This is the online version of Partee (2004): 1-25. The above quotation was excised from the version published online a year later. It is worth noting that Partee’s observation about computational linguists being mostly anti-Chomskyan is consistent with the recent history of music theory. With the exception of the Princeton Schenker Project formalists, the interest in computational music theory, including some popular current trends in computer-aided corpus analysis (e.g. Rohrmeier and Cross (2008), DeClercq and Temperley (2011), Quinn and Mavromatis (2011)), and the probabilistic modeling of musical pieces (e.g. Temperley (2004a, 2007, 2009)), all seem to be essentially anti-Schenkerian in their attitude.

What is most relevant for our present purposes though is the fact that these later formalisms were explicitly anti-Chomskyan in many ways too, including in their rejection of the transformational aspect of Chomskyan theory (Pollard and Sag (1994): 2), and as can be seen in Gazdar et al.'s general dismissal of Chomskyan, naturalistic approaches to language above as having the "greatest pretensions and the fewest reliable results".

A final physicalist paradigm in language scholarship worth discussing briefly is also the most mathematically-oriented of them all. Unlike the above approaches from within formal semantics, or phrase structure grammar (which at least emerged from, or had ties to, generative theory), there have been formalisms that have emerged directly from within the mathematical sciences themselves, such as Martin Nowak's work in mathematical biology (e.g., Nowak (2006)). Nowak has explored a variety of issues in biology using mathematical models, and particularly issues in evolution – a field called "evolutionary dynamics". Interestingly, some of this work has involved the study of language evolution too. But the approach to the mathematical investigation of language evolution here is one that rejects the Chomskyan claim, which we explored in section 1.1.2, that a search for (particularly Darwinian) explanations for the origins of language is misguided (e.g. see Nowak, Plotkin and Jansen (2000)).

The moral of the story from the preceding pages therefore seems to be that very similar, mathematically-driven physicalist paradigms exist in both music and language, but both of which similarly reject psychological explanations for music and language's origins and structure, specifically the kind of psychological explanation inherent in generative theory. Moreover, much of the scholarship in these physicalist paradigms seems to be aimed at challenging either Schenkerian proposals about musical structure or Chomskyan proposals about linguistic structure. If anything, this seems to re-affirm the overlap between certain modes of reasoning in both musical and linguistic scholarship, and the central role of Schenkerian/Chomskyan thought in all of this – which seems to provide just more evidence for the identity of music and language, or at least of their theories.

(3) Pedagogy: Like perception and physicalism, the third anti-generative paradigm in music/linguistic scholarship that I would like to discuss is also generally scientific in its orientation, and specifically cognitive as well. In its musical manifestation, however, and unlike the two preceding paradigms (and unlike generative theory), this paradigm is more attentive to the historical and stylistic aspects of music. That is, it is more concerned with music as the locus of historically- and culturally-situated communication – and therefore something that must be *learned* from the contexts in which it is made. The result is a particularly Empiricist paradigm in music scholarship, i.e. one that is opposed to the idea of musical knowledge being innate, in contrast to the Rationalist and nativist positions that generative theory takes on this issue. The emphasis on learning in music has also made some scholars within this paradigm focus specifically on the issue of music *pedagogy* as well, and so this is the label by which I shall refer to this area of music-theoretic scholarship. (Not all theorists in this paradigm have been explicitly interested in the issue of pedagogy, so this label is a bit of a misnomer – an alternative might be to call it the tradition that theorizes musical *performance* (i.e. as opposed to competence, in the Chomskyan sense), given its interest in the broader, culturally-situated, aspects of musical communication. However, this term is equally if not more confusing, since it can be misunderstood as having to do with performance in its ‘artistic’ sense, i.e. what musicians study in a musical conservatory and do on stage – hence my commitment to the, admittedly problematic, label of “pedagogy”, in keeping with the “p” theme I have been outlining here.)

I see this pedagogically-oriented paradigm in cognitive music theory as emerging primarily in the works of the eminent composer, philosopher and music theorist Leonard Meyer. Meyer taught for many years at the University of Chicago, and many of the theorists within this paradigm have been his students, or students of his students, at Chicago. So, we might refer to this third branch of anti-generative music scholarship as the “Chicago School of music theory” too – especially since its ideas continue to be developed at the University of Chicago in the research of Lawrence Zbikowski, who has proposed a culturally-situated cognitive theory of musical structure (e.g. see Zbikowski (2002, 2008)) that is influenced explicitly by language scholars working in the anti-generative “cognitive linguistics” tradition,

such as George Lakoff (1990). However, there have been other insitutional homes for pedagogical music theory, such as the University of Pennsylvania, where Meyer taught for many years after his tenure at Chicago, as did his student Eugene Narmour, who did his doctoral work with Meyer at Chicago. Another more recent home for Meyer-ian theory is Northwestern University, where a large, essentially anti-Schenkerian cognitive music theory program has developed in the last ten or fifteen years, seen most prominently in the work of Robert Gjerdingen, who also happened to be a doctoral student of Meyer at the University of Pennsylvania.

Given its focus on music cognition, many scholars in the pedagogical tradition have proposed cognitive models of music. But given their Empiricist foundations, these models usually reject a hierarchical understanding of musical structure in the way that more Rationalist, and particularly generative, approaches do. Instead, they understand music in terms of how it is processed and learned from a perceived musical stimulus, as Empiricists do more generally – which leads to a greater emphasis on the musical surface in this paradigm, and specifically an emphasis on how musical surfaces are perceived (as opposed to being generated from, or structured in, abstract hierarchical terms). This can be seen in Eugene Narmour’s perceptual theory of melody (Narmour (1990)), which models melodic perception in terms of five principles of melodic structure,⁹⁰ and in Justin London’s perceptual theory of rhythm (London (2004)). (London is another Leonard Meyer protégé, having done his doctoral work with Meyer at U. Penn too.)

Both Narmour and London’s theories model musical perception on the basis of how a listener thinks a given musical surface will proceed, which reveals a certain real-time, statistical (and particularly

⁹⁰ These five principles are (1) Registral Direction, which governs what direction (upward or downward) a melody will continue in, given the size of the preceding interval, (2) Intervallic Difference, which governs the size of a realized melodic interval on the basis of the size of the previous interval, (3) Registral Return, which requires that the difference between the first tone of an interval that requires continuation and the second tone of the interval that continues it should not be more than a major second, (4) Proximity, which governs the conjunct-ness of a melody, leading to a greater expectation for smaller melodic intervals and, (5) Closure, which governs differences in the registral size and direction of an expected and a realized interval. As is evident, this model is founded strongly on the belief that aspects of musical structure, e.g. the pitches in a melody, create *expectations* about the other structural elements that will follow them, which is based on ideas Leonard Meyer borrowed from Gestalt psychological theories of perception (in Meyer (1956)). Narmour further developed this idea in the above theory, to suggest that melodic pitches and intervals have certain *implications* that are subsequently *realized* or *denied* – leading to the “Implication-Realization” title of his model.

probabilistic) approach to musical perception. This is an approach shared with other, statistically-inclined and anti-generative, approaches in music (e.g. Temperley (2007)), and as we will see in a moment, many Empiricist approaches to language learning too. So, even though these are theories of perception, they are markedly different from the perceptual theory proposed by Lerdahl and Jackendoff in GTTM, since the latter models how listeners perceive a surface whose structure is *already* known, in a sense, due to a listener's innate preferences for certain (hierarchical) musical structures (that the preference rules model). (Although one could argue that we have a preference for certain musical structures because these are the ones that are made possible by (an intuitively known) grammar – which was not the focus of GTTM of course, but is the focus of a truly generative theory, like Schenkerian theory.)

The above focus on how people overtly hear musical pieces, and how they understand musical structure through experience, also reveals this pedagogical paradigm's focus on issues of style, and on the cultural contexts for different musical listening and learning experiences (see Meyer (1996) for an eloquent and explicit example of this attitude). Such a cultural or stylistic focus goes hand in hand with this paradigm's surface-based, Empiricist attitude to musical learning too. This is because in an Empiricist paradigm the musical surface has to provide a listener with a much richer source of information about that surface's structure, and the structure of that surface's idiom in general, so that the listener can learn that idiom in the absence of any innate knowledge of musical structure (which Empiricists deny listeners). And cues about characteristic stylistic features in a musical surface can provide such a richer source of information about that idiom, hence the emphasis on style in this branch of music theory.

This emphasis can be seen most clearly in Robert Gjerdingen's style-centered approach to issues in music cognition and learning. Much of Gjerdingen's work has focused on the concept of a musical schema, which he began with Gjerdingen (1986), leading up to the recent text *Music in the Galant Style* (Gjerdingen (2007)). In this text, Gjerdingen defines schemata as “stock musical phrases employed in conventional sequences” (Gjerdingen (2007): 6). Moreover, a schema is:

“...a shorthand for a packet of knowledge, be it an abstracted prototype, a well-learned exemplar, a theory intuited about the nature of things and their meanings, or just the attunement of a cluster of cortical neurons to some regularity in the environment. Knowing relevant schemata allows one to make useful

comparisons or, as the saying goes, to avoid “comparing apples with oranges.” Experts in a particular subject may distinguish more relevant schemata than non-experts. Becoming acquainted with a repertory of galant musical schemata can thus lead to a *greater awareness of subtle differences* [my emphasis] in galant music. The music may seem to develop more meaning.” (Gjerdingen (2007): 11)

So, by explicitly learning a set of schemata, a student of music can learn how a certain musical idiom works – i.e. s/he can learn how musical phrases are structured in that idiom, how one phrase is “subtly different” from another. In other words, overt knowledge of schemata can serve as a way of acquiring knowledge of a musical idiom in the absence of innate knowledge of musical grammar.

The Empiricist orientation of Gjerdingen’s proposal is clear – but so are its pedagogical implications, for *learning* a given set of schemata can be a way of *teaching* music students how to write, understand, and perform music in a certain idiom too. In fact, this has been a central theme in Gjerdingen’s recent work, where he has attempted to demonstrate how music students in various pedagogical environments in 18th century Europe learned how to write and comprehend tonal pieces, based on their instruction in certain schematic structures called “partimenti”. A partimento is a:

“...bass to a virtual ensemble that played in the mind of the student and became sound through realization at the keyboard. In behavioral terms, the partimento, which often changed clefs temporarily to become any voice in the virtual ensemble, provided a series of stimuli to a series of schemata, and the learned responses of the student resulted in the multivoice fabric of a series of phrases and cadences. From seeing only one feature of a particular schema – any one of its characteristic parts – the student learned to complete the entire pattern, and in doing so committed every aspect of the schema to memory. The result was fluency in the style and the ability to “speak” this courtly language.” (Gjerdingen (2007): 25)

So, through his theory of partimenti in Classical music pedagogy, Gjerdingen has attempted to give a more historically- and stylistically-informed account of music learning and cognition, than the ones proposed by generative theorists.

The pedagogical benefits of Gjerdingen’s approach are explicit and manifold, which is why the interest in partimenti has seen a revival in contemporary music theory – many theory teachers even using

Gjerdingen's treatise as a textbook in the classroom,⁹¹ continuing a tradition that has existed now for several centuries.⁹² Speaking from a generative perspective, however, there are at least a couple of points that can be made challenging the claims of partimento theory (and schema theory in general) as being a theory of musical learning, or even of musical cognition. For example, could it be the case that musical schemata provide a necessary (but not sufficient) musical stimulus that merely triggers an innate process of musical knowledge acquisition – as opposed to this knowledge resulting directly from *learning* partimenti, and other such schemata? Remember in this regard how many growth processes in the natural world require exposure to some external stimulus for their proper development. For example, a mammal needs to be exposed to light at an early age for its visual system to develop properly. But this does not mean that the animal is 'taught' how to see by the light source – the development of the visual system, and therefore visual knowledge, is largely innate, i.e. specified in the organism's genes. Similarly, one might argue that musical knowledge is innate too (or "intuitive", to use the Rationalist term that even Schenker subscribes to), but needs exposure to some musical stimulus for it to develop properly – which is certainly the argument that generative linguists make about language acquisition.

To develop this point further, consider the fact that many galant partimenti arise from fairly straightforward harmonic and contrapuntal principles. For example, the "Prinner", a galant schema that Gjerdingen named after the 17th century pedagogue Johann Jacob Prinner, involves two voices descending in parallel 10ths, normally to the tonic (Gjerdingen (2007): 45-60). This can be seen in the chord progression IV – I⁶ – V^{4/3} – I, in which the upper voice, i.e. the melody, traces scale degrees 6 – 5 – 4 – 3 (i.e. the pattern la – sol – fa – mi), while the lower voice, i.e. the bass, traces the bass line 4 – 3 – 2 – 1. The pedagogical view here is that a student would memorize this pattern, and then would combine this pattern with other schemata to learn how to write complete phrases. But this pattern, and in fact any such

⁹¹ Such as Daniel Trueman, in the Princeton music department's sophomore counterpoint course, Music 206 (personal communication).

⁹² In this context, Gjerdingen remarks (while discussing one of the prominent partimento teachers of the 18th century, Fedele Fenaroli), "Fenaroli's partimenti were praised by Giuseppe Verdi, studied by the eight-year-old Luciano Berio, and assigned by Nadia Boulanger to Walter Piston, future American author of a widely read textbook on harmony". From Robert Gjerdingen, "Monuments of Partimenti: Fedele Fenaroli", at <http://faculty-web.at.northwestern.edu/music/gjerdingen/partimenti/collections/Fenaroli/index.htm>. Accessed July 3, 2012.

pattern of descending parallel 10ths, is what Schenkerian theorists call a “linear intervallic” progression or pattern (Forte and Gilbert (1982): 83-100, Salzer and Schachter (1969): 190-199), i.e. a progression of voices moving in a certain contrapuntal configuration (in this case parallel 10ths). Furthermore, a linear intervallic progression, as is the case with contrapuntal patterns, serves to compose out a harmony – in this case tonic harmony. In other words, schemata like the Prinner are not primitives, but complex objects, which arise from composing out harmonies with voice-leading patterns such as a parallel 10ths linear intervallic progression. So, rather than taking musical schemata to be the building blocks of an idiom, one could analyze them further to reveal simpler musical building blocks, e.g. triads and the voice-leading patterns through which they are composed out, according to general principles of harmony and counterpoint. This of course leads to a generative picture of tonal structure, in which more complex structures are generated, hierarchically, from simpler ones. So, it is understandable why theorists who reject such hierarchical descriptions of musical structure, such as most Empiricists like Gjerdingen, also reject a generative description of musical structure.

Now, linear intervallic patterns, and the voice-leading principles that govern them are not specific to any idiom or style – which is why they are *general* principles of voice leading. As Allen Forte and Steven Gilbert say, “linear intervallic patterns exist over a broad spectrum of tonal music and are not restricted as to musical period, style or genre” (Forte and Gilbert (1982): 83). This means that these patterns are manifested in *idiom-specific* ways in different idioms. If we put this in the language of Principles and Parameters theory, this means that the general voice-leading principles that allow complex objects to be generated from simple harmonies, via certain voice-leading patterns like the above parallel 10ths progression, have idiom-specific parametric settings. These parametric settings can only be acquired through exposure to a given idiom, which is why a child’s innate ability to acquire a language will result in his or her becoming a native speaker of English if s/he grows up in an English-speaking environment (and thus acquires the parametric settings for English), whereas it will result in his or her becoming a native speaker of Swahili if s/he grows up in a Swahili-speaking environment (and thus acquires the parametric settings for Swahili).

In this light, a generative approach to the subject of music learning – as opposed to the Empiricist one inherent in Gjerdingen’s schematic approach – could argue that students exposed to schematic structures like the partimenti Gjerdingen discusses are not *learning* the tonal idiom from them, but are just *acquiring* the parameters of Western common-practice tonality, which helps them fix idiom-specific parameters on the general principles of music grammar they *innately* know – thus giving them native fluency in this idiom. This suggests that exposure to certain musical stimuli, such as those found in partimenti, is necessary for acquiring musical competence – but not sufficient, since this is not how children learn music. (That is, sufficiency comes from having an innate ability to acquire some musical idiom.) Of course this whole argument is speculative (albeit backed up by a rich body of Minimalist reasoning) – but it does challenge the pedagogical implications and assertions of partimenti theorists, and of schema theory more generally.

Another point that can be made to challenge the pedagogical paradigm in cognitive music theory, arises from the above idea that schemata like the Prinner are complex objects – i.e. they are not primitives in a grammatical description of tonal structure. In particular, they are not epistemological primitives; i.e. they are not what a student’s ability to acquire musical competence depends on initially – that role is performed by the child’s innate knowledge of the general principles of harmony and counterpoint, and *ex hypothesi* some notion of musical Merge.

This means that the relevance of schemata for a cognitive theory of musical structure can be rejected for the same reasons that the *Ursatz* can be rejected as a primitive in an axiomatic system of musical structure.⁹³ That is, any kind of schematic structure should be thought of as the *result* of a generative process, rather than the basis for musical learning or phrase formation. Lerdahl and Jackendoff make this point too, although they couch it in terms of their preference rules for perceptual analysis, rather than in terms of grammatical principles of phrase generation:

⁹³ In fact, if you think about it, Schenker’s *Ursatz* is a schematic structure too – it is a harmonic-contrapuntal complex, an “abstracted prototype” to use Gjerdingen’s words, that can help one make “useful comparisons” between tonal structures, and knowledge of which endows music with more meaning as well. Schenker definitely believed all of this; it is just that the *Ursatz* operates at a level of abstraction far beyond that of a partimento, or any of Gjerdingen’s other schemata.

“We propose ... that archetypal patterns emerge as a consequence of the preference rules for the four components of the musical grammar. A passage in which the preference rules maximally reinforce each other, within each component and across components, will be heard as “archetypal”. As more conflict appears among preference rules, the passage deviates more from archetypal form and is heard as musically more complex.” (Lerdahl and Jackendoff (1983): 288)

In other words, it is really the principles of grammar that give rise to certain *usable* (i.e. interpretable) structures, which then can be used in many ways, in different contexts. And *this* is what makes these structures statistically frequent in a corpus, or of value in pedagogy – when the same patterns can be used again and again in various contexts, it makes the creative process ordered and simple, but without sacrificing variety because of the manifold ways in which the pattern can be used.

Despite the above points, it is clear that the pedagogical paradigm in cognitive music theory has inspired a greater concern for stylistic detail in a theory of musical structure, which should be commended. But for our present purposes, it is worth noting again that this paradigm stands in clear opposition to the generative one proposed by this dissertation. The anti-Schenkerian aspect of this is evident even in Leonard Meyer’s earlier writings (e.g. Meyer (1956): 52-54, Meyer (1976): 753), and is probably associated most with Eugene Narmour’s critique of Schenkerian theory (in Narmour (1977, 1983-84) – although see Allan Keiler’s review of Narmour’s critique, in Keiler (1978a)). More recently, Justin London has argued that the study of musical structure, or the comparison of music with language should focus on “schemas, not syntax” (in London (2012)).

Most remarkably though, this anti-Schenkerian position has had an exact parallel in the anti-generative approach to language. Part of this approach has even focused on schematic explanations of linguistic structure (e.g. Rumelhart et al. (1986)), but this paradigm has seen its most famous manifestation in the rise of “connectionist” theory, which is a field in the cognitive sciences that models human behavior, including language learning, with computer models called “neural networks”. Connectionist networks attempt to demonstrate that human abilities like language are learned from overt, statistically-rich, environmental stimuli, rather than being acquired on the basis of innate knowledge. Connectionism is therefore explicitly Empiricist, and in the case of language scholarship, aimed primarily

at generative linguistics (e.g. see Rumelhart and McClelland (1986), Elman (1993), Elman et al. (1996), Plunkett (1995), Clark and Eyraud (2007)). So, this acts as clear evidence for Identity Thesis A – not only do the Empiricist paradigms in music and language show striking similarities (through their mutual emphasis on schematic, overtly-learned knowledge), they are both largely focused on providing an alternative explanation for musical/linguistic structure than the one provided by Schenkerian/Chomskyan theory. (Robert Gjerdingen has even worked with connectionist neural networks to model *musical* structure (in Gjerdingen (1989, 1990)), which shows an even greater overlap between this paradigm in both music and language scholarship.)

(4) Poetics: The last anti-generative paradigm I would like to discuss is, unlike the first three, also representative of the anti-scientific strain in music/linguistic scholarship. As we have discussed before, many music scholars see the task of a musical study as being one that situates music in its various historical, cultural, political contexts – and which therefore makes music a “poetic” object, an artform that reflects society, as opposed to an object of formal, physical or natural scientific investigation. Such an approach is inherent in language scholarship too, seen most famously in recent years in the rise of literary ‘theory’ and of postmodernism in general.

Now the label “poetic” for this form of music/linguistic scholarship is not unproblematic, since there is no inherent opposition between understanding music and language in aesthetic terms and also in scientific terms. The more Empiricist paradigm in cognitive music theory we just explored clearly illustrates this, with its joint interest in scientific models of music that *also* countenance the cultural aspects of music. (This just goes to show that the four Ps of anti-generative scholarship that I have been discussing in this section are not watertight categories, since the projects of many scholars have multiple descriptions – even though they seem to be united in their opposition to Schenkerian or Chomskyan ideas. Examples are Eugene Narmour’s work in perception, which also falls broadly in the pedagogical tradition of Leonard Meyer, and Richard Cohn’s mathematical work in music theory, which has also invoked ideas

from psychoanalytic theory, in a manner that has been more typical of certain postmodernist thinkers such as Jacques Lacan.)

However, I am using the term “poetic” here to refer more to the branches of musical and linguistic scholarship that actually *reject* a more scientific approach to issues of music structure and function. So, the core of the poetic tradition in music/linguistic scholarship is made up of scholars who not only reject Schenkerian or Chomskyan theory, but who are also often opposed to the various mathematical or psychological approaches to music or language inherent in projects such as GTTM or certain kinds of mathematical music theory. More generally, this paradigm represents thinkers, speaking specifically of music now, who are affiliated more with historical or ethnomusicological institutions in the academy, or with music-theoretic institutions that would like to forge greater associations with the humanities rather than the sciences.⁹⁴

Now, within our current musilinguistic context I think it is clear, once again, that the poetic approach also manifests itself in strikingly similar ways in both music and language scholarship, and also in direct opposition to the generative paradigm in both disciplines. So, this is the final bit of evidence I will present in this chapter for the identity of musical and linguistic theory, and therefore of music and language. But this discussion will have to be brief, given the vast body of scholarship that takes an essentially poetic approach to music and language, and the uncountable multiplicity of perspectives within this tradition – even to attempt a constructive critique of this approach vis-à-vis generative accounts of music and language requires a much more detailed investigation of this issue, which would take us far beyond our current Minimalist concerns.

In language scholarship, the anti-scientific core of the poetic approach is most visibly present in post-structuralist attacks on structural linguistics, particularly by thinkers influenced by philosophers in

⁹⁴ I have mentioned before the fact that many music theorists interpret Schenkerian theory in poetic terms too – which I think is perfectly legitimate, hence my admission that the current, scientific, approach to Schenker might even be considered a neo-Schenkerian project instead. Moreover, poetic approaches can in general be reconciled with more scientific ones, of which I gave some examples above. So, one could conceive of a poetic interpretation approach of Schenkerian theory that has scientific elements as well. However, I think this leads to various inconsistencies and contradictions, a couple of which I have discussed in this chapter. This is why this dissertation interprets Schenkerian theory as an essentially scientific theory of musical structure.

the continental tradition. One can see traces of this in the philosophical writings of Theodor Adorno, but more so later in the full-blown postmodernism of various French intellectuals, most notably Jacques Derrida. As the philosopher Peter Dews says:

“Over the past few years an awareness has begun to develop of the thematic affinities between the work of those recent French thinkers commonly grouped together under the label of ‘post-structuralism’, and the thought of the first-generation Frankfurt School, particularly that of Adorno. Indeed, what is perhaps most surprising is that it should have taken so long for the interlocking of concerns between these two philosophical currents to be properly appreciated... In the English-speaking world, it is the relation between the characteristic procedures of deconstruction developed by Derrida and the ‘negative dialectics’ of Adorno which has attracted the most attention: a common concern with the lability and historicity of language, a repudiation of foundationalism in philosophy, an awareness of the subterranean links between the metaphysics of identity and structures of domination, and a shared, tortuous love-hate relation to Hegel, seem to mark out these two thinkers as unwitting philosophical comrades-in-arms.” (Dews (1994): 46-47)

This emphasis on the “historicity of language” is important because it is what has driven an equivalent interest in the historicity of *music* in music scholarship – and one that has reacted against the anti-historical, synchronic formalisms of more scientifically-oriented music theories. This can be seen partly in the writings of Adorno himself, who also wrote on music – and compared music and language (e.g. in Adorno (1998)) – but is more evident in the later postmodernist tradition in music scholarship. Arguably the most famous document in this tradition is Joseph Kerman’s article “How We Got into Analysis, and How to Get Out” (Kerman (1980)), which was published in the journal often considered the intellectual home of postmodernism – viz. *Critical Inquiry* – and which was one of the foundational texts in this new musicological tradition, which is often referred to, aptly, as the “New Musicology”. Despite the title, Kerman’s point in this and other writings was not to advocate an end to analyzing musical structure, but rather to include considerations of style, history, textuality and so on in one’s analysis too. In this respect, Kerman’s suggestion mimics the similar move from structural analysis to what has come to be known as the “deconstructive” analysis of a text, popularized by Jacques Derrida in *On Grammatology* (Derrida (1998)), although Kerman does not seem to have been directly influenced by Derrida’s writings in at least the above paper. However, the wider influence of postmodernist thinking in music scholarship, and of Derrida, and other French intellectuals like Michel Foucault, Jean-François Lyotard, Jean Baudrillard,

Gilles Deleuze, Félix Guattari, and the later works of Maurice Merleau-Ponty and Roland Barthes in particular, can definitely be seen in the quick emergence of the New Musicology as a popular subdiscipline within music scholarship – this paradigm also becoming the musical home to many of the new areas of study within the wider postmodernist humanities, such as feminism, sexuality, disability, and so on (e.g. see McClary (1991), LeGuin (2006), Brett (1994), Lerner and Straus (2006)).

So, what I am calling the “poetic” tradition in the contemporary humanities, seems to have strong parallels in the way it arose in both music and language scholarship. Can we see this as a joint response to generative approaches in both music and linguistic theory too? Well, I claimed earlier that this paradigm arose as a response to structuralism in the early 20th century, so in this sense it predates the rise of generative linguistics and generative music theory (via the Princeton Schenker Project) in the 1950s and 60s. Moreover, generative theory itself arose out of a critique of structuralism, as we saw in the brief review of Noam Chomsky’s evaluation of structural linguistics a few pages ago – although the ‘post-structuralist’ aspects of generative theory are radically different from how that term is more commonly understood in the humanities.

But one could argue that at least the *recent* history of humanistic scholarship in music and language is in many ways a direct response, respectively, to Schenkerian theory and Chomskyan theory, given the towering presence of these figures in their respective disciplines, and how they are both seen as the archetypal ‘formalist’ enemy on whom the anti-formalist humanities scholar must set his/her sights. This is clear at least in Joseph Kerman’s writings. In the aforementioned “How to” paper, Kerman attacks several structurally-oriented music theorists, especially those of ‘Germanic’ lineage and interests, such as Eduard Hanslick, Rudolph Réti, and to a lesser extent, Alfred Lorenz, and Arnold Schoenberg – and even Donald Francis Tovey – but he reserves the greatest censure for Schenker, devoting several pages to a critique of Schenker’s analysis of a couple of songs from Robert Schumann’s *Dichterliebe*, something he does not even begin to do with the other theorists just mentioned (Kerman (1980): 323-326). Not surprisingly, many of the rebuttals to New Musicologists like Kerman have come from Schenker-affiliated music theorists too (e.g. Agawu (1992b, 2004, 2006), Van den Toorn (1995)).

In the case of poetic language scholarship, the vast diversity of individuals and approaches here makes it difficult for any one individual to emerge as the central figure, and Derrida – to take but one example – does not seem to have referred ever by name to Chomsky, especially as an intellectual antagonist. However, the philosopher Daniel Dennett does say that:

“Philosophers of language were divided in their response to [Chomsky’s] work. Some loved it, and some hated it. Those of us who loved it were soon up to our eyebrows in transformations, trees, deep structures, and all the other arcana of a new formalism. Many of those who hated it condemned it as dreadful, philistine scientism, a clanking assault by technocratic vandals on the beautiful, unanalyzable, unformalizable subtleties of language. This hostile attitude was overpowering in the foreign language departments of most major universities. Chomsky might be a professor of linguistics at MIT, and linguistics might be categorized, there, as one of the humanities, but Chomsky’s work was science, and science was the Enemy – as every card-carrying humanist knows.” (Dennett (1995): 386)

Also, the English and comparative literature scholar Christopher Wise has said that though he admires Chomsky’s well-known left-wing political activism, his own interest in language scholarship has been quite anti-Chomskyan – which prompted him to write an entire book that problematizes Chomskyan linguistics from a deconstructive perspective (i.e. Wise (2011)). In the first few pages of this text, Wise says that Chomsky has always appeared to him:

“...an exemplary oppositional figure in Edward W. Said’s sense, one worthy of careful attention if not emulation. On the other hand, Chomsky’s scornful attitude toward major theorists such as Jacques Derrida, Michel Foucault, and others essential to the field of postcolonial studies left me somewhat bewildered. It also seemed paradoxical that, in his regular attacks upon philosophical rivals like Derrida, Foucault, and Julia Kristeva, Chomsky never included the name of Edward Said, although the latter’s views about language are far closer to those whom Chomsky reviles than to his own. One problem with criticizing Chomsky’s views, especially for those who tend to agree with his courageous analyses of U.S. foreign policy, is that one risks undermining political objectives shared with him... [But] in the U.S. setting, students and faculty who adhere to Chomsky’s linguistic theories are often indifferent, when they are not openly contemptuous of Chomsky’s political views. Hence, I began to feel that Chomsky’s colleagues in linguistics in U.S. academe were not only selectively reading him, but that there might be something inherently wrong about his orientation to the study of language.” (Wise (2011): 1-2)

So, in Wise’s perspective on Chomsky we see exactly the kind of attitude that Dennett alludes to in the previous quote, i.e. in scholars of foreign languages and literature who are more sympathetic to postmodernism, and particularly to the views of Derrida, but still see Chomsky as a central figure in language scholarship, who they must respond to (if for no other reason than the fact that they share his political convictions, which many in the postmodernist humanities do).

Finally, Chomsky has participated in some well-publicized debates with at least two eminent thinkers within the poetics tradition, viz. Michel Foucault and more recently, Slavoj Žižek. So, in sum, the parallels between the anti-generative response to generative approaches to both music and language again seem to be conspicuous in this tradition as well.

Now, as I said at the end of the last section, my goal in presenting the discussion of this section is less to criticize the various anti-generative paradigms I have just discussed, but rather to show the striking similarities between generative and anti-generative approaches in both music and language scholarship, and how central Schenkerian and Chomskyan thought is to all of this. I believe the preceding discussing makes this point quite convincingly, justifying in the process the identity of musical and linguistic theory – which is itself predicated on the identity of music and language. If this presents a convincing case for the two Identity Theses for music and language, or at least sets up the plausibility of these theses – which I will continue to justify from a more technical music-theoretic perspective in the subsequent chapters – then my job here is, at least tentatively, done.

However, I cannot but help make one last comment before we move on, in defense of the joint Minimalist Program for music and language being proposed by this dissertation. Given the foundations of this project in generative musical and linguistic theory, this project is therefore opposed to the various anti-generative proposals we have explored in this section too. This does not de-legitimize these opposing perspectives in any way – but it does create a problem for them when they make incorrect statements about music, language, or the relationship between the two. This becomes a bigger problem when these statements go on to become accepted as fact by an uncritical intellectual community, based on certain prejudices it has against generative theory.

This is particularly true of the last, poetic, paradigm in anti-generative scholarship, given its rejection of not just generative approaches to music and language, but of scientific approaches *per se*. An example can be seen in the rather common dismissal of universals in much humanistic music scholarship (e.g. see McClary (1991): 25-26). In a way this brings us full circle, since this dismissal lies behind much of the cultural relativism seen in anti-generative and anti-scientific paradigms in music scholarship too –

including within the field of ethnomusicology, as we explored and critiqued right at the outset of this chapter. But as Leonard Meyer says in this regard, “One cannot comprehend or explain the variability of human cultures unless one has some sense of the constancies involved in their shaping” (Meyer (1998): 3).

And this point is so obvious that almost all scholars who have investigated the matter seriously agree with it. So, for example, Lerdahl and Jackendoff agree with it, given their commitment to a nativist, universalist perspective on musical structure (Lerdahl and Jackendoff (1983): 278-281) – but so does the anti-generative (and anti-GTTM) music theorist Eugene Narmour. Right at the beginning of his anti-Schenkerian, implication-realization based text on melodic structure, Narmour says “the theoretical constants invoked herein are context free and thus apply to *all* [my emphasis] styles of melody” (Narmour (1990): ix). The situation applies to language too (re-affirming again the identity of music and language in this regard) – not only do Chomskyan linguists believe in (linguistic) universals, given their nativist, universalist perspective on linguistic structure, so do *anti*-Chomskyan linguists (e.g. see Scholz and Pullum (2006): 60). So, the debate is not about whether there are universals in music or language, but what kind of universals these universals are (e.g. grammatical, as generative theorists believe, versus perceptual, as some anti-generativists, like Narmour, believe). That there are universals is not even a question – it is a *fact*, and a widely accepted one at that too. In other words, an anti-universalist attitude toward music and/or language, often based on a fundamentally anti-scientific attitude towards these systems, really just reveals an ignorance of the relevant issues on the part of those who possess this mindset. In other words – to put it bluntly – it is just plain silly.

In this section, I have tried to illustrate some of obstacles that have prevented more progress on the Princeton Schenker Project, because of various intellectual trends that have developed in music scholarship in recent years. It is my hope, though, that this chapter will reveal that there is much to be gained from pursuing the challenges of this project, as it seems to be the best way of answering Bernstein’s Unanswered Question, and shedding light on the resilient connection between music and

language that has inspired so many in the course of intellectual history. As I have said earlier, the remaining chapters of this dissertation will discuss several more connections between music and language, and their respective theories, as seen from a Minimalist perspective. But ultimately this project is beyond the abilities of a single individual; so, hopefully the renewed interest in a naturalistic music theory in the academic musical community will see many more advances in this ‘neo-Bernsteinian’ (or neo-Schenkerian) project in the years to come.⁹⁵

⁹⁵ On a personal note, I must admit that, prior to my joining it as a graduate student, I was unaware of the critical role played by the Princeton music department in formulating and advancing the research questions posed in this dissertation – particularly those of the Princeton Schenker Project. So, discovering these contributions by the very institution I have been affiliated with for some time now has been a source of great pride for me. This is why I sincerely hope that this community will both maintain its commitment to theoretical musical scholarship in the future, and resist succumbing to the same trends that have hamstrung scientific approaches to musical questions in recent years, including those that reject a search for a universal human endowment for music.

Chapter 1.2

Minimalist Musical Grammar: (i) Computational System

In the previous chapter, I tried to illustrate how compelling the music/language analogy has been for a variety of thinkers, especially when both music and language are considered essential aspects of human nature. However, I also tried to show how this analogy often breaks down under further examination – unless one looks at the matter from the perspective of generative music/linguistic theory. Specifically, there still seems to be a reason to countenance an identity between music and language, but only through a Minimalist approach to musical and linguistic grammar. I also suggested in the previous chapter that the study of musical grammar from a Schenkerian perspective already adopts such a Minimalist posture, despite there being no conscious, historical collaboration between Schenkerian theorists and Minimalist linguists – implying not only the possible identity of music and language but also of their respective theories, which constitutes the two identity theses for music and language proposed in the last chapter.

However, just making suggestions about how and where one might find an identity between music and language, and their respective theories, isn't enough; such an endeavor has to be actually undertaken – and the proof for music/language identity lies in details of this endeavor. So, to this end I turn in this chapter. The tone of the last chapter was more philosophical and methodological, and also interdisciplinary, so from this chapter onwards I will get into the technical nitty-gritty that make up contemporary music and linguistic theory. This chapter will focus in particular on the specific constitution of the computational system of human language (C_{HL}), proposed by generative linguists as the basis for human linguistic competence. It will also focus on the constitution of the proposed musical analogue to C_{HL} , viz. the computational system of human music (C_{HM}). In the process, I will review some technical ideas from current Minimalist linguistics and compare them with ideas from Schenkerian music theory, primarily to demonstrate the overlap, and the possible identity, of the two systems (although, being a dissertation in music theory, the chapter will focus more on musical matters than linguistic ones).

Despite this emphasis on Schenkerian theory, the long history of Schenkerian approaches to music theory, and the controversy over whether one can relate this enterprise to the work of Noam Chomsky and his followers, requires hammering in something I already asserted in the last chapter. This is the fact that this dissertation's main purpose is to present a Minimalist approach to musical grammar, rather than an authoritative interpretation of Heinrich Schenker's ideas. If a more historically- and contextually-informed reading of Schenker's works suggests that he was indeed already proposing what I am presenting in this dissertation that would provide an interesting historical precedent for the present project. However, I am quite happy to consider this project as more of a supplement to traditional Schenkerian theory or a contribution to "neo-Schenkerian" theory – rather than a historically-accurate reconstruction of the ideas of Heinrich Schenker. The ultimate goal here is to justify the two identity theses for music and language, rather than the specific, historically-situated ideas of the master.

There is an old belief, essentially correct, that a language consists of a grammar and a dictionary. In the context of generative linguistics, "grammar" essentially means (a theory about) C_{HL} . So, to understand the grammatical component of language from a Minimalist perspective, one needs to provide a description of how language's computational system works, and also a Minimalist explanation for why it works this way. A similar consideration arises if one wants to understand musical grammar. This challenge will engage us for most of this chapter. But what about the "dictionary" component of language? A language's dictionary is really just made up of "familiar bundles of primitive features: garden-variety words" (Uriagereka (1998): 100), known more technically as a "lexicon", and it is from words/the lexicon that sentences are generated according to grammatical principles. The primitive features that are 'bundled up' to form words are things like their gender, their number (i.e. whether they are singular or plural), their tense etc. When the features of two words 'agree' in a certain way, they can be combined according to some grammatical principle to generate a sentence. For example, the sentences "cats drink milk" and "a cat drinks milk" can be generated because "cats" is plural and so agrees with the plural verb "drink",

whereas the singular “cat” agrees with singular “drinks”, which results in the generation of a different sentence.

In the above manner, the lexicon provides the input to the grammar, from which grammatical outputs (i.e. sentences) are generated. Therefore, in order to properly understand how musical and linguistic grammar work, we have to first understand what the lexicon is that they are working on, and what kind of outputs they are generating from this lexicon. So, what is this ‘dictionary’ that musical grammar operates on, if there even is such a thing, and what are the ‘sentences’ it consequently outputs? The next two sections will address these questions, which will hopefully inform, in turn, the subsequent, detailed exploration of C_{HM} and C_{HL} that will occupy us for the remainder of the chapter.

1.2.1. The Computational System for Human Music (C_{HM}) – The Input

The issue of whether music has a lexicon like language is a complicated one. For one, music certainly does not have nouns like “cat” or verbs like “drink”, so a musical lexicon will not be identical to a linguistic one. Based on this, many scholars believe that music does not have a lexicon at all; it does not have ‘parts of speech’ (e.g. Lerdahl and Jackendoff (1983): 5, Patel (2007): 263, Katz and Pesetsky (2011): 2, Bashwiner (2009)). If this were really true, it would be a problem for a grammatical theory because without a lexicon musical grammar would not have any inputs to combine, and thus generate musical ‘sentences’ from. But popular opinions about musical lexicons notwithstanding, the idea that musical sentences are generated from a commonly agreed upon set of inputs has always been implicit in music theory. In the case of Western tonal music, these inputs are generally taken to be chords. Both Schenker’s and Lerdahl and Jackendoff’s theories, in addition to many non-cognitive theories of Western music, take chords to be the building blocks for musical sentences, since they are combined according to certain relationships that exist between them to generate hierarchical musical structures – similar to what we know to be the case for sentence generation in language. Moreover, chords themselves can be ‘inflected’ by melodic pitches much in the way words are inflected by suffixes and prefixes in language. This is clearly evident in traditional, and especially Schenkerian, descriptions of chord grammar, for

example in Schenker's description of "diminutions" (Schenker (1979): 93-106). Finally, the way at least Schenkerian theory describes how chords are prolonged at surface levels of structure in musical passages is strongly connected to (Schenkerian) descriptions of musical meaning, rhythm, and formal design too, among other things. So, chord structure seems to play a role that is at least analogous to what words do in language because linguistic meaning, and to a lesser extent linguistic rhythm (i.e. prosody), are strongly related to linguistic word structure and linguistic grammar too, especially in the descriptions of these phenomena in Minimalist linguistics. (Part II of the dissertation will deal exclusively with this issue.)

In this light, the idea that music has a lexicon does not seem so absurd – as long as one can demonstrate that chords play essentially the same role in musical grammatical operations as words do in linguistic ones (despite the fact that they clearly lack some of the features that words have, such as their 'noun/verb-ness', their gender etc.). Therefore, demonstrating the lexical status of chords is an important issue in a generative grammatical theory of music. The best contemporary defense for the view that chords form a lexicon in Western tonal music seems to be Fred Lerdahl's (2001) description of chord structure, in which he argues that chords are combined into larger musical structures by virtue of their inherent "tonal pitch space" properties. That is, chords can be said to 'agree' on the basis of their relative proximity within a spatial representation of chord structure, which Lerdahl describes and explores with eloquence in his text. Importantly, Lerdahl himself does not note any explicit similarities between word agreement in language and chord agreement in his pitch space theory, and the broader theory of music that he developed with Ray Jackendoff denies, as has been discussed earlier, that music has a lexicon – or that musical grammar even parallels linguistic grammar in many important respects. Therefore, the question of whether chords really constitute the lexicon of Western tonal music remains unresolved.

There is one aspect of chord structure that has to be demonstrated if chords are to be considered lexical, which seems to have a more positive and conclusive answer though. To understand this, consider that if generative grammar is universal (as Minimalism believes for language and as generative theories like Lerdahl and Jackendoff's often claim for music too), then the lexicon of the grammar has to have universal properties too. So, even though different languages have different words, these words have

features that are universally shared. Features like tense, gender, number etc. are not present in the lexicon of only one or two languages – they characterize the words used in *all* languages. If we assume for the moment that chords constitute the lexicon for Western tonal music, then for musical grammar to be universal, all musical idioms must possess chord-like structures with similar properties too.

Now this is the kind of phenomenon whose existence many music scholars have rejected – in fact we looked at Harold Powers’ specific rejection of it in the last chapter, when he asserted that “no two natural languages of speech could differ from one another in any fashion comparable to the way in which a complex monophony like Indian classical music ... differs from a complex polyphony like that of the Javanese gamelan”. However, I would like to demonstrate in the remainder of this section that quotes like the above reveal a fundamental misunderstanding of the inputs to musical grammar, particularly when they distinguish them as being “monophonic” in one idiom and “polyphonic” in another. I would like to argue that the chords that make up the input to polyphonic idioms are actually quite monophonic in their properties, so that there exists no real difference in the inputs to the grammars of monophonic and polyphonic idioms. In other words, all idioms are essentially monophonic, and this results from their being generated from essentially monophonic inputs (which is how I will define chords in the next few pages) – implying that this feature of chords characterizes the inputs to all musical idioms. Given that this universality is a requirement for a true lexicon, as stipulated in the last paragraph, this aspect of chord structure strengthens the case, in my opinion, of chords being considered truly lexical, and thus the basis for a true musical lexicon.

To begin this discussion, let us examine a particularly famous piece of music, Beethoven’s “Kreutzer” sonata for violin and piano, on which Tolstoy based his eponymous novel. The second movement of this sonata is a set of theme and variations in F major. Let us examine the actual music of this theme and its variations. In the interests of space, I will only discuss the first phrase of the melody of the theme and its variations, shown in Example 1.2-1. Note that the violin sometimes shares the main melody with the piano, as happens, for example, in variation I. Example 1.2-2 shows us that the variations contain a

number of pitches from the theme, as shown by the circles around them. However, importantly, some of the notes of the main theme are omitted in one or more variations, such as the penultimate note G, shown in the rectangular box (which is not heard in variations 1 and 2). (Please ignore the box in variation III. I will return to that later in the section.)

The explanation for these similarities and differences is a harmonic one. By examining the chord progression shown on the bass staff of Example 1.2-3a, we see that the circled pitches in the main theme are all chord tones, and it is these pitches that are shared between the theme and its variations. For the same reason, some of the notes in the main theme, such as its penultimate G, can be omitted from the variation melodies because they are absent in the underlying chord progression.¹

Importantly, the harmonic relationship between theme and variations is an abstract one, since the notes of this chord progression are never realized in exactly the same way in Beethoven's score. If we examine the piano accompaniment for the theme and its variations, we will see that it changes each time. So, what is common to all the passages here is not an *actual* chord progression, seen in the actual texture of the piano part, but rather an *abstract* progression that can be inferred from the different melodies and their corresponding piano accompaniments.

The harmonic relationships between the theme and its variations are abstract because they concern the grammatical *function* of chords in generating musical sentences. This is not concerned with how these chords are actually realized in the *structure* of the piece, i.e. in the exact notes played by the piano. For example, the harmony in the first four measures of the piece is a dominant seventh harmony, realized as a C⁷ chord in the score. The instability of this chord, which partly owes to the tritone in it (between E and B-flat), directs the chord to resolve to the more stable, tonic, F major harmony that follows in measure 5. This yields a grammatical structure, a constituent made up of the C⁷ to F

¹ Incidentally, there are other notes in the main theme that appear in all of the variations but which are not chord tones, such as the D in the first measure of the theme. The important thing to realize here is that the D could have been omitted from a following variation, just as the penultimate G was, without really rupturing its association with the main theme, at least not as much as an omission of the first E would have done. This is why only chord tones need to be retained in the variations of a theme for their connection to the theme to be understood.

Example 1.2-1. Beethoven, Violin Sonata #9 “Kreutzer”, Op. 47/ii: Melody of the theme and its variations (first eight bars each)

Theme

Variation I

Variation II

Variation III

Variation IV

Example 1.2-2. Beethoven, Violin Sonata #9 “Kreutzer”, Op. 47/ii: Common tones in the theme and its variations

Theme

Variation I

Variation II

Variation III

Variation IV

Example 1.2-3. Beethoven, Violin Sonata #9 “Kreutzer”, Op. 47/ii: Main theme, mm. 1-8.
(a) Chord tones (b) Structural motives

(a) Chord tones: The bass line shows a sequence of chords: V⁷ (F major), I (F major), IV (D minor), V⁷ (C major), and I (F major). The treble line shows a melodic line with various rhythmic patterns and accidentals.

(b) Structural motives: The treble line shows structural motives indicated by horizontal lines above the notes. The bass line shows the same chord sequence as in (a).

progression. So, the harmonic phenomenon we are seeing is really a grammatical one. But it is one that arises out of the abstract grammatical function of the chords in these passages, not from the actual structure of the notes that make up the passages. If this were not the case, the different structures of the theme and its variations would not reveal a common harmonic progression present in all of them.

For the above exploration of the abstract nature of harmony, Heinrich Schenker’s discussion of harmony, beginning with his *Harmonielehre* treatise (Schenker (1973)) and leading up to his monumental theory of musical grammar in *Der freie Satz* (Schenker (1979)), is especially relevant for two reasons. First, Schenker argued that the abstract grammatical functions we have been discussing only apply to specific sonorities in the surface of a musical piece – not to every vertical chord in the surface. We can understand this idea by considering Example 1.2-3 again. We know from our previous discussion that the grammatical function of each chord in the surface of the theme and its variations is given by the abstract harmonic progression in the example that, in turn, is inferred from the different melodies and their

corresponding piano accompaniments in the theme and variations. We can therefore label every chord in this abstract progression on the basis of its grammatical function in the abstract progression, e.g. through Roman numeral analysis. In this manner, we could label all the chords in the first four measures of Example 1.2-3 as V^7 chords, since the chord in each of those measures realizes a dominant seventh C^7 harmony.

Instead of labeling each chord according to its grammatical function, Schenker argued that the grammatical function of all of these measures can be ascribed to a single, even more abstract, entity called a *Stufe* (i.e. scale “step”, “level” or “degree”) (Schenker (1973): 133-153). A *Stufe* “is a triad that serves in the harmonic foundation of a passage or composition” (Cadwallader and Gagne (1988): 65). So, the first four measures of Example 1.2-3 can be thought to be founded on the C-major *Stufe*, since these measures all realize a (dominant seventh) version of C-major harmony. The reason why this harmony is a scale-step is because of the important connection between the root of this abstract entity (i.e. the pitch class C) and the tonic F of this entire passage in the F-major scale – the connection being that of scale-degree 1 (the tonic F) and scale-degree 5 (the dominant C), one of the most important grammatical relationships between two pitch structures in tonal music. The *Stufe* of a musical passage can be embodied within that passage as a specific sonority or in a string of sonorities. (So, all of the sonorities in the four measures we have been considering embody the C-major *Stufe*, which is why only one Roman numeral (the V^7) followed by a horizontal line is sufficient to represent it in Example 1.2-3a.) When a string of sonorities embody a single *Stufe*, that entire passage can be called a “*Stufengang*”, i.e. a scale-step area or region.

Since *Stufen* are abstract entities they have to be realized in actual music in a certain way. Schenker argued that they are realized melodically. To understand this idea, consider Example 1.2-3b. We see here that the chord tones of a *Stufe* serve to arpeggiate that *Stufe* by appearing successively in a melody – in the way the first four melodic tones of the example arpeggiate a C-major *Stufe*. Since these arpeggiations form a *Stufengang* with a certain temporal duration in the actual composition, Schenker argued that they expand the *Stufe* by melodic means in the actual fabric of a piece:

“But in all cases we do not need three voices to produce these consonant intervals [i.e. the intervals between the chord tones of a *Stufe*]; i.e., the concept of the triad is not tied, as one might think, to the concept of real three-phony. Rather, it may be fulfilled by two voices, even by a single one. In the latter alternative, Nature as well as art is satisfied if the course of a melody offers to our ear the possibility of connecting with a certain tone its fifth and third, which may make their appearance in the melody by and by.” (Schenker (1973): 133)²

Since these arpeggiations form motives within the melody (illustrated by the horizontal brackets in Example 1.2-3b), we can redescribe these abstract chord progressions as a sequence of actual melodic motives in the piece. (Note that these motives can be made up of single notes too, as the final F5 motive is. In such cases, the underlying chord has not been arpeggiated but is merely represented by one of its constituent pitches – though it does have the potential to be arpeggiated if the composer chose to do so.)

It is important to note that such motives are *structural* melodic motives, since they embody the underlying harmonic structure of the piece. For this reason, they should be distinguished from the garden-variety motives that are often discussed in music theory, which could be any group of melodic pitches, not merely those that arpeggiate *Stufen*. (See Burkhart (1978) for the classic discussion of such *non-structural* motives from a Schenkerian perspective.) Before Schenker developed his harmonic theory, he was deeply interested in such motives as an organizing force in tonal music, and persevered with this interest later in his life too. However, the grammatical role played by non-structural motives is *not* the same as those played by structural ones (Keiler (1989): 278-292), and often raises problems for those who have sought to incorporate them into the more harmonically based grammar being considered here (the classic description of these problems can be found in Cohn (1992b)). So, for our present purposes we will only consider motives that arpeggiate *Stufen*, and thus have a structural, harmonic basis. Note that Schenker did not use the term “structural motive” himself; but the unique wedding of harmony and

² In fact, Schenker had a more abstract, ‘metaphysical’ conception of melody than many of his peers, in which melody is in many ways the locus of human musical creativity – i.e. the locus of a *natural* human predisposition for music. Since such a predisposition is part and parcel of the concept of generative grammar too, Schenker’s conception of melody just reinforces the connection between melody and grammatical structure in music. As Allan Keiler says, “One original feature of Schenker’s discussion [about melody, in the “Der Geist” essay] is that he emphasizes melody not as a musical parameter so much as the creative melodic impulse. Indeed, the emphasis is not unlike that given to the properties of universal musical competence that forms the subject matter of the first half of the essay. It is not only that Schenker characteristically describes melody as an inherent property of the musical instinct toward creativity; the primeval character of melodic creation is described in such a way as to give the feeling of great antiquity.” (Keiler (1996): 176-177)

melody in Schenkerian theory demonstrates how melodies have a deeper harmonic structure based on the *Stufen* they expand.³

Schenker's second relevant contribution to our present discussion on the inputs to musical grammar was his theory of prolongation. As we can see from Example 1.2-3a, the chord tones that make up the structural motives in a piece are often non-adjacent in the actual music of a piece, and are interspersed with non-chord tones. For one, this makes their motivic relation a concealed one. Further, it allows these motives to be inflected by non-chord tones in various ways – in fact, this is exactly how the different variations of the Kreutzer theme arise. Schenker called this process prolongation,⁴ and was thus able to show how varied melodic surfaces arise by elaborating a scale-step in different ways.

But Schenker's prolongational theory was not merely concerned with how *Stufen* are elaborated locally. He showed how complex phrases can be generated from the larger, harmonic relationships *between Stufen* too. We saw earlier that chords have hierarchical relationships based on their relative stability or instability, as a result of which *Stufen* can be prolonged by other *Stufen* as well. Since Schenkerian theory describes how *Stufen* are realized by melodic means in actual music, these hierarchical, prolongational relationships apply to entire melodic spans (what I am calling structural motives here). This leads to complex motivic hierarchies in music, which can thus generate larger hierarchical structures in music. The tree diagram shown in Example 1.2-4 illustrates exactly such a hierarchical structure for our Kreutzer theme. In this diagram, the taller branches represent hierarchically-superior structures, the shorter branches hierarchically-inferior ones (Lerdahl and Jackendoff (1983): 112-117).

³ Given the problems inherent in the term “motive” in Schenkerian theory, I could have just used the term “arpeggiation” instead of “structural motive” here. But that term really makes sense only within a triadic context such as Western tonal music, and cannot be easily applied to other, non-triadic, musical idioms – hence my preference for “structural motive”.

⁴ This is the more common way of interpreting “prolongation”, but is actually closer to Schenker's notion of *Auskomponierung* (i.e. composing-out). For another way of interpreting this term, see the discussion in the last chapter.

Example 1.2-4. Beethoven, Violin Sonata #9 “Kreutzer”, Op. 47/ii: Tree diagram of theme, mm. 1-8

The image shows a musical score for the first eight measures of the 'Kreutzer' theme. The score is written in G major (one sharp) and 2/4 time. The melody is in the treble clef, and the bass line is in the bass clef. A tree diagram is drawn above the staff, showing the hierarchical structure of the melody. The root node is at the top, and branches lead to individual notes. The chord symbols below the staff are V7, I, IV, V7, and I, indicating the harmonic progression. The notes in the melody are G4, A4, B4, C5, B4, A4, G4, F4. The notes in the bass line are G2, B2, D3, E3, G3, B2, D3, E3.

Finally, not only are *Stufen* realized melodically in actual music in Schenkerian theory, they are joined to other *Stufen* by melodic means as well. If a *Stufe* is realized as a string of pitches in a passage, one pitch is taken to be the primary one because it is this pitch that maps on to a pitch in an adjacent *Stufe*, thus joining these *Stufen* into a larger structure. But in the process, a *melodic* line of primary pitches is realized in the fabric of the passage too – so, musical ‘sentences’ are formed from scale-step progressions by melodic means, i.e. by means of a melodic line that joins the chords in the progression.

Example 1.2-5 illustrates this phenomenon in our Kreutzer passage. As the example shows, each bracketed structural motive has one primary pitch, notated with a white note.⁵ These pitches are primary because they help join the different scale steps of the passage together, to yield the theme, while also forming the melodic line F5-G5-A5-B-flat5-G5-F5. The melodic line also results from its constituent pitches being connected through *stepwise* melodic motion, an exception being the leap down from B-flat5 to G5, which is explained by the B-flat5 being an incomplete upper neighbor to the preceding A5 – hence its peculiar eighth-note notation in the example (a standard Schenkerian convention for neighbor notes).

⁵ This is consistent with the earlier observation that structural motives can often be made up of just one pitch – in which case this pitch is usually the primary pitch of that motive.

The G5 is not actually there in the melody, but is implied by the underlying chord progression, hence the parentheses around it. (The fact that pitches are often implied in Western tonal music is a characteristic that I will return to in section 1.2.4.)

Example 1.2-5. Beethoven, Violin Sonata #9 “Kreutzer”, Op. 47/ii: Voice leading structure, mm. 1-8

The overall contour of this melody is that of an initial ascent from F5 to A5 (that Schenker called an *Anstieg*), followed by a descent back down to F5. Schenker called this entire melodic line the *Umlinie* (i.e. “Fundamental Line”) since the line reveals the underlying grammatical structure of the passage. This particular *Umlinie* is a “scale-degree 3 line” because it is characterized by a descent from scale degree 3, A5, also known as the *Kopftone* (i.e. “head tone”) of the *Umlinie*. The stepwise motion of the *Umlinie* also shows how the chords of the passage are connected in a melodically economical way – which reflects the economical nature of grammar in general, a characteristic that I shall return to later in the chapter.

By extending his prolongational theory to abstract harmonic relationships, and thus illustrating how musical sentences are generated from harmonic structure, Schenker described a prolongational,

generative grammar for Western tonal music. So, in light of the preceding discussion, we could say that the musical structures of Western Classical tonality are just strings of structural motives in hierarchical, grammatical relationships with each other.

As the preceding discussion illustrates, Schenkerian grammatical theory can be construed as a primarily melodic theory – a theory of structural melodic motives – that only get their grammatical *function* from an abstract, harmonic foundation. This harmonic foundation could be concretized within a musical texture by means of a bass line, which then would give Western music its ‘vertical’ nature; but it does not have to be, or else monophonic textures in the Western common practice (such as the solo violin and cello pieces by Bach) would not have any harmonic structure. Bass lines or vertical, chordal sonorities in Western music really serve to disambiguate the harmonic structure of a passage for a *listener* (e.g. is a passage in C major or A minor?) – a composer undoubtedly knows what the harmony of a passage is in his/her mind *prior* to confirming it with an explicit bass line written into the score, so harmonic structure should not be construed as being dependent on the existence of a bass line in the score. So, what we are dealing with is what composers know when they write music, not with how listeners hear music – I am not focusing on the *perceptual* aspects of harmony, in which bass lines and vertical sonorities must be considered, as cues to help us parse a musical surface, and help disambiguate it for the listener. It is for this reason that I am adopting a Schenkerian approach to musical grammar, rather than that of Lerdahl and Jackendoff, for whom the grammatical project was one of describing how listeners *perceive* music (Lerdahl and Jackendoff (1983): 6).

This point is illustrated further in Example 1.2-6. The top of the example displays the first four measures of the first movement main theme of Beethoven’s A major cello sonata, which is performed by the cello without accompaniment at the very beginning of the piece. Throughout the rest of the exposition, these measures are either repeated by the solo cello, or played by the right hand of the piano doubled in octaves by the left hand. Therefore, the melody is sounded throughout the first part of this sonata

Example 1.2-6. Beethoven, Cello Sonata #3, Op. 69/i: Harmonic structure of the main theme

Allegro, ma non tanto

Cello

p dolce

m. 152

dolce

fp

Piano

fp

Harmonic Reduction A

(cello)

I V vi (iii) ii⁶ vii^{o4}/V V⁶

Harmonic Reduction B

10 - 5 - 10 - 5 - 10

Tonic Tonic-P Dom-P Dominant

movement without an accompanying bass line – in fact, without any sort of vertical, harmonic information at all. Does this mean that this melody has no harmonic structure, and therefore no grammatical organization? Not at all – for anyone well versed in Western tonal music, the harmonic structure of this passage is clear even in the absence of a bass line in the score. We can even describe this harmonic structure, although, without an explicit bass line, this description will be abstract, probably even more so than a scale-step description of the passage’s structure.

So, we can say that the first measure clearly realizes tonic harmony, which is reinforced by the A3-E3 structural motive here, which arpeggiates the tonic triad. This might not be evident to the listener when s/he first hears this measure. So, the harmonic function of this measure might only become evident to her/him *retrospectively*, after hearing the rest of the passage. However, there can be no doubt that Beethoven himself knew what the harmonic function of this measure is, or else he could not have composed the passage in the first place. Along similar lines, the second measure of the passage seems to realize a tonic-prolongational function; therefore the first two measures form an opening tonic-prolongational span in the main theme. This is complemented by the dominant-prolongational span of the next two measures, with the third measure prolonging dominant harmony by means of some other triad, the dominant harmony itself arriving in measure 4.⁶

As I just said, this description of the passage’s harmonic structure is abstract, since it involves the abstract notions of “tonic” and “dominant” – *functional harmonic* notions that are more abstract than even abstract scale-steps, since multiple scale-steps can realize a single harmonic function. (For e.g. scale-steps V and VII can both have dominant function, and II and IV can both have predominant function.) So, even if my description of the grammatical functions of the first four measures of the Beethoven cello sonata are

⁶ Note that one could obviously debate the specific harmonic functions I am attributing to these measures, but that would be a conflict between two grammatical theories, not a conflict in Beethoven’s mind about the harmonic structure of this passage. The composer obviously knows what specific function each measure has implicitly, which is different from the (explicit) knowledge the music theorist has to reveal about the structure of the passage. The theorist has to reveal this knowledge by getting into the composer’s mind, so to speak. In this respect, the job of the music theorist is not different from that of the linguist who tries to get into a 2-year old child’s mind to understand the knowledge of his or her mother tongue that the child already has (which allows young children to speak their native languages fluently at a young age) – even though different linguists can debate what this knowledge is within a theory of linguistic grammar.

correct, this will not determine which specific triads would appear in the accompaniment of the cello melody, if Beethoven were to write one. Indeed, one of the most famous functional harmonic theories in Western music, that proposed by Hugo Riemann, ascribes one of essentially only three functions to all chord structures in a musical passage. According to this view, the dominant-prolongational function of measure 3, for example, could be realized by any predominant harmony, though voice leading and other considerations will preclude the use of some chords in the passage. At this level of abstraction, ascribing a function to a sonority in a musical passage becomes an interpretive task too, and is not solely dictated by chord structure. This is actually an important point, since interpretation plays an important role in Minimalism. In fact, in later sections I will argue that (Riemannian) harmonic functions have more to do with the semantic, interpretive side of musical grammar, unlike Schenkerian scale-steps, which are more purely syntactic in the role they play in musical structure.

Beethoven finally reveals how he himself conceived of this passage (confirming my functional description of the passage in the process) only at the onset of the recapitulation in measure 152 of the movement, where he provides a contrapuntal accompaniment to the cello melody for the first time in the piece. This is shown in the second stave system of Example 1.2-6. The third stave system in the example (Harmonic Reduction A) reduces the piano part of mm. 152-155 to show the chord succession that forms the accompaniment to the cello melody in these measures. The cello melody is shown in the bass stave here, since it itself forms the bass line to this passage now.⁷ Of note here is the ii^6 harmony in measure 3, which turns out to be Beethoven's predominant harmony of choice for prolonging the following dominant harmony of measure 4. The final stave system in the example (Harmonic Reduction B) revoices the chords of the previous reduction to reveal the voice leading and the prolongational, grammatical structure of the passage. Here we see that the tree diagram under the stave system verifies my previous description of the grammatical function of each measure. The tonic-prolongational role of measure 2 is verified by

⁷ Incidentally, the bass register of the cello often allows cello melodies to have a clearer harmonic structure. For example, see measure 51 of the first movement of Beethoven's first Razumovsky string quartet, Op. 59 No. 1. There the solo cello plays a measure's worth of notes that could easily be heard as the bass line in a IV (or II) – V – I cadential harmonic progression. Maybe tonal composers tended to think of cello melodies in bass line terms because that is how they are used to writing for the instrument in polyphonic music.

the way the tonic sonority on the downbeat of measure 1 is prolonged by a classic 10 – 5 voice leading sequence between the outer voices. This sequence is halted in measure 3, where a voice exchange (of the pitches B and D) between the bass and alto voices helps prolong the ii harmony of this measure. This verifies the function of this harmony as a left-branching dominant prolongation of the V in the following measure, which is reinforced by the applied diminished seventh chord on the downbeat of measure 4.

Even though we have to wait until measure 152 for Beethoven to provide an accompaniment to the main theme, and thus clarify the grammatical structure of the passage, hints of its structure are already given to us earlier in the movement. For example, consider the transition from the main theme to the secondary theme of the movement that commences in measure 23. The passage is shown in Example 1.2-7. This transition begins with a 2-bar phrase that states a varied version of the main theme in the parallel minor with an accompaniment that clarifies the harmonic structure of the phrase. This version of the theme is played by the right hand of the piano. The first four notes in the treble stave (A4-E5-F5-E5) clearly mirror the first four notes of the main theme, even though the pitches have been altered to suit the parallel minor key prevailing in these measures. Following the implied harmonization of these pitches in Example 1.2-6, we would expect these four pitches to be part of a right-branching prolongation of the initial tonic – and this is exactly what the tree structure in Example 1.2-7 suggests. Similarly, the three notes D5-C5-B4 starting on the second beat of measure 24 should be part of a left-branching prolongation of dominant harmony, just as they were in measure 3 of the main theme, and this is what the branches of the tree in Example 1.2-7 suggest too.

Clearer hints about the harmonic structure of the main theme appear in the development section. Let us see how this happens for the two pitches in measure 4 of the theme, the pitches A2 and G#2. Even though this movement is in A major, the A2 is not a structurally important pitch here, but rather an appoggiatura to the following G#2. This is evident from the way the A is harmonized by a diminished 7th harmony in mm. 155 that clearly prolongs the following E major dominant harmony (of which G# is a chord tone), as shown in the harmonic reductions of Example 1.2-6.

Example 1.2-7. Beethoven, Cello Sonata #3, Op. 69/i: Structure of transition in mm. 23-24

m. 23

Cello

Piano

f *sf* *sf* *sf*

f *sf*

$i \frac{5}{3} \quad - \quad \frac{6}{4} \quad - \quad i \frac{5}{3} \quad ii^6 \quad V \frac{6}{4} \quad - \quad \frac{5}{3}$

The four-bar phrase starting at measure 107 in the development section, given in the first stave system of Example 1.2-8, is analogous to the four-bar extract from the main theme we have been considering. This is clear from the way the first two measures of the phrase prolong the local tonic of F# minor and the next two the local dominant of C# major. Mm. 107-109 of the piano's right hand part are clearly analogous to mm. 2-4 of the cello's main theme, as seen in both the identical rhythmic structure of both passages and the similar descending melodic contour seen in the latter part of both passages (i.e. E3 to G#2 in mm. 3-4 and B5 to E#5 in mm. 108-109). So, these measures constitute the development of the main theme in F# minor. Now, consider the two pitches in measure 109 with the box around them, viz. the F#5 and E#5. These pitches are analogous to the A2 and G#2 discussed in the previous paragraph for the reasons just stated. The A2 and G#2 are not accompanied by a bass line, but the fact that the analogous F#5 and E#5 in measure 109 are, and that this proves that the F#5 is an appoggiatura to the E#5 (given the C# major harmony of the measure) demonstrates the appoggiatura status of A2 in measure 4 as well. A similar event happens two bars later in measure 111 of the cello part (shown by another rectangle in Example

1.2-8), where the first note B3 is an appoggiatura to the following A3 given the prevailing F# minor harmony in the measure.

Example 1.2-8. Beethoven, Cello Sonata #3, Op. 69/i: Development of the main theme

The image displays two systems of musical notation for Beethoven's Cello Sonata #3, Op. 69/i. The first system, labeled 'm. 107', shows the Cello and Piano parts. The Cello part begins with a half note B3, which is an appoggiatura to the following A3. The Piano part features a complex triplet pattern. Below the notation, harmonic analysis is provided: '(F# minor) Tonic' for the first two bars, 'Dominant' for the next two bars, and 'Tonic' for the final two bars. The second system, labeled 'm. 127', shows the Cello and Piano parts in a different key. The Cello part begins with a half note B4, which is an appoggiatura to the following A4. The Piano part features a complex triplet pattern. Below the notation, harmonic analysis is provided: '(C# minor) Tonic' for the first two bars, 'Dominant' for the next two bars, and 'Tonic' for the final two bars.

The second system of Example 1.2-8 shows the phenomenon discussed in the previous paragraph being repeated in mm. 127-131, this time in the local key of C# minor.⁸ And there are more examples of this. For instance in mm. 113 the E5 in the piano's right hand is an appoggiatura to the following D#5 in the

⁸ Note that in all of the passages being discussed here, there is one important difference between the first four bars of the main theme and its later development in mm. 107-110 and 127-130. This difference lies in the placement of the cello melody (or its restatement by the piano's right hand) within the first four bars of the main theme and its later appearance in the analogous parts of mm. 107-110 and 127-130.

dominant seventh (B⁷) harmony of the measure, something that is repeated by the piano's left hand in mm. 117. Subsequent to this, Beethoven dispenses with the appoggiatura altogether, choosing instead to go directly to the pitch that resolves the appoggiatura on the downbeat of the measure where the appoggiatura used to appear previously. So, in measure 119, where the main theme is being developed in E minor by the piano's right hand (slightly obscured by the rapid sixteenth-note octave work), the downbeat of the measure sounds the chord tone G⁵ directly without sounding the expected appoggiatura A⁵ first. The same thing happens in the next four measures while the main theme is being developed during a modulation from E minor to B minor, and is played alternatively by the piano's left and right hands. Example 1.2-9 shows the voice leading sequence that realizes this in mm. 119-123. The circled notes in the example show the chord tones that would normally resolve the appoggiatura in the main theme, but now appear directly on the downbeat of each successive measure sans appoggiatura.

Example 1.2-9. Beethoven, Cello Sonata #3, Op. 69/i: Voice leading sequence in mm. 119-123

m. 119

6	-	6	-	6	-	6	-	6
8	-	10	-	8	-	10	-	8
5	-	7	-	5	-	7	-	5

Em V7 of G G V7 of Bm Bm

Since all of these passages are developments of the main theme, and they all share similar structural features, the grammatical structure of the unaccompanied main theme in mm. 1-4 is now beyond doubt – even though it has no vertical accompaniment to clarify its harmonic structure. The conclusion being that a melody does not need a vertical accompaniment or a bass line to have a harmonic, grammatical structure. This is true even if the latter can help clarify the harmonic structure of a passage to a listener, as they certainly do in the case of the main theme of the cello sonata when it is developed and recapitulated

later in the movement. In other words, melodies have a harmonic structure implicit in them, in the form of the structural melodic motives from which the melody is generated.

As a result of this, pitches can be implied in melodic lines in Western tonal music too, as shown by the parentheses in Example 1.2-5. This explains a curious feature of Example 1.2-2. In that example, the minor mode variation (variation III) of the main Kreutzer theme omits several pitches from the main theme in mm. 4-5 (shown by the rectangle around them). The F#5 and G#5 in measure 4 of the theme are non-structural, so their omission in variation III is not unusual, and can be explained along the same lines used to explain the omission of the G5 in the last measure of the theme in several of the subsequent variations. However, the other pitches in those measures are structural chord tones, so the explanation for their omission in variation III is different. The two pitches in measure 5 of the theme, A5 and F5, are actually there in variation III but are played by the piano part, and the A5 is changed to an A-flat5 to accommodate the change in mode. The G5 is not played in variation III at all, but is replaced by a B-flat5 instead (although it could be argued that G is sounded by the piano's left hand on the last beat of measure 4). All of these changes and substitutions are explained by the fact that the chord tones in variation III belong to the same *triad* as those in the theme, which makes their substitution by other tones from the same triad possible.

So far I have been discussing how chords can be considered essentially monophonic entities, and thus the 'lexical' basis not only for polyphonic musical idioms (as they always have been), but even for monophonic idioms, based on their essentially monophonic qualities – thus endowing them with the universal status that a true lexicon should have. If this is indeed true, all musical idioms should have *Urlinie*-like structures in them, as Schenker proposed for the grammatical structure of Western tonality, although in other idioms these *Urlinie*-like structures need not be identical to Schenkerian *Urlinien* – that is, they need not be descending, that too from only the three scale-degrees that can serve as the *Kopftön* in the Schenkerian *Urlinie*, and they need not be stepwise either as the Schenkerian *Urlinie* is. All they need to be are the hierarchically-organized, fundamental melodic structures that lie at the foundation of

grammatical generation in an idiom, which are made up of harmonically-derived structural motives. We could refer to these structures as idiom-specific examples of a universal, ‘generalized’ *Urlinie*, for which the idiomatic exemplar in Western Classical tonal music would be the Schenkerian *Urlinie*. In this context, Arnold Schoenberg made the interesting observation that:

“In homophonic-harmonic music, the essential content is concentrated in one voice, the principal voice, which implies an inherent harmony.” (Schoenberg (1967): 3)

Schoenberg called this voice the *Hauptstimme*, which literally means “head voice” (although this should not be confused with the better known singer’s term). Admittedly, Schoenberg’s conception of the *Hauptstimme* is quite different from Schenker’s *Urlinie*, mainly in that it was not conceived as a hierarchically-organized grammatical structure. However, the idea that the main harmonic (and therefore grammatical) content of a musical passage is connected to one fundamental melodic line in it is shared by both approaches.

So, combining both Schoenberg’s *Hauptstimme* and Schenker’s *Urlinie*, I propose the more neutral term *Headline* to refer to the idea of an idiom-neutral, generalized *Urlinie* described above.⁹ In section 1.3.2 of the next chapter I will illustrate just such a Headline, from which the fundamental melodic structures of both Western Classical tonality and North Indian music might be derived. This Headline hypothesis of course depends on the larger argument, as we have seen, that chords form a ‘lexicon’ for musical grammar in both monophonic and polyphonic musical idioms. (After all, it is from chordal building blocks that a Headline is derived, albeit monophonically, given the Schenkerian marriage between chordal harmony and melodic structure.) But this argument has teeth only if the flip side of the equation is also true. That is, chords can be considered as a universal musical lexicon not only if they act as inputs to monophonic idioms, but also if the building blocks of these idioms display chord-like qualities. And this is not necessarily true. So, just because a polyphonic idiom like Western tonality is

⁹ The term “Headline” also invokes the notion of a news headline, i.e. something that catches our attention. This association is relevant to my conception of a musical Headline, since the Headline is what catches our minds’ attentive gaze, as the locus of grammatical content in a musical passage – resulting in its also being what the mind processes when comprehending the grammatical structure of the passage.

made of up triadic chords that are realized monophonically (and might thus yield a Headline, specifically the Schenkerian *Urfinie*), that does not make it identical to a monophonic idiom like North Indian music, and it does not imply that North Indian music is made up of triadic chords, albeit realized monophonically. In the end, however, I believe the lexical status of chords still obtains because a monophonic idiom like North Indian music *can* be described in chordal terms, and which, in turn, also yields a Headline for North Indian music. The next chapter is devoted to tackling this issue exclusively, and as mentioned above, will go on to illustrate an actual Headline for both North Indian and Western Classical tonal music.

So I now turn to the other end of grammatical generation in music, viz. what the outputs of musical grammar are, once its lexical inputs have been operated upon by the computational system.

1.2.2. The Computational System for Human Music (C_{HM}) – The Output

My discussion of what the outputs of musical grammar are will be relatively brief and inconclusive compared to the preceding discussion about the inputs to musical grammar. This is because among the many topics in musical grammar this is the one, in my opinion, with the least satisfactory conclusion. As stated before, the outputs of musical grammar would be analogous to the sentences of language, and defining what a musical ‘sentence’ is continues to be a frustratingly difficult problem, even within a Schenkerian approach to musical grammar.

If musical sentences are analogous to linguistic ones, they should be entities that are made up of grammatically-related lexical items (just as grammatically-related words are combined to form linguistic sentences). Like linguistic sentences, they should also be complete in a certain sense, i.e. a musical sentence is the final output of grammatical generation – all subsequent generation is of new sentences that are not grammatically dependent on previous ones. (This aspect of sentence structure is captured by the lay notion that sentences are defined to a large extent by the periods (or question marks, exclamation points etc.) that end them, and the capital letters that subsequently begin new ones.) Finally, sentences are said to have a certain logical structure too, i.e. they often contain a predicate that is logically related to the

other parts of the sentence, which are called its *arguments*, examples of which are the subject, and the direct and indirect objects, of the sentence. (“Predicate” is being used here in the way it was used in Gottlob Frege’s work, which is also the way it is used in modern logic. An older, Aristotelian, use of the term also exists in which a “predicate” is the second half of a sentence whose first half is the subject.)

Although musical grammarians have long discussed how sentence-like musical structures are generated by combining things like chords (which I have proposed as being the foundation for the musical lexicon), determining when such a generative phenomenon is complete, and whether the output of this phenomenon has a logical structure (both purported attributes of linguistic sentences) remain tricky problems for music theory – thus leading to the inconclusive nature of this discussion. I will deal with the logical aspect of this issue first.

To look at the matter in slightly technical terms, when a predicate assigns a property to one of its arguments, it is in fact assigning what is called a *thematic relation* to it, which is a semantic property of the given argument. So, the predicate “smiled” can assign the thematic relation of *agency* to its subject “Leila” in the following sentence, if Leila is indeed the one doing the smiling:

(2a) Leila smiled.

Given the above structure, a sentence can be given an analysis that reveals how the different parts of the sentence are logically related to each other (this is the so-called “predicate calculus”). Such an analysis also amounts to a semantic analysis of the sentence, since the thematic relation assigned to the subject of (2a) by its predicate asserts something meaningful about it, i.e. that Leila is the doer of a certain action. Finally, given this connection to meaning, the logical structure of a sentence also asserts something about its truth. That is, (2a) can only be true if there is actually something or someone called “Leila” that exists, and who also performed the action of smiling – otherwise the sentence would be false at best (if Leila did not smile), or meaningless at worst (if no such thing/person as Leila exists), depending on the theory of meaning one subscribes to. A *formal semantic* analysis of the sentence therefore depends on the sentence having a logical structure, which can in turn be analyzed using the software of the predicate calculus.

Minimalist linguistics has a particular take on linguistic meaning, so it does not necessarily accept all the details of the logical approach to sentence structure – especially those parts of this approach that assert that a sentence’s logical structure directly connects language to objects in the external world (such as “Leila”), and in particularly propositional or truth-functional terms. Moreover, Minimalist linguists do not necessarily see the relation between a predicate and its arguments in exclusively semantic terms as Frege did, since the semantic mapping between a predicate and the arguments it takes is not always clear. For instance, the predicate “eat”, like “smile” above, can occur in an intransitive form (as in “Leila has eaten”), but it can also occur in a transitive form, where it takes two arguments, viz. a subject and an object (as in “Leila has eaten her dinner). Semantically speaking, both sentences have to do with Leila eating something, though the nature of this something is not mentioned in the first sentence with the intransitive predicate. So, the difference between the sentences cannot be teased apart in purely semantic terms, and depends partly on the grammatical context of the given sentences too. (Moreover, the presence of a subject in a sentence has to do with a purely grammatical constraint on sentence generation called the “Extended Projection Principle”, which we will explore later in this chapter.)

As a result, Minimalism treats predicates in more grammatical terms. For example, it stipulates that every predicate assigns a bundle of thematic relations to each of its arguments, called a “theta role”, and there is a one-to-one mapping between the arguments of a sentence and their theta roles. And this has a grammatical basis. So, if a sentence has more arguments than theta roles, or more theta roles than arguments, it will be ungrammatical. (This is known as the “Theta Criterion” (Carnie (2002): 168-172).) In (2a), the predicate “smiled” is an intransitive verb, and therefore assigns only one theta role, to the subject of the sentence.¹⁰ So, if a sentence has an intransitive verb such as “smile” as its predicate, but also has two arguments, it will be ungrammatical:

(2b) *Leila smiled the sandwich. (The * here is the conventional symbol for an ungrammatical sentence)

¹⁰ The phenomenon of verbs having only a certain number of arguments, and that these arguments must conform to certain grammatical categories (such as subject, direct object etc.) too, is called *subcategorization*.

The two arguments here (i.e. the two noun phrases, “Leila” and “the sandwich”) outnumber the single theta role that the predicate “smiled” assigns, leading to a violation of the Theta Criterion and the ungrammaticality of the sentence. However, a predicate that takes two arguments (any transitive verb, such as “ate”) would work perfectly well in place of “smiled” in (2b):

(2c) Leila ate the sandwich.

Now, if we replace “ate” with a ditransitive verb (such as “gave), which takes *three* arguments, the number of theta roles assigned by this verb would outnumber the arguments in (2c), also leading to a violation of the Theta Criterion, and an ungrammatical sentence. (This is the opposite scenario to the one presented in (2b)):

(2d) *Leila gave the sandwich.

This can be fixed by adding another argument to the sentence to match the remaining theta role:

(2e) Leila gave the sandwich to John.

So, logical structure matters in how we define a sentence, even if not defined in exclusively semantic terms as Frege did. Moreover, this logical structure has something to do with the lexicon too, since the argument structure of a sentence (also called the “theta grid”) is said to be contained within the lexicon, in addition to the meaning, pronunciation and syntactic category information of a word (i.e. whether it is a noun, verb, adjective etc.) (Carnie (2002): 173). Now musical structures clearly do not have predicates and arguments, since they do not have verbs and nouns. However, it is possible that music has a lexicon, especially a chordal one as argued previously – and Aniruddh Patel says that music does have a logical structure, and this has something to do with its chordal makeup too (at least in the case of Western harmonic tonality). Specifically:

“The harmonic function of a chord derives from the context and its relation to other chords rather than from intrinsic properties of the chord itself. Typically, three such functions are recognized: tonic, subdominant and dominant, prototypically instantiated by the I, IV, and V chords of a key, respectively.

The same chord (e.g., C-E-G) can be a tonic chord in one key but a dominant or subdominant chord in other keys, and empirical research shows that listeners are quite sensitive to this functional difference ... Conversely, two distinct chords in the same key – for example, a IV chord and a II⁶ chord – can have the same harmonic function by virtue of the way in which they are used. The salient point is that a chord's harmonic function is a psychological property derived from its relation to other chords. Thus music, like language, has a system of context-dependent grammatical functions that are part of the logical structure of communicative sequences.” (Patel (2007): 266)

In saying that harmonic function does *not* derive from the intrinsic properties of a chord, Patel seems to be suggesting that this, logical, aspect of musical structure does not derive from the purported musical lexicon. The linguists Jonah Katz and David Pesetsky concur with this view, since they specifically state that music does not have a lexicon, but that it (or least Western tonal music) does have harmonic function (Katz and Pesetsky (2011): 57-64). Moreover, they say that harmonic functions characterize the semantic aspect of music, since they must be interpreted from pitch structure rather than being directly represented in it, much in the way that the semantic aspects of a linguistic sentence are interpreted from the words that make it up. (We shall explore their argument a bit more in a subsequent section.) So, Aniruddh Patel seems to believe that music has a logical structure, and Katz and Pesetsky strengthen this argument by giving a semantic explanation for harmonic function – which is similar to the kind of explanation that linguists give for the logical structure of language.

This still leaves the problem of the logical structure of music not having a lexical connection, which is the opposite of what is said to be the case in language. However, Martin Rohrmeier has argued that harmonic function in music – the alleged locus of music's logical structure – is indeed connected to chord structure (Rohrmeier (2008)), which is my alleged locus of music's lexicon too. Unfortunately, Katz and Pesetsky's above model of harmonic function challenges Rohrmeier's conclusions – so the matter of whether music has a logical structure, whether this structure is inherent in harmonic function, and whether any of this has anything to do with the purported musical lexicon (i.e. chords) remains unresolved. In my later discussion of Katz and Pesetsky's model, I will argue that the situation is slightly different, viz. that, in line with Rohrmeier and contra Katz and Pesetsky, aspects of harmonic function *are* present in chord structure, but not harmonic function in the Riemannian sense as both authors seem to

understand it. Rather, it is harmonic function in the Schenkerian scale-step sense that is part of chord structure – and in contrast, the Riemannian sense of harmonic function has more to do with the semantic aspects of music, as Katz and Pesetsky correctly assert. However, none of this has any bearing on the connection between chords, harmonic function and logical structure seen together – and we will have to leave it at that.

So, it is possible that musical ‘sentences’ have a logical structure, and it is likely that this has to do with the relation between harmonic functions like tonic and dominant – although this argument is inconclusive at the moment. But what about the stipulation that musical sentences be “complete”? In the case of language, the simplest complete sentences are often considered to be those with a subject and a predicate, i.e. a *clause* (Carnie (2002): 33). (“Predicate” is being used here in its Aristotelian sense, mentioned earlier, which is distinct from the way it is used to talk about thematic relations and theta roles.) Following the analogy between harmonic function and logical entities such as subjects and predicates, we could say that the simplest, complete musical sentence would be one with, say, a tonic and another tonic, or a tonic and a dominant.

But this opens up a huge can of worms. Schenker himself thought of musical sentence structure in “tonic-dominant” terms, but his views on this varied a lot throughout his life – to the extent that his later views on this subject (which are also the ones he is best known for) consider entire pieces (or movements within multi-movement pieces) to be musical ‘sentences’ of a sort. In other words, the later Schenker understood entire pieces as being generated from a simple I-V-I *Ursatz* form. Schenker analyzed many complex works by Beethoven, Mozart, Haydn, Brahms and others to reveal *Ursatz* structures in them, particularly in his later texts such as the *Five Graphic Analyses* (Schenker (1969)) and *Free Composition* (Schenker (1979)) – through which he hoped to show how a complex piece, organized into multiple hierarchical levels and with a very intricate, recursive layering of phrases, is ultimately generated from a three-chord fundamental structure. He believed that such a revelation would highlight the deceptive

simplicity of the masterwork, and – as we saw in the last chapter – the organic, unified approach to composition (and ultimately the creative genius of the composer) required to create such a masterwork.

This makes it really hard to conclude whether Schenker's *Ursatz* is really a grammatical structure, i.e. a musical sentence, or whether it is really a poetic structure, akin to a literary work, such as a poem or a novel (or a stanza or chapter within these). As discussed in the last chapter, this is also what makes it hard to accept the *Ursatz* as a primitive in an axiomatic system of musical grammar. A related problem is that whereas a sentence (in language) is usually a relatively short structure (the mathematical linguist András Kornai estimates median sentence length to be in the vicinity of 15 words (Kornai (2008): 188)), a poem or a novel, or even a section thereof, can be many pages long. Therefore, a long musical piece, or movement thereof, would seem more analogous to a literary work – and an *Ursatz* form that describes such a piece would therefore appear to be more a poetic structure than a sentence.

Now sentences can obviously be much longer than 15 words. Steven Pinker cites a sentence by George Bernard Shaw that is 110 words long (Pinker (1994): 77). Such complex, florid constructions are not unusual when written by literary figures for use in literature, as opposed to in quotidian discourse, where 'getting to the point' might be more important. Given that Schenker's analyses are also of works of art, the fact that the constructions he examined are longer, and more complex than is usual should probably not be surprising. And even Noam Chomsky has remarked on the lack of relevance sentence length has to grammatical theory:

"There is no longest sentence (any candidate sentence can be trumped by, for example, embedding it in 'Mary thinks that. . .'), and there is no non-arbitrary upper bound to sentence length." (Hauser, Chomsky and Fitch (2002): 1571)

One could respond that another issue here is that of memory – i.e. if a construction is too long it cannot be held in memory and thus loses its practical use in communication. But it is worth remembering that the description of music and language we are considering here is of the human computational system of music and language – i.e. the system behind musical/linguistic *competence*. So, the use of musical or linguistic constructions in communication, and performance limitations (like those of memory) on such a

phenomenon, are not of any particular relevance to our discussion – what is relevant is what is *possible*, what C_{HL} and C_{HM} are capable of generating when left to their own devices. Also, the music Schenker examined was largely written down in the form of notated scores, where memory limitations are not an issue anyway (since one can just flip back the pages of the score to refresh one’s memory of a musical passage, something that cannot be done when engaged in verbal conversation with someone). So, the length of the musical ‘sentences’ that Schenker analyzed, as arising from a background *Ursatz* form, does not necessarily impede an interpretation of this aspect of Schenkerian theory in grammatical terms.¹¹

However, it just happens to be a fact that even the sentences that generative linguists examine are normally of the ‘approximately 15 words’ variety, rather than elaborate prose renderings from the world of literature. So, basing a theory of musical grammar on long pieces of music seems to be at least a different pursuit in degree, if not in kind, than that pursued by linguists and other cognitive scientists. In fact, this is one of the reasons that Lerdahl and Jackendoff rejected Schenkerian theory as the ultimate basis for their own work in musical grammar:

“Although this *a priori* construct [i.e. the *Ursatz*] was understandably central to Schenker, a thinker steeped in 19th-century German philosophical idealism, its status made little sense to a modern, scientifically inclined American. ... The *Ursatz* is too remote from a musical surface to be picked up and organized by a listener who is not already predisposed to find it.” (Lerdahl (2009) : 187-188)

Despite the above caveat, Lerdahl and Jackendoff’s own definition of a musical sentence is not any more conclusive or ‘post-Schenkerian’ in its formulation – which just goes to show that rejecting Schenkerian constructs on ideological grounds in favor of a more ‘modern’ approach to musical grammar is not necessarily an easy task, given how difficult it is to establish what a musical ‘word’ or a musical ‘sentence’ is. This is evident from the way Lerdahl and Jackendoff formulate their second Grouping Well

¹¹ In this context, it is worth reviewing the Schenkerian music theorist Poundie Burstein’s reminder that, “It should be noted, however, that the popular association of large structures with Schenkerian analysis is an exaggeration. Schenkerian analysis tends to put no more emphasis on large structures than do many other popular methods of tonal analysis. Many other analytic systems evoke structures that are as large or larger than ones discussed by the typical Schenkerian analysis. For instance, many non-Schenkerian analytic approaches propose huge tonal plans than embrace multi-movement compositions or even entire operas. In contrast, a typical Schenkerian analysis discusses a single movement or a passage within a single movement, and most of Schenker’s own published analyses focus on works or passages that last not much more than a minute at most.” (Burstein (2010): 9)

Formedness Rule (which states that “a piece constitutes a group”), which seems to reveal the same concern for (poetic?) unity in describing the structure of an entire piece that Schenker had (and that made Schenker formulate the *Ursatz* to begin with):

“The second rule expresses the intuition that a piece is heard as a whole rather than merely as a sequence of events.” (Lerdahl and Jackendoff (1983): 37)

Of course, one can conceive even of smaller constructs like phrases and sentences as unified wholes – in fact, that is the very point behind the idea that a sentence should be complete. The only problem is in deciding how complex such a structure has to be before being considered sentence-like in the case of music. As I suggested above, the matter seems to be about deciding the grammatical status of sentence-like structures without regard for the coherence of the poetic entities that result from these structures. In other words, it seems to be a matter of deciding whether certain constructions are sentences based on whether they have grammatical closure (i.e. are well-formed), without concern for whether the larger, literary entity one creates with their help has poetic closure (i.e. unity). (The musicologist Nicholas Cook has referred to this dilemma as the problem of differentiating music as a language versus music as literature (Cook (1989b)).) Unfortunately, neither later Schenkerian theory nor Lerdahl and Jackendoff’s approach seem to be able to resolve this dilemma, since neither is able to arrive at a description of smaller musical constructions without extending this description *ex hypothesi* to entire pieces too.¹²

However, there are reasons to doubt the sentence-like status of *Ursätze* that last for entire, long, pieces though, on purely grammatical grounds. One of these reasons can be found in the work of certain scholars who resist Schenkerian approaches to musical structure, such as those who work within the paradigm that is known as neo-Riemannian theory. In the last chapter, we briefly explored the ideas of Richard Cohn in this context, and Cohn has suggested that it is better to look at entire pieces as ‘star clusters’ of tonal areas, rather than as unified wholes, because each area often has no grammatical

¹² On a slightly different note, it is worth noting that the grammar-versus-poetics issue has attracted some interest in language scholarship too, given the interest shown by some linguists in the notion of a “discourse grammar” (e.g. Van Dijk (1972, 2003), Polanyi (2003)), which contrasts with the generative-grammatical interests of the Chomskyan tradition.

relationship with the others – and so they have to be seen as separate regions or sections within a piece (e.g. see Cohn (1999)). (This is akin to how the sentences in a paragraph have a complete grammatical structure of their own, but not to each other – and therefore each sentence forms a discrete expression within a larger narrative.) The reason tonal areas often have no grammatical connection with each other is because the standard functional relationships of tonic, dominant etc. cannot be applied between them, which renders impossible a unified grammatical analysis of the whole piece in which these areas occur.

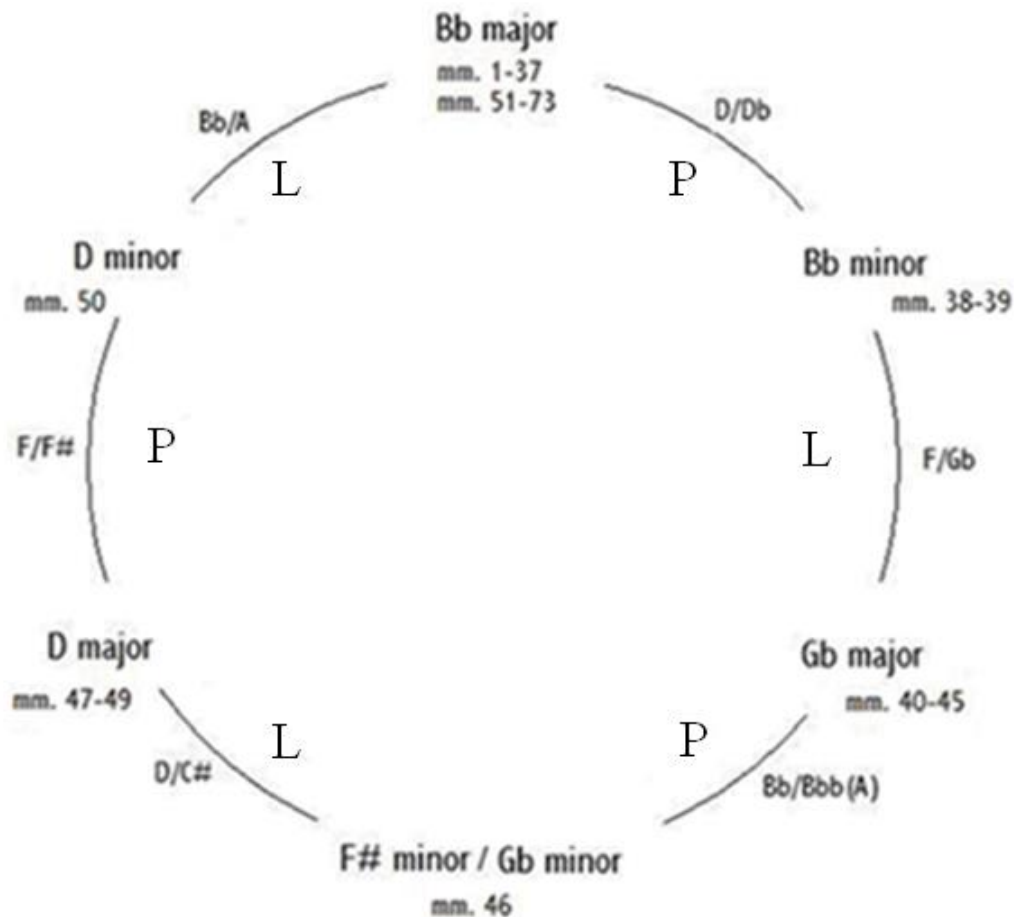
To understand this, consider Example 1.2-10, which represents the different tonal regions in the second movement of Beethoven's fifth sonata for violin and piano, the Op. 24 "Spring" sonata. This movement is 73 measures long, and the first 37 measures are in the home key of B-flat major. In measure 38, the music modulates to B-flat minor, by exchanging the D-natural of the B-flat-major triad for the D-flat of the B-flat-minor triad.¹³ Since B-flat minor is the parallel minor of B-flat major, the change from B-flat major to minor is called a "parallel transformation" (P) in neo-Riemannian theory (not to be confused with the *Parallelklang* in the original theory of Hugo Riemann, which actually refers to a transformation between a major key and its *relative* minor key, or a minor key and its relative major). Two measures later B-flat minor modulates to G-flat major. This is accomplished by exchanging the F of the B-flat-minor triad with the G-flat of the G-flat-major triad. Since F is the leading tone of G-flat major, which is exchanged with G-flat to effect the modulation, this transformation is called the "leading tone exchange transformation" (L) or *Leittonwechselklang* in the original theory of Hugo Riemann.

After this transformation, another parallel transformation takes us from G-flat major to G-flat minor, where the exchanged pitch B-double-flat is enharmonically reinterpreted as A-natural, and the G-flat minor tonal area as F# minor (Richard Cohn refers to this as traveling through the "enharmonic seam"). Another L transformation turns F# minor into D major, and then another P transformation takes us from D major to D minor. Finally, one last L exchange returns us to the home key of B-flat major from

¹³ To change the key of B-flat major into B-flat minor, the G-natural of B-flat major has to be lowered to the G-flat of B-flat minor too. But to change the tonic *chord* – the governing chord of B-flat major – into the tonic chord of B-flat minor, only the D/D-flat exchange needs to happen. This point will apply to the further chord transformations explored in the above Beethoven movement.

D minor. In all, we travel through six key areas, and each transformation from key area to key area involves the exchange of just one triadic pitch with another just a semitone above or below it – which is an example of very parsimonious or smooth voice leading between triads. Therefore, the cycle of keys the

Example 1.2-10. Beethoven, Violin Sonata #5 “Spring”, Op. 24/ii: “Cohn Cycle” distribution of keys



movement goes through is called a “maximally smooth hexatonic cycle” (Cohn (1996)). (Given Richard Cohn’s contributions to the theory of maximally smooth hexatonic cycles, they are often referred to as “Cohn cycles” as well.)

Just by looking at the key areas of this movement we can see that the harmonic functional relationships between them are not of the tonic/dominant sort, but rather involve P and L relationships.

This already reveals how difficult it might be to describe the grammatical structure of this entire movement in a unified way. Schenkerians do warn though that key areas should not be treated as equivalent to structurally important events in the grammatical structure of a piece, so even if the relationship between, say, the first two key areas is P, it could be the case that there is an F major harmony somewhere in the B-flat minor section that acts as the dominant to the home key of B-flat major – and can therefore form a hierarchical, grammatical relationship with it. But even if that is the case, once we get to the enharmonic seam between G-flat major and G-flat minor/F# minor, things change, because F# minor is not a scale-step in B-flat major, and so *cannot* have a grammatical relationship with it. G-flat minor on the other hand, can, being the flat-6 scale-step of B-flat – but if we continue reading G-flat minor as G-flat minor, and not F# minor, then its relation to the subsequent L-related D major falls apart.¹⁴ So, we might be able to force a traditional, scale-step based functional reading on this Beethoven movement, but it would have to be at the expense of the, fairly explicit, hexatonic organization of its tonal structure. In other words, we might be better off understanding this piece not as one unified musical sentence, but rather as a compound (or network) of six sentences, each separated by a musical ‘punctuation mark’, viz. the smooth voice leading exchanges that take us from one sentence to the next.¹⁵

In fact, each key area might very well be treated as a sentence when we look at some of the other features of the piece, such as its thematic material and meter. Consider Example 1.2-11, which depicts mm. 38-45 of the movement, the measures that realize the B-flat minor and G-flat major key areas of the piece. These measures quite clearly contain a varied form of the eight-bar main theme of the movement, albeit transposed to a different key. So, the material here is melodically complete (it’s the complete

¹⁴ Moreover, if we continue reading G-flat as G-flat and not F#, then we will have to analyze, for consistency, future occurrences of F – the dominant of the home key B-flat – not as F but as G-double-flat. In other words, the supremely important tonic-dominant relationship between B-flat and F, in places like the final perfect authentic cadence of the piece, would be reduced to a strange, chromatic relationship between B-flat and G-double-flat. Of course, one could say that under equal temperament F and G-double-flat are enharmonically equivalent – but then why not interpret the earlier G-flat minor as the enharmonically-equivalent F# minor too, as neo-Riemannian theory says we should anyway?

¹⁵ Richard Cohn explores a similar ‘network-based’ approach to the issue of unity in an earlier paper with the philosopher Douglas Dempster (Cohn and Dempster (1992)).

Example 1.2-11. Beethoven, Violin Sonata #5 “Spring”, Op. 24/ii: Harmonic structure of mm. 38-45

mm 38 40 42 44

tr

Violin

Piano

Bb-minor: i ——— i⁶ ——— v³ ——— i ——— IV ——— V⁴ ——— i ——— 8/6 ——— 7 ——— 5 ——— I

Gb-minor: iii ———

theme), and it is metrically complete too, as it is in the form of a canonical eight-bar phrase, which might be subdivided into two, well rounded, four-bar phrases.¹⁶ Harmonically, it contains the entire B-flat minor and G-flat minor key areas, but we could interpret the whole passage as being in G-flat major, as the example illustrates, in which the initial two bars in B-flat minor are analyzed as iii in G-flat major. This would imply that the harmonic structure represents the complete G-flat major tonal region of the piece, which is perfectly correlated with the well-formed melody and meter of this passage.

Finally, even though there may not be a grammatical relationship between some of the key areas of the movement, the material *within* each key area reveals a conventional, hierarchical tonal structure. In fact, non-hierarchical, neo-Riemannian phenomena are known to appear more at *middleground* levels of musical structure, meaning that they usually involve sonorities that are often far away from each other in the surface of a piece, such as the sonorities that mark different sections of a piece. This means that sonorities closer to each other in the musical surface are usually hierarchically-related as well, and therefore have a conventional grammatical structure. Example 1.2-11 shows this to be the case for G-flat major, since the sonorities in this passage are all conventional tonic, dominant and subdominant harmonies (except for the B-flat minor iii harmony at the beginning), which therefore have the conventional hierarchical organization that such harmonies do.

It might be worth remembering at this point that Schenker did not always view a whole piece as one unified structure, especially in his earlier thinking, and we can conceive of musical sentence-like structures that have a ‘tonic-dominant’ construction but that do not encompass entire pieces. In the previous pages, this seems to be the case for at least key regions within a piece. So, even within a Schenkerian perspective we could propose that a musical sentence is a section of a piece that is in one key – defined clearly by structurally-important harmonic events before and after the section that act like

¹⁶ Note my use of the language-influenced term “phrase” here. This is a conventional usage, especially when describing much-discussed and analyzed structures like the eight- and four-bar ones I discuss above. However, it is also a problematic usage because of the overlap with its use in describing linguistic sentence structure. Therefore, my use of the word requires a clarification, which I shall give in the course of the next few pages.

punctuation marks, such as sentence-ending perfect authentic cadences and sentence-starting modulating progressions – both of which we see in the case of the G-flat major passage in Example 1.2-11.

It is interesting though that music theory has always referred to musical passages that have such punctuation marks (e.g. those that end with cadences) with terms that suggest an implicit acceptance of sentence-like structures in music. Consider the concept of the Classical period, with its constituent antecedent and consequent *phrases*, each ending with a cadence, the second giving more closure than the first. (The term “phrase” being one I used myself a few paragraphs ago, when describing the metrically-, harmonically-, and melodically-complete structure of the Beethoven Spring G-flat major passage above.) Even more explicit is the notion of the Schoenbergian *Satz*, which is usually translated as “Sentence”. The *Satz* is also defined unequivocally as a structure in which only one cadence appears, at the end of the structure, to give it harmonic and melodic closure (Caplin (1998): 45).

Now, these uses of “sentence” and “phrase” are not equivalent to their linguistic use of course. For one, we have already seen the difficulty of applying the “sentence” label to anything in music. Moreover, a “phrase” in linguistic terms is a construction that ‘centers’ around a word – in the way the noun phrase “the brave soldier” is built up around the noun “soldier”, which is what the phrase is about. (As a result of this, a tree-diagrammatic representation of this phrase would make “soldier” the most hierarchically-superior structure in the phrase, represented by the tallest branch in the tree. Such a hierarchically-superior structure is called the *head* of the phrase in the technical vocabulary of linguistics.)¹⁷ If we continue my previous analogy between words in language and chords in music, a musical phrase should then be a structure that is built up around a chord. An example of this might be the cadential 6-4 progression, which is ‘built up’ around a dominant chord – since it involves two sonorities, the first one (the actual 6-4 sonority) serving to prolong the dominant chord that normally follows it. So, the term “phrase” should properly be applied to the “V6/4 – 5/3” complex that represents a cadential 6-4 progression, and the progression itself should be called a cadential 6-4 phrase. However, this is unheard

¹⁷ The idea that a noun phrase is really about the noun, and thus headed by the noun, is controversial in generative linguistics, since many linguists take the determiner to be the head of the phrase – which should therefore be a determiner phrase. I will return to this issue in the next section.

of in music theory. Instead, “phrase” is used for structures like the antecedent and consequent parts of the Classical period, which are both actually made up of multiple phrases (in the linguistic sense), each phrase prolonging the different scale-steps that constitute these structures.

Given the popular and frequent use of “phrase” (in the musical sense) in music theory, I will continue to use it from time to time, e.g. when referring to antecedents and consequents. However, I will sometimes use the term in its true linguistic sense as well, when describing a specifically grammatical aspect of musical structure. The context should clarify which use of the term is intended at the moment.

But what the above discussion implies is that a better term for antecedents and consequents is probably the very term we have been trying to define in this section – viz. “sentence”. Antecedents and consequents are usually in one key (our working definition of “musical sentence” above), and they usually have a conventional hierarchical harmonic structure (which therefore reveals a grammatical relationship between its constituents – another feature of sentences). To the extent that these constituents have tonal-harmonic functions like tonic and dominant, which they usually do, antecedents and consequents might be considered to have a logical structure too. Finally, since they are defined to a large extent by the cadences that end them, they are complete structures as well. Which implies that the sentence-like nature that has long been implicitly ascribed to antecedents and consequents seems to be justifiable in more explicit, linguistic-theoretical terms too.

The situation is even better for the Schoenbergian *Satz*; not only has it always been referred to as “sentence”, William Caplin actually compares it to actual linguistic sentences, given that its first part (called the “presentation phrase”) is unclosed and seems to set up a thought, like the subject of a linguistic sentence, while its second part (the “continuation phrase”) ends with a cadence and completes the previously set up thought (Caplin (1998): 45).

Finally, Caplin also states that “it is rare for a period to be embedded within a period, or a sentence, within a sentence”.¹⁸ This is an important point because it relates to the critical, recursive nature of musical/linguistic grammar. In the discussion in the last chapter, we saw how recursion allows subordinate clauses to be embedded within main clauses – all within the generative process of sentence construction. However, a main clause is almost never recursively embedded within another main clause, since the limits of the main clause of a sentence represent the limits of that sentence. That is, once a main clause, along with all of its subordinate clauses, has been generated the sentence is complete – the generative process then moves on to the next sentence.¹⁹ So, the fact that Classical periods and especially Schoenbergian sentences are almost never embedded within each other suggest that they are akin to main clauses – i.e. they represent complete structures, where the generative process has reached its limit. In this light, they are the closest we can come to defining what a musical “sentence” is.²⁰

Despite this positive outcome, defining the musical sentence, as stated at the beginning of this section, will have to be inconclusive. This is because even if periods and Schoenbergian sentences display many of the characteristics of a true (linguistic) sentence, there does not seem to be a way to prevent these characteristics of theirs from being generalized to entire key areas in a piece (which we also considered “sentences” earlier in this section). This is similar to one of the limitations of later Schenkerian approaches to sentence structure, as we saw earlier too. That is, if a smaller sentence-like structure is described as such, there is nothing in later Schenkerian theory that prevents this description from being

¹⁸ William Caplin, “Classical Form and Recursion”. Accessed at <http://lists.societymusictheory.org/pipermail/smt-talk-societymusictheory.org/2009-March/000097.html>, on August 5, 2012.

¹⁹ The situation is a bit more complicated than this. As we will see in the next section, generative linguists consider main clauses to be essentially subordinate clauses, but without a complementizer like “that” before them – or rather, main clauses are subordinate clauses with a *null* complementizer. So, a main clause with an *overt* complementizer *can* be embedded within other main clauses in language. But Caplin’s point in the case of musical main clauses would still apply until we demonstrate what a ‘musical complementizer’ is, because without such an entity musical main clauses cannot be embedded within other musical clauses.

²⁰ This is also the closest we will come to distinguishing my concept of the Headline from the Schenkerian *Urlinie* on which it is based. That is, an antecedent or consequent phrase, or a Schoenbergian *Satz*, can exemplify a Headline, in that they appear to be truly grammatical, sentence-like structures. Schenker’s *Urlinie* on the other hand, can span an entire work, and therefore is quite possibly a poetic, rather than grammatical, entity. (Although, as we have seen, even smaller structures that are antecedents etc. in their own right could be treated as *Urlinien* in earlier Schenkerian theory – in which case *Urlinien* would be equivalent to Headlines.)

Example 1.2-12. Beethoven, Piano Sonata #19, Op. 49 No. 1/ii: Sentence structure of main theme

The image displays a musical score for the main theme of Beethoven's Piano Sonata #19, Op. 49 No. 1/ii. The score is written for piano and is in 6/8 time with a key signature of one sharp (F#). The music is divided into two main sections: 'Sentence 1 = Antecedent' and 'Sentence 2 = Consequent', which together form a 'Period'. The 'Antecedent' section begins with a piano (*p*) dynamic and ends with a half cadence (HC). The 'Consequent' section begins with a fortissimo (*sf*) dynamic and concludes with a perfect cadence (PAC). The score is presented in a grand staff format, with the treble and bass clefs joined by a brace. The notation includes various rhythmic values, accidentals, and dynamic markings.

extended to the entire piece in which the smaller structure occurs as well – in fact Schenker would probably encourage such a generalization, given his interest in revealing organic unity in a masterwork. And we have the same problem here – all the characteristics of the period and the *Satz* that makes them sentence-like seem to be extendable to entire key areas too, since key areas are also made up of functional harmonic characteristics, also display closure (often through the same cadences that close the phrases, in the musical sense, within them), and are in one key by definition.

The situation does not improve when we consider the phenomenon of recursion in this context. Periods and *Sätze* cannot usually be embedded within themselves, but neither can key areas in many instances, for example in the Beethoven “Spring” violin sonata example we looked at a little while ago. However, this stipulation is not hard and fast, since in other instances key areas are often considered to be ‘embeddable’ within other key areas. (In fact, the organic unity of a piece in later Schenkerian theory is based on this very characteristic, viz. of secondary key areas being embedded within a larger home key that ‘bookends’ the entire piece.) And periods and Schoenbergian sentences can *also* be embedded within each other as well, if not within themselves. Take, for example, the theme from the second movement of Beethoven’s Op. 49 No. 1 piano sonata, shown in Example 1.2-12. This theme is a conventional period, made up of an antecedent and a consequent phrase – but both of which are Schoenbergian sentences themselves.

So, there is no conclusive way of defining a musical sentence in period/*Satz* terms that cannot also be applied to key areas. Therefore, the definition of a musical sentence, in the end, is still up for grabs. For the sake of convenience in this dissertation, however, I will at least assume that small, conventional structures such as 4- and 8-bar antecedents, consequents, and Schoenbergian sentences fit the bill of a musical sentence adequately. For this reason, the vast majority of musical examples in this dissertation, particularly those that attempt to illustrate grammatical phenomena in music, will be of the ‘4 or 8 bar antecedent or consequent’ variety.²¹

²¹ There is one specific, stylistic, advantage in taking 4- or 8-bar structures to be musical sentences. This has to do with the stylistic fact that expanding simpler 4- or 8-bar structures into more complicated ones, usually through

Having discussed the inputs and outputs of musical grammar, it is now time to get our hands dirty with the actual operations of the computational system of music – the specifics of musical grammar itself. In order to facilitate this discussion, I will now provide a brief history of ideas and techniques developed by generative linguists to understand *linguistic* grammar, as they developed from the earliest days of generative linguistics to the current Minimalist Program. This history will give us a toolkit with which to explore musical grammar more rigorously, and also with which to compare it to linguistic grammar.

1.2.3. A Brief Overview of C_{HL} in Linguistic Theory²²

To start, let us review some basic premises and goals of generative linguistics, which have persisted to this day. First of all, generative linguists argue that humans have an innate knowledge of language (i.e. I-language or Language), which comprises the human psychological faculty of language, and which allows them to acquire their native languages (i.e. E-language) unconsciously and effortlessly, as long as this happens within a critical period of youth. It is this unconscious, innate knowledge of language that makes humans competent in their native languages at an early age too. This occurs even when no explicit instruction in the language is provided – which could happen, for example, when a child grows up in a society where it is uncommon for children to be addressed at all (as happens in some cultures), or if s/he is only corrected when something socially-inappropriate, as opposed to ungrammatical, is said, as has been documented as well. It is also this knowledge of language that allows us to generate and comprehend, in theory, infinitely long and complex sentences – which are impossible to learn consciously from one's environment in principle, given the finiteness of time and human life.

elaborate chromatic voice leading, became a prevalent feature of much music in later eras of the Western Classical tonal idiom. Examples of this are Brahms' well-documented use of 5-bar phrases, and many late Romantic passages where the completion of a phrase takes much longer than usual because of the way the final cadence of the phrase is pushed further and further back. Wagner's much-analyzed Prelude to *Tristan and Isolde* is a good instance of the latter, since its first phrase does not end until bar 17 – that too deceptively, since Wagner never allows for a full authentic cadence to appear in this Prelude. So, considering 4- or 8-bar structures as musical sentences helps us distinguish the use of such normative structures in the Classical period from the more elaborate structures of the Romantic era – something we cannot do when we define a full piece as a sentence, as happens in Schenker's more mature conception of the *Ursatz*.

²² This section is essentially a summary of the ideas presented in Andrew Carnie's *Syntax: A Generative Introduction* (Carnie (2002)). Therefore, the reader is referred to this text for a more thorough treatment of the ideas presented in this section.

Since Noam Chomsky's earliest writings on the subject, the human faculty of language has also been thought of as a computational system, C_{HL} , which recombines the building blocks of language, viz. lexical items, to generate sentential outputs, according to certain grammatical principles. And as discussed in the last chapter, more recent Minimalist approaches to grammar have described this computational system as displaying a certain economy and underspecification in how it operates, of being comprised only of components that are conceptually necessary (for it to meet certain constraints imposed on it by the conceptual-intentional and sensorimotor systems), and generating, as a result of all of this, outputs that are discrete and potentially infinite. Therefore, the first and foremost task of an adequate theory of grammar is to give an account of C_{HL} , and specifically the grammatical principles that it operates by in an economical and conceptually necessary way.

In the first work on generative grammar, done in the 1950s and 60s, C_{HL} was thought to operate according to three kinds of rules that govern how sentences are generated. (This early work in grammar is often known as the "Standard Theory", and also the "Aspects" model because it was introduced most famously in Chomsky's seminal work, *Aspects of a Theory of Syntax* (Chomsky (1965)) – much of which was a revision of *Syntactic Structures* (Chomsky (1957)), where the three above rule types were first introduced.). The three rule types are *phrase structure rules*, *transformational rules*, and *morphophonemic rules*.

Phrase structure rules (PSRs) are rules that operate on *constituents* to generate larger constituents. A constituent is nothing but a word, or a group of words that seem to work together as a unit in the way a single word does. So a noun is an example of a constituent, and so is a noun phrase, since a phrase is a group of words that is built up around a single word (a noun in this case) and acts as a unit around that word.²³ A PSR therefore operates on smaller constituents like single words to generate larger constituents

²³ As Andrew Carnie points out, constituents are not merely theoretical postulates that linguists have devised to make their theories work – they are real (more specifically, they have psychological reality), since psychological experiments have confirmed that people parse sentences into their constituents even when they are prompted not to in the course of an experiment (Carnie (2002): 31). The fact that groups of words form constituents can be easily demonstrated by just looking at some common grammatical phenomena. We have already explored the notion of movement in grammar, where parts of sentences are moved around to form, for example, questions (in the case of

like phrases. In turn, these rules can operate on these larger constituents to form even larger ones, such as phrases that contain other phrases in them (such as verb phrases, which can contain a noun phrase) and ultimately clauses, which are usually made up of a number of phrases and behave like simple sentences. (Remember that this ‘phrase within a phrase’ organization brought about by phrase structure rules reflects the hierarchical nature of linguistic grammar.)

Now words, as we know, are part of the lexicon. So, a word is made up of various lexical features, such as the meaning of the word, its pronunciation, its theta grid – and also its syntactic category, of which the most important are nouns, verbs, prepositions, and adjectives.²⁴ The Standard Theory therefore states that PSRs operate on words that possess syntactic category features of one of the above four kinds, to generate phrases from them whose names are determined by the kind of word they are operating on, viz., adjective phrases (AP), noun phrases (NP), preposition phrases (PP), and verb phrases (VP).²⁵ This leads to four PSRs for English phrases:

AP → (AP) A

NP → (D) (AP+) N (PP+)

PP → P (NP)

VP → (AP+) V ({NP/S'}) (PP+) (AP+)

wh-movement). When movement occurs, entire groups of words often move together – thus proving that they form a constituent, since they act as a unit in their movement behavior. As an example, consider the movement involved in transforming an active into a passive. The sentence *Leon ate juicy red strawberries for breakfast*, when transformed into its passive becomes *Juicy red strawberries were eaten by Leon for breakfast*, not **Strawberries were eaten by Leon juicy red for breakfast*. In other words the group of words “juicy red strawberries” (which happens to be a noun phrase) moves together in transformations – and therefore acts as a unit.

²⁴ This category can be considered to include adverbs, since many linguists do not distinguish adjectives and adverbs.

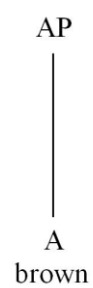
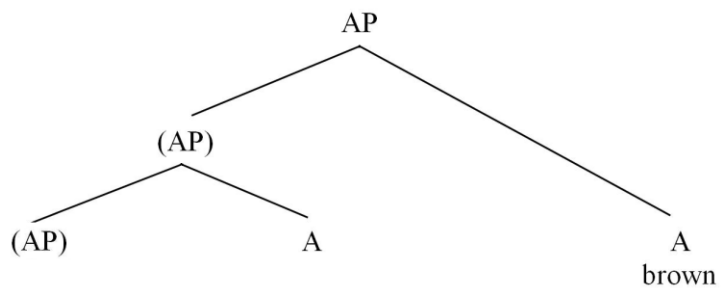
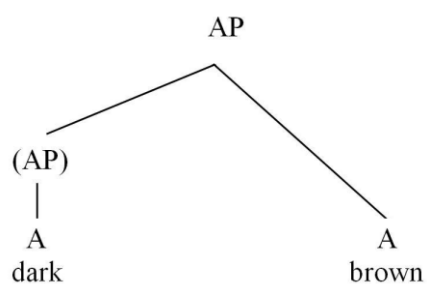
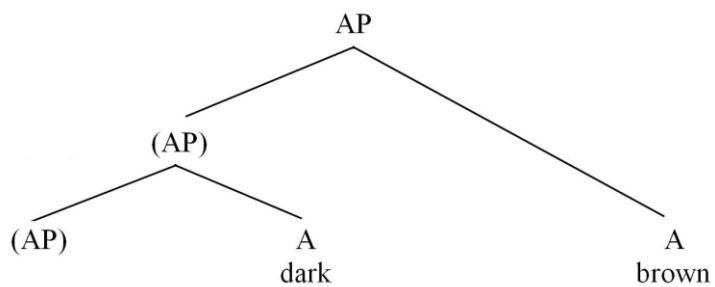
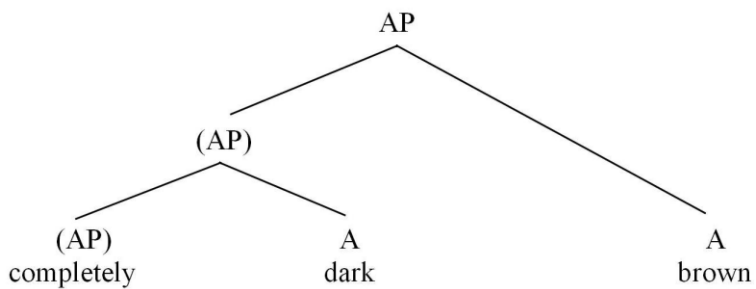
²⁵ Since the syntactic category information of a word is different from its meaning, it is important to note that “nouns”, “verbs” etc. are not semantic notions. That is, they are not determined by the meaning of the words that exemplify them. So, nouns are not persons, places or things, and verbs are not actions – since all of these are semantic characterizations of these words. Rather these categories are determined on structural grounds alone, such as what other words precede or follow them, and what kinds of prefixes and suffixes they can take. For example, in English, adjectives often end with the suffix “-ish” and appear in between a determiner and a noun, and verbs often end with the suffix “-ed” and appear after the subject noun or noun phrase. If we were to use semantic criteria to determine syntactic category membership, it would confuse everything. For example, “panting” might be considered a verb based on the fact that it appears to be an action, e.g. something a dog does. But in “The panting dog chased the postman with great gusto”, “chased” is clearly the verb (both because of its -ed ending and its appearance after the subject noun “dog”), whereas “panting” is an *adjective* that appears between the determiner “the” and the noun “dog”.

The first rule captures the intuition that an adjective phrase always has an adjective A (by definition), which is the head of that AP. However, this adjective can be modified or ‘elaborated’ with another adjective as in “the *dark brown* book”, where “dark” modifies “brown” – it says something more about the adjective “brown” itself. Moreover, an adjective like “brown” can be modified by more than one word – even an entire phrase, such as in “the *completely dark brown* book”, where “brown” is modified by a constituent that is itself an adjective phrase, i.e. “completely dark”.²⁶ In other words, an adjective phrase is a phrase that contains at least one adjective by definition, in this case “brown”, but it can optionally have another adjective phrase (which appears before it in English) which modifies the first, non-optional, adjective “brown”. The phrase structure rule “AP \rightarrow (AP) A” formalizes this by stating that an AP is formed by combining an optional AP (exemplified above by “completely dark”) with a non-optional A, with the optional nature of the AP “completely dark” being captured by the parentheses around it. (Also, notice how the definition of an AP in terms of another, optional AP in the above PSR reveals the recursive nature of APs – which is an important characteristic of language in general.)

Example 1.2-13 presents tree-diagrams that illustrate the three kinds of APs we get by applying the PSR for APs to the adjectives and adjective phrases in the previous paragraph. The tree at the top depicts the full structure “completely dark brown”, in which an entire AP (“completely dark”) itself modifies the head adjective “brown” to get the adjective phrase “completely dark brown”. But an AP can also be formed in which the head adjective “brown” is not modified by an entire phrase but only minimally by the adjective that defines *this* AP, i.e. “dark”. Therefore, the AP “completely” that modifies “dark” to form the optional AP “completely dark” is absent here, which can be represented by an empty AP branch in the tree. Alternatively this empty AP branch can be omitted altogether. These two possibilities are shown by the graphs in the middle of Example 1.2-13.

²⁶ Notice also how “completely” in “completely dark” is itself an AP, which modifies the adjective “dark” *within* the larger AP “completely dark”. This is similar to how the larger AP “completely dark” itself modifies the adjective “brown” in the even larger AP “completely dark brown”. This leads to a larger point in grammatical theory, which is that constituents that modify other constituents are always phrasal in nature.

Example 1.2-13. The AP in English: The PSR “AP → (AP) A” in tree-diagram form



Importantly, “dark” here still constitutes the optional AP in the PSR for APs – and not a single adjective A. This is because as a single (optional) adjective, it would be represented in the PSR as AP → (A) A, but this confuses it with the other single adjective “brown” that is the head of the entire phrase. (Although “dark” *can* be the single adjective head in the AP phrase “completely dark”, which itself modifies the adjective “brown”.) The moral of the story is that its being a *phrase* is what makes “dark” or “completely dark” optional, since phrases modify heads – the latter being non-optional and also single words, which, as we saw earlier, are either nouns, verbs, adjectives or prepositions. (This point will have important consequences later in our discussion of X-bar theory.)

Finally, the optional AP can be omitted altogether, so the entire AP is made up of just the single adjective that heads it, viz. “brown”. The bottom of Example 1.2-13 shows the two ways of representing this tree – either with the optional AP represented as an empty branch, or with it omitted altogether.

Given the above description of the PSR for APs, the PSRs for NPs, PPs and VPs should now be easy to understand. So, an NP is made up minimally of the (single word) noun that heads it, such as the proper name “Ofra” in (3a):

(3a) *Ofra* ran up the hill.

The head noun of an NP can also be optionally modified by a determiner D, such as “the” in (3b):

(3b) *The cat* ran up the hill.

In English, NPs can only have one determiner, when they have them at all, but they can be modified by more than one AP. So, we can have:

(3c) *The nimble cat* ran up the hill. But also:

(3d) *The nimble, crafty cat* ran up the hill.

Here “nimble” and “crafty” are two separate (and optional) APs, which both modify the head noun “cat”.

The fact that they are separate APs can be seen through another example (3e):

(3e) *The very crafty cat* ran up the hill.

In (3e) “very” is an optional AP, within the larger AP “very crafty” that modifies the head adjective of this phrase “crafty”. However, in (3d) “nimble” is not modifying “crafty” but rather the head *noun* “cat” – just as it does in (3c). So, “crafty” in (3d) is actually another AP that also modifies the head noun “cat”, alongwith “nimble”. (And remember that only phrases can modify, not single words, which is why “nimble” and “crafty” are both APs, not single adjectives).

Just as it can have multiple, optional APs, an NP can have multiple, optional PPs too, such as:

(3f) *The nimble, crafty cat, with the neat whiskers, from Ofra’s shelter* ran up the hill.

Here, “with the neat whiskers” and “from Ofra’s shelter” are both PPs (which we will discuss in a second), both of which modify “cat” – since they both say something about the cat. All of this is formalized in the PSR for NPs:

NP → (D) (AP+) N (PP+)

This just states that an NP is formed by combining a noun (the head of the phrase), with an optional (again represented with the parentheses) determiner, and multiple, optional APs and PPs (denoted by the + symbols). Moreover, in English these constituents also occur in a specific order, which is the order they appear in the PSR.

Turning now to PPs, we have already seen two examples of these in the last paragraph, viz. “with the neat whiskers” and “from Ofra’s shelter”. We already know that a PP must have at least a preposition, which will be the head of the phrase. In the two above sentences, these are “with” and “from”. Now, if we take away these prepositions from the above PPs all we are left with are two NPs – “the neat whiskers”

(which has the form [(D) (AP) N], corresponding to an NP as we just saw, with the optional PP omitted) and “Ofra’s shelter” (which has the form [(D) N], since “Ofra’s” is really some kind of determiner, as we will see later, which modifies the head noun “shelter” of this noun phrase). So, the PSR for PPs is really simple, viz. a non-optional, head preposition followed (in English) by an optional noun phrase:

PP → P (NP)

Note again how the PSRs for NPs and PPs display the recursive nature of linguistic grammar, since we defined an NP in terms of an optional PP, which itself is defined in terms of an optional NP. That is, we can rewrite the PSR for NPs:

NP → (D) (AP+) N (PP+), as

NP → (D) (AP+) N ([P (NP)]+) – through which an NP (on the left hand side of →) is defined in terms of another NP on the right. Similarly, we can rewrite:

PP → P (NP), as

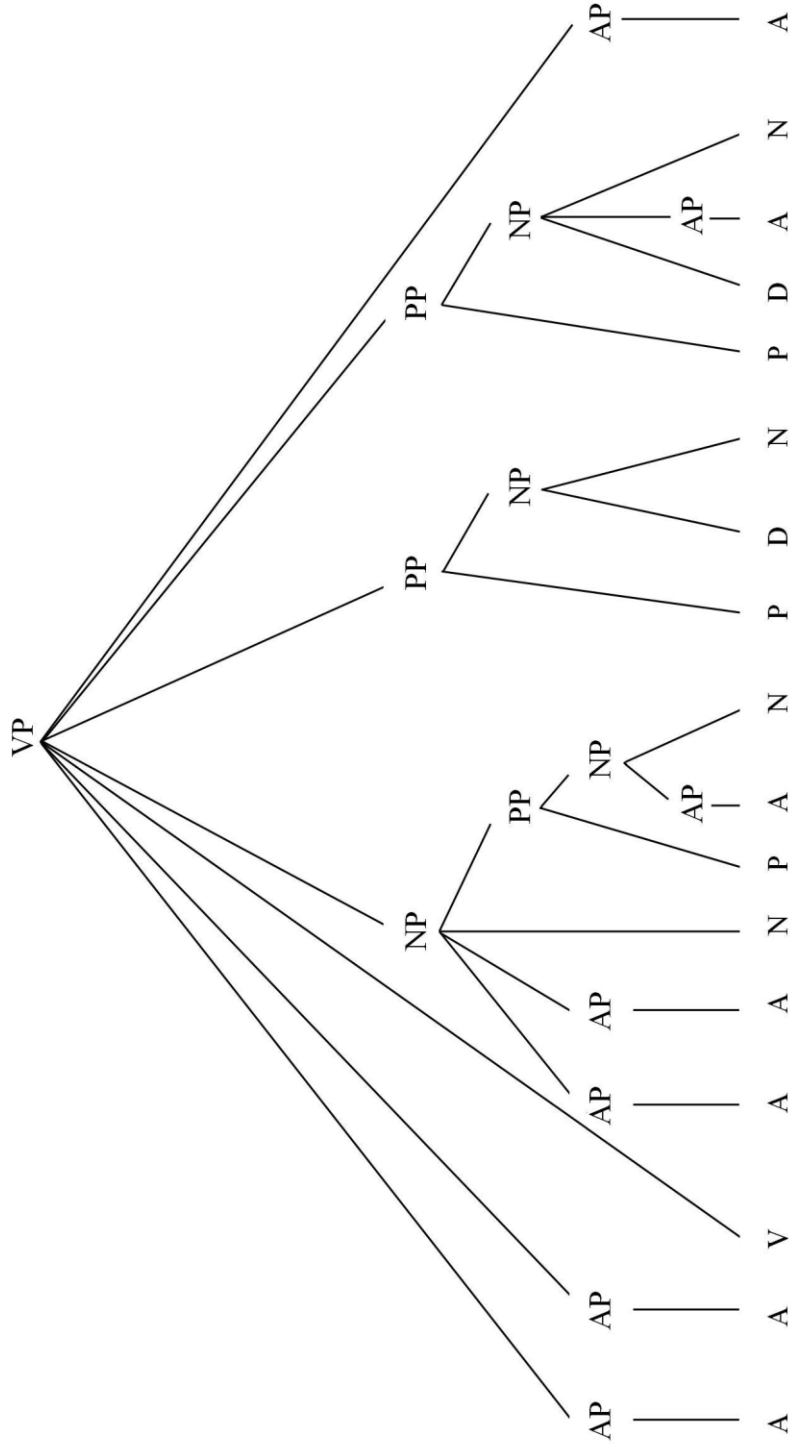
PP → P [(D) (AP+) N (PP+)] – through which a PP (on the left hand side of →) is defined in terms of another PP on the right.

This finally brings us to the PSR for VPs. VPs obviously have a single verb that heads the phrase, and this verb can be modified by multiple APs either before or after it (often called “adverbs” – or more specifically “adverb phrases”). Moreover, the head verb can be followed by an optional NP and an optional PP. This can give us the wonderfully complicated VP after the NP “the girl” in (4a):

(4a) The girl [[_{AP} always] [_{AP} quickly] [_V adopted] [_{NP} nimble crafty cats with neat whiskers] [_{PP} from Ofra’s shelter] [_{PP} with her pocket money] [_{AP} joyfully]]

The tree-diagram of Example 1.2-14 depicts the structure of this VP more clearly. (Notice how the PP “from Ofra’s shelter” is being taken to modify the verb “adopted” here and not the noun “cats” as it did in

Example 1.2-14. The VP in English: The phrase structure of a VP in tree-diagram form



always quickly adopted nimble crafty cats with neat whiskers from Ofra's shelter with her pocket money joyfully

(3f) above. That is, “from Ofra’s shelter” is being taken as the place where the girl did the adoption from, rather than where the cat with the neat whiskers is from. The two ways of using this PP leads to the issue of ambiguity in sentence structure, which I shall examine in the next section.)

In the Standard Theory’s statement of the PSR for VPs, there can be an optional subordinate clause (S’) in place of the optional NP. (This choice is represented by the curly brackets around NP/S’ in the PSR for VPs.) For example, let us look at a simpler version of (4a):

(4b) The girl always adopted crafty cats with neat whiskers from Ofra’s shelter.

We can replace the NP “crafty cats with neat whiskers” with an S’ as in (4c) (making the requisite change of the verb “adopted” to one that works with subordinate clauses too, i.e. “said”):

(4c) The girl always said that the other girl adopted crafty cats with neat whiskers from Ofra’s shelter.

Here, the entire constituent “that the other girl adopted crafty cats with neat whiskers from Ofra’s shelter” is a subordinate clause. So, unlike (4b), in which the VP’s structure was $[[_{AP} \text{always}] [_{V} \text{adopted}] [_{NP} \text{crafty cats with neat whiskers}] [_{PP} \text{from Ofra’s shelter}]]$, the structure of the VP in (4c) is $[[_{AP} \text{always}] [_{V} \text{said}] [_{S'} \text{that the other girl adopted crafty cats with neat whiskers from Ofra’s shelter}]]$. In order to account for this possibility, that of subordinate clauses within VPs, we need PSRs for clauses too. The Standard Theory proposes two such rules in English, one for main clauses (S) and one for subordinate clauses (S’):

$$S \rightarrow \{NP/S'\} (T) VP$$
$$S' \rightarrow C S$$

We have already discussed how a clause is a simple sentence with a ‘subject’ and a ‘predicate’ (in the non-technical, non-theta grid sense of the terms). In grammatical terms, this translates into a sentence/clause being made up minimally of an NP (the ‘subject’) and a VP (the ‘predicate’). The PSR for S captures this very idea, since it states that an S is formed by combining an NP (or alternatively an S’ – just as in happens in the PSR for VPs) with a VP. However, a sentence/clause can have an optional

auxiliary or modal verb (that marks, among other things, the tense of the sentence) and which is denoted by T. The two following sentences illustrate this, with the T in italics:

(5a) The girl *will* adopt crafty cats with neat whiskers from Ofra's shelter.

(5b) That the girl adopts crafty cats with neat whiskers *will* make Ofra happy.

(5a) has an NP in the 'subject' position. (5b) presents the other choice given by the PSR for S, in which the subordinate clause "that the girl adopts crafty cats with neat whiskers" is present in the subject position instead, embedded in the larger (main) clause that is (5b). This subordinate clause omits the optional T, which is present in (5c):

(5c) That the girl *will* adopt crafty cats with neat whiskers *will* make Ofra happy.

So, adding the optional (T) constituent to the rule gives us the PSR for main clauses. The PSR for subordinate clauses is now easily derived because all we have to do to create an S' is to add a *complementizer* (such as the word "that") to the beginning of a main clause. So:

Main clause: The girl will adopt crafty cats with neat whiskers.

Subordinate clause: *That* the girl will adopt crafty cats with neat whiskers.

This gives us $S' \rightarrow C S$

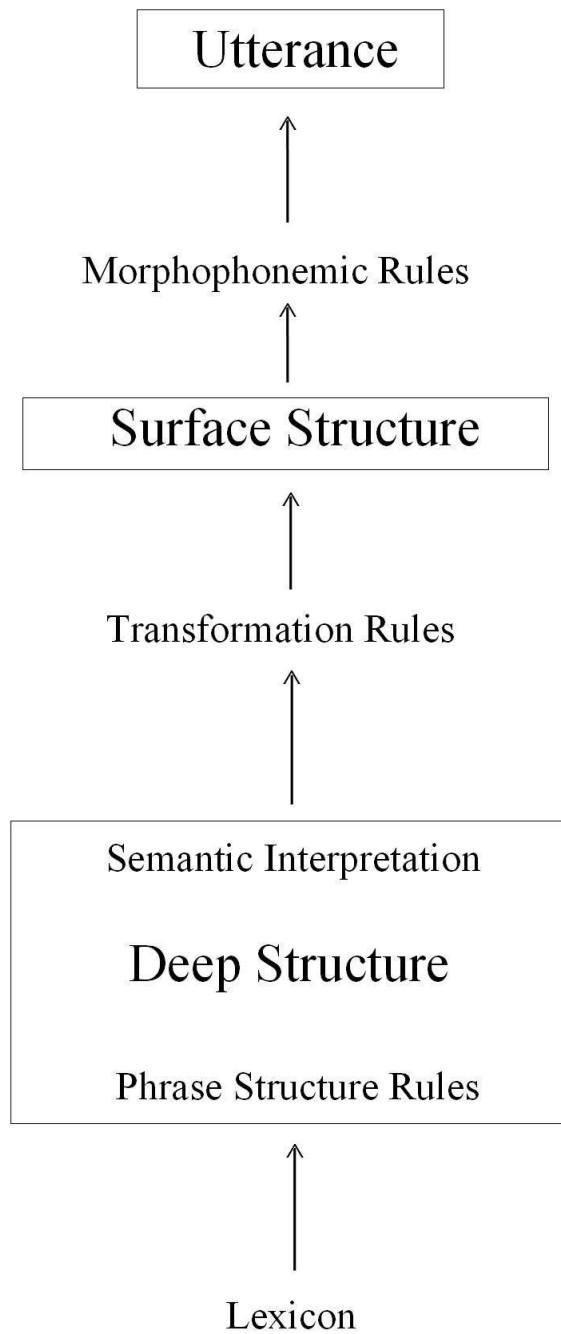
The above PSRs are what the Standard Theory proposes to generate the phrases and sentences of language. But language contains many structures that are not just simple phrases and sentences, which arise from transformation operations, such as the wh-movement transformation we explored in the last chapter. So, PSRs are not enough to generate all the grammatical structures of language. Therefore, the Standard Theory proposes certain rules for transformation as well, which allow more structures to be generated through transformation operations like wh-movement. (This is why the Chomskyan project in linguistics has always been one of *transformational* as opposed to *phrase structure* grammar.) To understand how such rules might be added to the PSRs to generate more grammatical structures, consider

Example 1.2-15, which gives a sketch of the computational system of language as proposed by the Standard Theory. At the bottom of this sketch we see the first step of grammatical generation in which words are taken from the lexicon and joined into phrases and sentences using PSRs, in the manner described in the preceding pages. This yields what are known as “deep structures” – abstract representations of phrases and sentences in the mind that are yet to be articulated by our vocal apparatuses, but which have completely undergone PSR-based computation.

Now, in the earliest versions of the Standard Theory deep structures were thought to have a certain semantic content, specifically of the kind shared by actual sentences that have the same meaning – like actives and passives (e.g. “the girl adopted the cat” and “the cat was adopted by the girl”, which have essentially the same meaning). Due to this semantic content, the active and passive forms of a sentence can be generated from the same deep structure – in fact, the very phrase “active and passive *forms of a sentence*” implies that there is an abstract sentence from which its active and passive forms can be derived, and it is this abstract sentence that the concept of “deep structure” captures. So, many linguists working within the Standard Theory hypothesized that the semantic content of a deep structure is analyzed once it has been generated by PSRs, to establish the kinds of semantically-related, actual sentences – like actives and passives – that might be generated from it. In other words, a deep structure receives a *semantic interpretation*, after which an actual sentence can be generated from it.²⁷ But this can

²⁷ Importantly, later versions of transformational grammar after the Standard Theory abandoned the idea that semantic interpretation happens in deep structure, and a more purely grammatical conception of this, more abstract, level of grammatical structure was eventually adopted – in fact, the very notion of “deep structure” in the *Aspects* sense has been abandoned by Minimalist linguistics, as mentioned in the last chapter. However, the idea that semantics plays a ‘deep’ role in sentence generation continued to play a role in the rival, anti-Chomskyan field of linguistics called “generative semantics” – and there are interesting historical parallels between the generative grammar vs generative semantics divide in linguistics on the one hand, and the pro-Schenker vs anti-Schenker divide in music theory on the other. In fact, much of the anti-Schenkerian rhetoric in music theory is a reaction to the perceived ‘hyper-formalism’ of Schenkerian theory, which is quite akin to the common anti-Chomskyan belief that generative grammar is too abstract and does not deal with the expressive, meaningful realities of language use in society. The latter view is quite common in field known as “cognitive linguistics” to which generative semantics was a precursor. And in line with the music/language parallel noted a few sentences ago, some of the ‘anti-formalist’ approaches to musical structure within music theory have been directly influenced by ideas from cognitive linguistics. The most notable example being Lawrence Zbikowski’s *Conceptualizing Music: Cognitive Structure*,

Example 1.2-15. An architecture of C_{HL} according to the Standard Theory



Theory, and Analysis (Zbikowski (2002)) – who I referred to in the last chapter as representing the “Chicago School” of anti-generative music theory.

happen only after the application of a *transformational rule* that determines the actual structure – the *surface* structure – of the sentence, i.e. whether it is active or passive, a question or its answer etc. After all, such surface structures related by meaning arise from transformation operations such as movement – a question and its answer are often related by wh-movement, as we saw in the last chapter, and an active and its passive are usually related by a transformation called “NP-movement”. So, after transformational rules are applied, surface forms of a sentence arise.

But these surface forms are still only grammatical representations of a sentence, albeit complete ones – they still need to be articulated by the vocal system. This is because the words that make up grammatical sentences are made up of abstract units themselves, called morphophonemes, that determine how a word should be pronounced – but which can be pronounced in different ways. So, which pronunciation applies to a morphophoneme in a particular word has to be determined first before a surface structure can be articulated. For instance the morphophoneme //z//, which often marks the plural ending of a noun in English, can be pronounced in three different ways, [s] as in cats, [z] as in frogs, and [ɪz] as in tortoises. Which of these three forms the morphophoneme will take depends on the previous morpheme to which it attaches, i.e. the morpheme at the end of the words “cat”, “frog” and “tortoise”. So, depending on these previous morphemes, the actual pronunciation of a word has to be determined, before the surface structure in which it appears can be articulated. This determination is done with the help of *morphophonemic rules*, which describe the dependencies between morphemes and morphophonemes. So, it is only after morphophonemic rules are applied to surface structures, which are themselves formed from deep structure after the application of transformational rules, is a sentence actually pronounceable.

The above Standard Theory of transformational grammar gave linguistics a very robust way of dealing with the structure of language. But it had some deficiencies too, which led generative linguists to improve upon it. For example, the PSRs we have seen so far work well for English but not necessarily for other languages. Take the rule for VPs, which states that an optional NP or S' can follow the head verb of the phrase. In a language like Turkish though, such an NP, and specifically the direct object in it (the head

noun) appears *before* the head verb of the VP, as in “Hasan kitab-i oku-du”, which rendered into English reads “Hasan the book read” (Carnie (2002): 129).²⁸ One solution to this problem is to postulate a different set of PSRs for every language. This is obviously an inelegant and cumbersome solution for a linguistic theory that aims to be scientific and universal. A much better solution would be a simple set of rules that applies across languages – and the attempt to find such increasingly simple and elegant ways of describing linguistic structure has therefore been a perennial goal in generative linguistics, which can be seen in the very name of the *Minimalist Program*.

Moreover, all of the four PSRs we discussed above have different makeups. They all have a head word, but have optional constituents that are different in kind and number – so a PP in the Standard Theory is postulated as having one optional constituent (an NP), whereas an NP itself is postulated as having three optional constituents (a D, one or more APs, and one or more PPs). So, there is a certain lack of elegance or simplicity in the formulation of the PSRs even *within* the single language of English. This really results, as we shall see in a moment, from the Standard Theory’s inability to capture some structural details within the four phrase types, which occur consistently across the four phrase types as well. In other words, a system of rules might be formulated that captures these similarities across phrase types, while also accounting for the structural intricacies within these phrase types in a manner that the four PSRs of the Standard Theory cannot. This quest for greater simplicity, elegance and (both inter- and intra-linguistic) explanatory power led to the formulation of the Extended and Revised Standard Theories developed in the later 1960s and 1970s. The main technical innovation that characterized these improvements was *X’ theory* (or *X-bar theory*), which was proposed by Noam Chomsky in 1970, and developed further by Ray Jackendoff (Chomsky (1970), Jackendoff (1977)).

To understand the essentials of X-bar theory let us consider the PSR for VPs again:

VP → (AP+) V ({NP/S’}) (PP+) (AP+)

²⁸ A similar example in Japanese was given in footnote 2 of the last chapter.

This rule presents what might be called a “flat structure”, i.e. all the constituents of the VP are on the same hierarchical level. This flat structure is evident if you examine Example 1.2-14 again. In the VP shown here, there are seven branches that join at the top VP node of the tree, viz. the first two optional APs, the non-optional head verb, the optional NP, the two optional PPs, and the final, optional AP – and they are all on the same hierarchical level, since they all join at the VP node and none of them is contained within another.

But what this PSR for VPs misses is that VPs do *not* have a flat structure – there are structural differences between how different constituents of the VP combine with the other constituents, and this leads to a more intricate, hierarchical organization of the constituents within the VP than suggested by the PSR. To see this, consider (4a) again:

(4a) The girl always quickly adopted nimble crafty cats with neat whiskers from Ofra’s shelter with her pocket money joyfully.

In the VP of (4a), there can be only one verb, viz. the head verb “adopted”, and only one NP (or alternatively, only one S’), but multiple APs and PPs. The PSR for VPs does account for this fact. However, it does not account for the fact that the PPs can be in any order; if we switch the order of the PPs in (4a), the sentence is still grammatically acceptable:

(6a) The girl always quickly adopted nimble crafty cats with neat whiskers with her pocket money from Ofra’s shelter joyfully.

But, these PPs cannot be switched with the single NP in the phrase:

(6b) *The girl always quickly adopted with her pocket money from Ofra’s shelter nimble crafty cats with neat whiskers joyfully.

The point is that the NP “nimble crafty cats with neat whiskers” seems to belong together with the head verb “adopted”, as a unit, and so they resist being split up, as happens in (6b). In contrast, the two PPs do

not seem to have that kind of connection to the head verb; they do not need to be next to the head verb, and can be switched around. Finally, the APs in the sentence seem to have a special connection to the head verb, like the NP, since they seem to be specifying something about it (in the way adverbs do anyway) – but they do not need to be next to the head verb in the way the NP does, since APs in VPs often appear at the end of the VP – as indeed the AP “joyfully” does in (4a) and (6a-b).

The fact that the [V ({NP/S'})] part of the VP acts a unit separate from the other constituents in the VP receives further evidence from a phenomenon known as “do-so replacement”. In the following sentence, the phrase “do so” (or rather its variant “did so”) targets only the [V ({NP/S'})] part of the VP (as shown by the corresponding italics), while excluding the other constituents of the VP – which demonstrates how [V ({NP/S'})] acts as an independent unit within the VP:

(6c) The girl always *adopted crafty cats with neat whiskers* from Ofra’s shelter but her brother only *did so* from the shelter run by Salim.

The PSR for VPs cannot generate these smaller, independent units within the VP, since it generates the entire, flat, structure of the VP in one fell swoop. X-bar theory provides a solution to this by describing phrase structure in a slightly different way (to account for these more detailed aspects of phrase structure), and by also providing a more intricate set of rules to generate this detailed structure. Rather than describing phrase structure in a flat way, it suggests that a phrase can be thought of in terms of three hierarchical levels of structure, made up of the smaller constituents of the phrase. I suggested earlier that this hierarchical microstructure exists consistently across the four main phrase types, and we shall see some evidence for this in a moment too. Therefore, rather than describing VPs, NPs, APs, and PPs independently, as the PSRs do, X-bar theory provides a unified description of the hierarchical microstructure of all the four main phrase types, represented by the general phrase XP. This is generated from a head word X (which stands for V, N, A, or P), and which contains intermediate structures called X', from which the theory gets its name, and which account for the three-leveled hierarchical

microstructure of XPs. The three levels of XP structure (and therefore all the levels of structure in all VPs, NPs, APs, and PPs) can be generated consequently by the three following, general X-bar rules:

$X' \rightarrow X$ (WP) (complement rule)

$X' \rightarrow X'$ (ZP) (adjunct rule)

$XP \rightarrow X'$ (YP) (specifier rule)

The first rule generates the basic unit of structure described in (6a-c) above, formed by the head of a phrase X and the other, optional constituent (WP) that seems to go together with it (such as the NP or S' in the case of a VP). This optional constituent (WP) is called the *complement*, so this rule that generates the [X (WP)] “head + complement” unit can be called the “complement rule”. Also, the [X (WP)] structure in the case of the VP discussed above would therefore be [V (NP)] (i.e. “adopted crafty cats with neat whiskers”), and the rule that generates it $V' \rightarrow V$ (NP) (or more generally, $V' \rightarrow V$ ({NP/ S' }), since there is a choice between an NP and an S' for the complement constituent, as we have seen earlier).

As we have also seen, the general structure that results from the application of the complement rule, i.e. X' (and V' in its specific application to VPs) forms a smaller constituent within the larger structure that is the whole phrase. Another way of saying this is that X' is an intermediate level in the three-level hierarchy of a phrase. In X-bar theory jargon, it is specifically called an intermediate level of *projection* or a *bar-level projection*. To understand why it is called this, consider that grammatical structures are made up of lexical items, so the information contained in the lexicon is represented in all of these structures. In other words lexical information *projects* upwards from the smallest structures to the largest ones. This is what “bar-level *projection*” means – i.e. the level of structure up to which information from the lexicon has projected. At a bar-level projection, lexical information has only projected to these bar-level structures, not to the whole phrase. From the discussion on Merge in the last chapter, you might remember that when two lexical items are combined, only the information from the hierarchically-superior item (i.e. the head of the resulting phrase) projects. In the case of X' level projections, only information from the head constituent X projects, which is why it is labelled an *X-bar*

and not a *WP*-bar after the complement (this is a phenomenon called *endocentricity*). So, in the case of VPs, V is the head of the resulting *V'* structure, and which is why it is labelled a *V*-bar and not an *NP*-bar or *S'*-bar.²⁹ Therefore, we get a *V*-bar level projection when a verb is combined with an optional *NP* or *S'*, because verb-related information projects upward in the generation of this structure, and ultimately in the generation of the complete *VP* that arises out of this.

With the complement rule we have been able to account for part of the microstructure of a phrase – i.e. the fact that the head of a phrase and its complement have a special connection, something the flat structure proposed by the *PSRs* was not able to account for. But now we have to account for the other parts of the microstructure of a phrase. So, *X*-bar theory now proposes another intermediate level of structure, i.e. another *X'* level of projection, in which the first *X'* is combined with one type of optional phrasal constituent in *XP* that still remains to be added to the structure, represented by (*ZP*) in the second *X*-bar rule above. These constituents are called *adjuncts*, so the rule that generates this second *X'* level of projection can be called the “adjunct rule”. In the case of VPs, the remaining optional PPs (such as “with her pocket money” and “from Ofra’s shelter”) are the adjuncts, and so they give rise to the adjunct level of structure in VPs, formalized by $V' \rightarrow V' (PP)$. That is, these PPs are combined, one at a time, with the *V'* structure generated earlier (i.e. “adopted crafty cats with neat whiskers”) to generate more complex *V'* structures like “adopted crafty cats with neat whiskers with her pocket money”, and ultimately “adopted crafty cats with neat whiskers with her pocket money from Ofra’s shelter”. (Remember, the verb is still the head of this second *V'* projection, not the PP (it is still a *V*-bar structure, not a *PP*-structure). So, *V* still dominates in the hierarchy over the PP – and it is *not* the case that the PPs of the adjunct level now dominate *V* or its complement *NP*, just because they constitute the second level of *X*-bar structure being proposed here.)

²⁹ In fact, we could never have something like an *NP*-bar or *S'*-bar structure because *NPs* and *S*'s are both optional in VPs and phrasal/clausal in nature. The head of a phrase is always non-optional, and never a phrase itself, as we saw earlier in this section. So, an *NP* or a *S'* can never be the head of a phrase – and they can never project a bar-level structure. Phrasal/clausal constituents like *NP* or *S'* can only be complements, as we have seen, or adjuncts or specifiers as we will see in a moment.

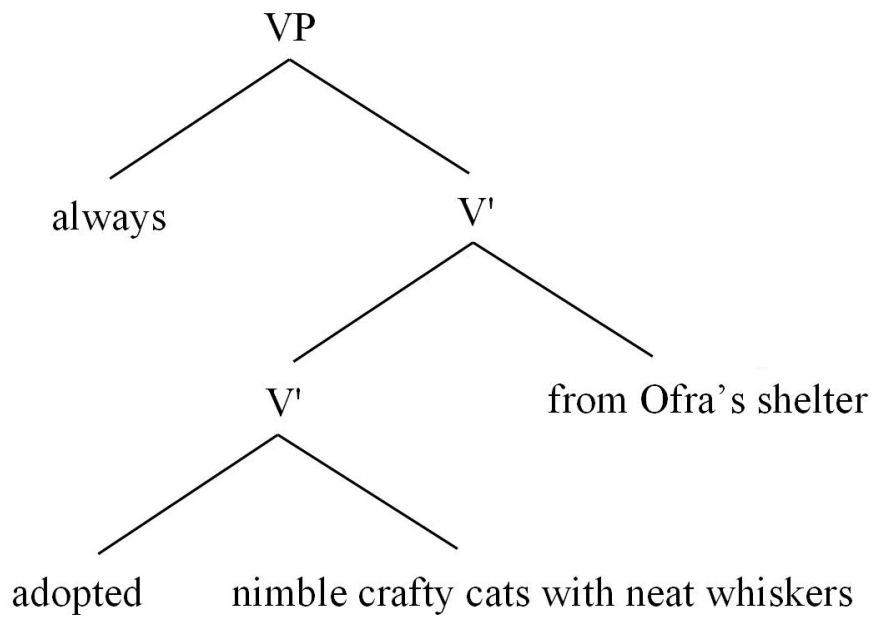
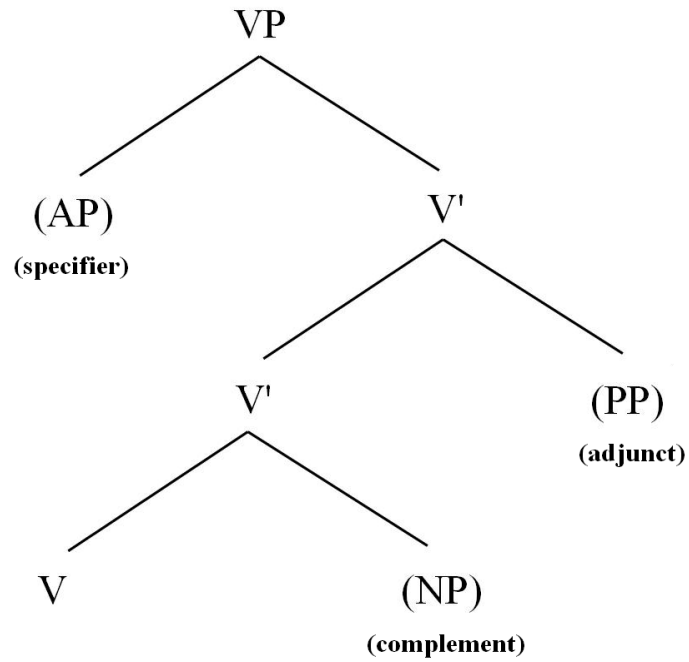
Finally this second X' combines with the last remaining optional constituent, called a “specifier” in X-bar theory, to give us the third and final level of structure – the complete XP – and this is described by the third and final X-bar rule, in which the specifier is represented by (YP). In the case of VPs, this “specifier rule” generates the complete VP from the second V' and its specifier – which is usually an optional AP (Fromkin, Rodman and Hyams (2009): 146), e.g. “always” in the sentence in (4a) we have been considering so far, and which is the last remaining constituent from the PSR for VPs yet to be added to the structure. So, the specifier rule for VPs looks like $VP \rightarrow V' (AP)$, or alternatively $VP \rightarrow (AP) V'$, since APs in a VP can both precede or follow the head verb, as we have already seen in the PSR for VPs. (Also, note that V still remains the head of the entire VP after the specifier rule has applied, since the optional, phrasal AP specifier cannot be the head of a phrase.)

The top figure in Example 1.2-16 depicts this three-leveled structure of a complete VP as proposed by X-bar theory. The lower figure shows how this model can be used to analyze the structure of the VP in (6d) below, which is the part shown in brackets. For simplicity's sake, I have left only one PP and one AP in (6d) compared to the two PPs and three APs in (4a). However, all of these can be added back to the structure of the VP; we just need to add more intermediate V' levels to account for these additional phrases (although we cannot have more than one specifier in an XP, so if we add more APs to the phrase, they will not be as specifiers):

(6d) The girl [[_{SPEC} always] [adopted] [_{COMP} nimble crafty cats with neat whiskers] [_{ADJ} from Ofra's shelter]].

The above X-bar model of phrase structure has a great advantage over the Standard Theory, in that it presents one unified description of how all phrases are generated, that too in terms of a more detailed “head-complement-adjunct-specifier” model, compared to the Standard Theory's flat structure approach. However, this model would only be a real improvement over the Standard Theory if this is in fact how all phrases are really structured and generated. We have seen how complement, adjunct, and specifier

Example 1.2-16. VP structure according to X-bar theory



phrases in VPs seem to behave in different ways, which justifies the use of the “head-complement-adjunct-specifier” model to describe their structure, and the X-bar rules to generate them. But what about NPs, APs, and PPs?

Actually, NPs, APs, and PPs *do* seem to make distinctions between their specifier, complement and adjunct components – so the general X-bar model of phrase structure *does* apply to them too.³⁰ In other words, the goal of consistency and elegance in the description of phrase structure that drove the revision of the Standard Theory was not merely an ideological one – it was indeed needed to account for the actual consistency of phrase structures in language. In the interests of space, I will just discuss the evidence for X-bar theory’s applicability to NPs, but there is evidence that even APs and PPs can be described in terms of the “head-complement-adjunct-specifier” model (for example, see Carnie (2002): 112-114).

In NPs, there seems to be a microstructure that distinguishes complements from adjuncts, as X-bar theory illustrates. In (7a-c), the first preposition phrase “of soup” seems to attach to the head noun “bowl” as a unit, and resists displacement by the other preposition phrases “with the garnishings” and “from Nanook’s kitchen” when they are switched around. However, the other preposition phrases can be moved around, and do not seem to require adjacency to the head noun in the way “of soup” does – identical to how the complement NP “nimble crafty cats with neat whiskers” and the adjunct PPs “from Ofra’s shelter” and “with her pocket money” acted in the VPs above.³¹ Also, the determiner “the” in the following sentences seems to do the job of specifying the head noun (as in *the* specific bowl of soup that was tasty), just as the specifier AP “always” does in (6d):

(7a) The bowl of soup with the garnishings from Nanook’s kitchen was tasty.

(7b) The bowl of soup from Nanook’s kitchen with the garnishings was tasty.

³⁰ In fact, the generative linguist Richard Kayne has stated in his famous “Antisymmetry” proposal (Kayne (1994)) that “specifier-head-complement” is the basic word order of *all* phrase structures in *all* languages.

³¹ In fact, one relatively reliable rule-of-thumb that decides that “of soup” is a complement, whereas “from Ofra’s shelter” and “with her pocket money” are adjuncts is that, in English, complement PPs almost always have the preposition “of” as their head, whereas adjunct PPs have other prepositions like “from”, “with”, “at”, “to” etc. as their heads.

(7c) *The bowl from Nanook's kitchen of soup with the garnishings was tasty.

So the “head-complement-adjunct-specifier” structure seems to obtain for NPs too, which means that the flat structure ascribed to them by the PSR for NPs will not do. Also, given that this is exactly the kind of microstructure VPs have, any description of the microstructure of NPs should capture this consistency between NP and VP structure. In other words, the X-bar description of phrase structure makes sense for NPs. This specifically leads to three X-bar rules for NPs, similar to the ones we have seen earlier, and which involve an intermediate level of projection called, in this case, N’:

$N' \rightarrow N (PP)$ (complement rule)

$N' \rightarrow N' (PP)^{32}$ (adjunct rule)

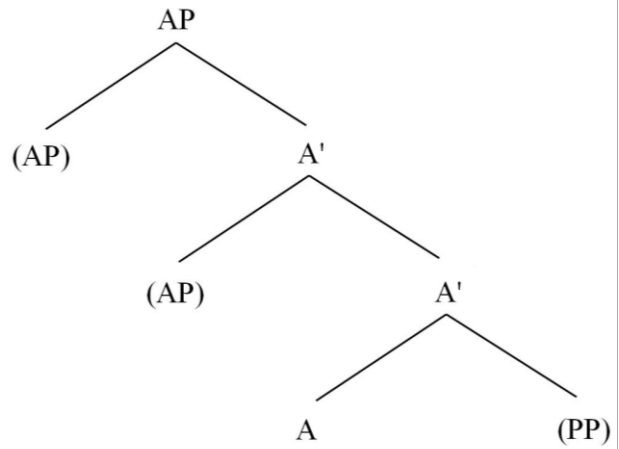
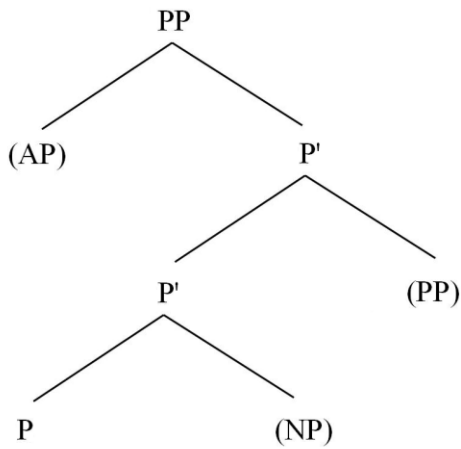
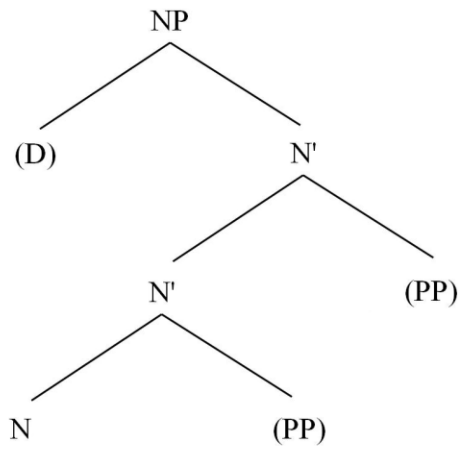
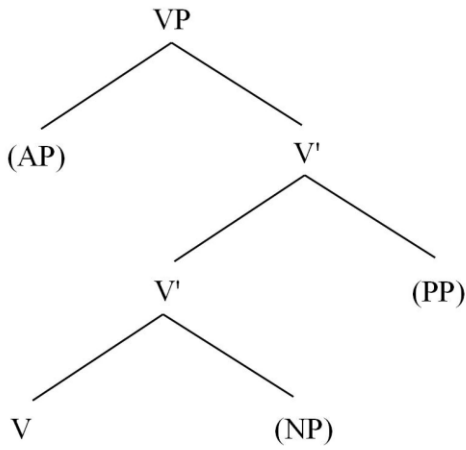
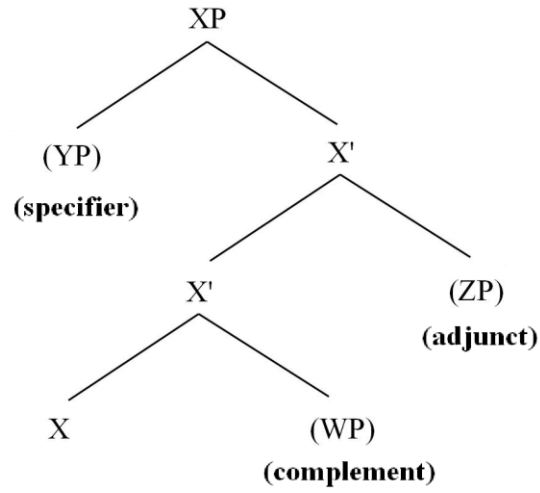
$NP \rightarrow (D) N'$ (specifier rule)

Here, the first rule describes how the complement PP “of soup” is added to the structure, and the second rule how the adjunct PPs “with the garnishings” and “from Nanook's kitchen” are added. The third rule generates the complete NP, by adding the (D) specifier to the previous N’.

So, X-bar theory seems to work for all phrase structures. Example 1.2-17 gives a summary of this conclusion; it shows the general X-bar model at the top, and how the model is realized in the specific VP, NP, PP, and AP structures lower down in the figure. If you compare the four X-bar descriptions of phrase structure in the lower part of the example with the PSRs for VPs, NPs, PPs, and APs discussed earlier, you will see that they essentially match up. For example, the X-bar model for VPs describes the exact same components of the VP that the PSR for VPs does – except that it does so through the much more detailed “head-complement-adjunct-specifier” model, which reveals the microstructure of VPs in a way

³² Given that NPs can have an optional AP constituent too, the second X-bar rule for NPs could also account for this AP rather than the optional PP, i.e. it could be stated as $N' \rightarrow (AP) N'$ as an alternative to $N' \rightarrow N' (PP)$. Both rules describe the adjunct level of NP generation, and can both be applied depending on whether the adjunct is a PP or an AP. If an NP contains both APs and PPs, multiple N’ levels will be required to add them all, one at a time, to the final NP phrase.

Example 1.2-17. X-bar theory: The general XP model, and its NP, VP, AP, and PP manifestations



the PSR for VPs cannot. The same applies to the X-bar description of NPs, as long as one keeps the caveat from footnote 32 regarding AP adjuncts in NPs in mind. The X-bar descriptions of PPs and APs is a bit more complicated than the ones given by the PSRs for PPs and APs, but only because it accounts for more complex PPs and APs such as the following (shown in brackets):

(8a) The artist was [_{PP} completely in love with his work].

(8b) The [_{AP} very fanatically serious about his work] artist is dead.

In (8a), the head of the PP is the preposition “in”, and its complement is the NP “love”. The PSR for PPs only accounts for this basic, complement-rule level structure in its formulation $PP \rightarrow P (NP)$. But it cannot account for a more complex structure that has an optional specifier, such as the AP “completely” in (8a), and an optional adjunct, like the PP “with his work” in (8a). The X-bar model for PPs in Example 1.2-17 can, however, account for these components of the PP too, which is why it describes a slightly different, but more complex, PP structure than the PSR for PPs does. Similarly, in (8b), the head of the AP is the adjective “serious”. According to the PSR for APs, $AP \rightarrow (AP) A$, we can generate an AP from “serious” by adding another AP to it, such as “very fanatically”. But adjectives can take a PP as a complement, such as the PP “about his work” in (8b). The PSR for APs does not account for this, but the X-bar model for APs at the bottom of Example 1.2-17 does, which is why it, again, describes a slightly different, but more complex, structure than the PSR for APs does. (Also notice how the X-bar description of APs in Example 1.2-17 is slightly different from the other XPs depicted, given the different shape of the AP tree. The difference in shape here owes to the fact that when an adjunct AP combines with an A’ level projection, it appears *before* the A’ (at least in English; which is why “fanatically” appears before “serious about his work” in (8b)). But adjuncts in other phrases, such as the adjunct PP “with his work” in (8a) appear *after* the bar-level projection – e.g. the P’ “in love” in the case of (8a). This leads to differently-shaped trees in APs versus other XPs.)

So, X-bar theory does represent a significant improvement over the Standard Theory in describing phrase structure. It also leaves us with some important conclusions about linguistic structure. First of all, it reveals the recursive nature of language again, given how X' level projections are often made up of other X' level projections. X-bar theory also shows us, very importantly, that linguistic structure is *binary-branching*. That is, by revealing the finer microstructure of phrases, X-bar theory shows that when constituents join to form larger X', and finally XP, structures, they always do so in twos. Look again at all the tree-diagrams in Example 1.2-17 – they are all made up of groups of *two* constituents whose *branches* join together to form a larger constituent, e.g. the head of a phrase and its complement that join to form an X' level projection, or the X' projection itself and the adjunct or specifier that joins with it to form either a larger X' projection or the complete phrase. This binary-branching character of linguistic structure will have some important consequences for our discussion of musical structure later in the chapter.

(Incidentally, paired, binary-branching constituents that join together to form a larger constituent, are called “sisters”, and the larger constituent they form, their “mother”. So, in a tree-diagrammatic representation of their structure, a head and its complement are branching sisters whose mother is X'. Such X' projections and adjuncts are sisters too, whereas a specifier is both a sister to an X' projection, and the *daughter* of XP. Finally, branching sisters join at the *node* of a tree diagram that represents their mother. In other words, mother nodes *dominate* their daughter nodes in a tree. X' projections form intermediate, *non-terminal* nodes in a tree, whereas the complete XP phrase forms the *root* node at the top of the tree,³³ and the smallest constituents at the bottom of the tree (normally individual words) form *terminal* nodes. With all of this in mind, we can now define a constituent more formally as “a set of nodes exhaustively dominated by a single node” (Carnie (2002): 72), which means that a constituent is a set of nodes all and only of which are dominated by a single node – in the way a V' dominates all and only its daughter nodes, the head verb and its complement NP/S', which therefore form a constituent as we have seen before.)

³³ Since linguistic trees are described in top-down fashion in tree diagrams, the top of the tree is where the assignment of structure *begins*. So, the top of the tree, somewhat paradoxically, becomes the ‘root’ of its structural description.

So far we have only examined X-bar theory's description of phrase structure. What about the structure of clauses in X-bar theory? We have seen two PSRs for clauses so far, but neither of them are in X-bar form:

$$S \rightarrow \{NP/S'\} (T) VP$$
$$S' \rightarrow C S$$

Can these rules be revised from an X-bar perspective? Let us try to answer this question with the rule for S first. We know from before that the head of a grammatical structure is both non-optional and non-phrasal/clausal. Do these criteria apply to any of the constituents of S? They cannot apply to NP and VP, which are phrasal, or S', which is clausal. This leaves us with T, which *is* usually a single word. And despite its placement within parentheses, it is actually non-optional too, although the T position is sometimes filled in an unusual way in sentences, as we will soon see.

In this light, the X-bar theoretic interpretation of S is as a TP (Tense Phrase), whose head is T. The VP of S then becomes the complement of T, both of which join together to form a T' level of projection. This T' then combines with the NP (or S') of S, now interpreted as a specifier, to give the complete TP (the tree-diagrammatic representation of this structure is given in Example 1.2-18a):

$$T' \rightarrow T VP$$
$$TP \rightarrow NP T'$$

Example 1.2-18b gives tree diagrams for the following sentences, the first with an explicit T (the auxiliary "will") and the second without one:

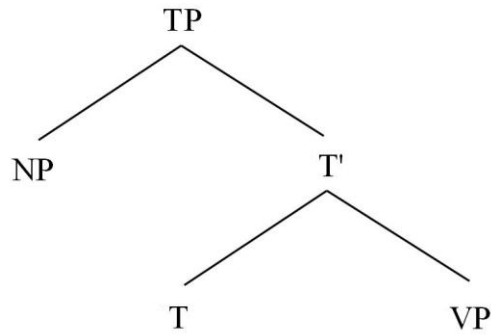
(9a) Uche will play the guitar.

(9b) Uche played the guitar.

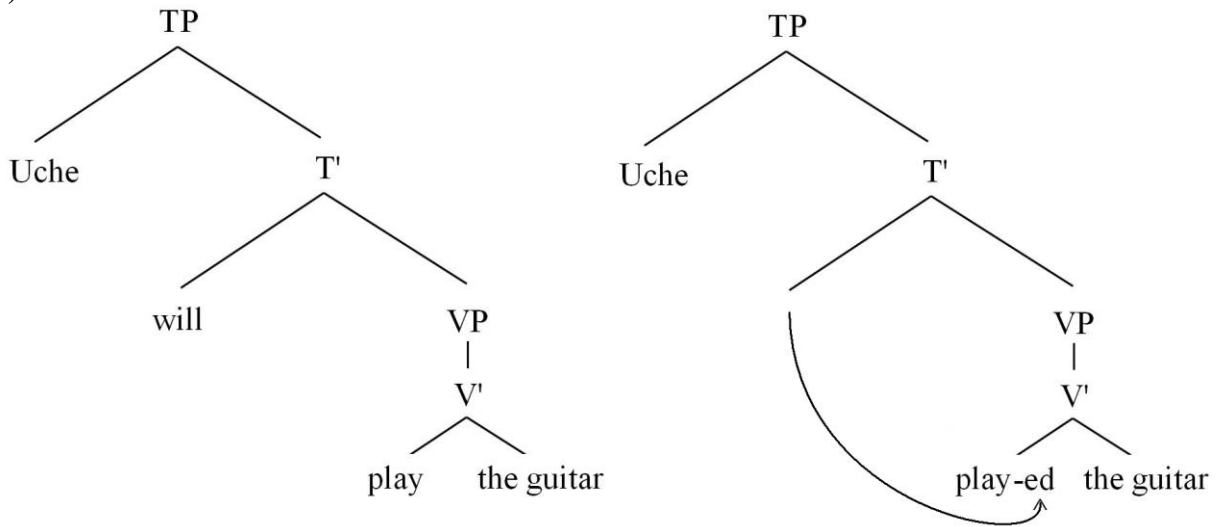
The structure of (9a) fits easily into the new TP interpretation of sentence structure, as the left image in Example 1.2-18b shows. But what happens when a sentence does not have an explicit auxiliary, as

Example 1.2-18. The X-bar sentence: (a) The general model, and (b) actual TPs with and without T

(a)



(b)



happens in (9b)? Well first of all, such sentences usually have a *tense inflection* on the verb – as in the “-ed” suffix after “play”, which renders this verb as being in the past tense. So, when a sentence does not have an explicit auxiliary, its main verb has tense inflection – but the reverse is also true, i.e. when verbs have tense inflection, the sentence will not have an explicit auxiliary:

(9c) *Uche will played the guitar.

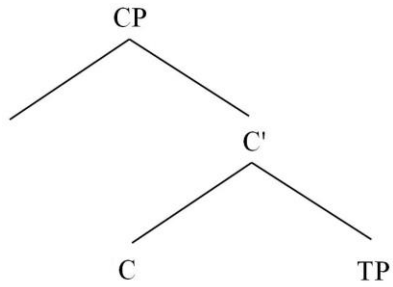
On the basis of this, generative linguists hypothesize that verb tense-inflectional affixes *are* a form of T, which could therefore appear in the T position in sentences, like auxiliaries. (In consequence, the T position is sometimes referred to as an I (i.e. “inflection”) position, and T’ and TP as I’ and IP respectively.) But unlike auxiliaries, tense-inflectional affixes cannot stand alone in a sentence – they need to attach to another word to be pronounced, i.e. a verb. So, when no explicit auxiliary appears in a sentence, a tense-inflectional affix might be considered to appear in its place – but this affix must *lower* to the main verb of the phrase and attach to it, so that it can be pronounced. This results in the grammatical phenomenon of *movement*, where a grammatical entity, in this case the T-affix, moves from one position in the sentence to another. (This particular phenomenon is called, unsurprisingly, “T-affix lowering”.) The T-affix lowering of the suffix -ed in (9b) is shown in its tree-diagrammatic representation in the right image of Example 1.2-18b. Note that this movement phenomenon involves the T-affix moving from one head position to another, i.e. from T to V, which is why it is considered an example of *Head-to-head Movement*. As a result, the T in a sentence is never optional, as appeared to be the case in the PSR for S. When an explicit T does not appear in a sentence, a tense-inflectional affix takes its place – which adds more evidence for T being the head of a sentence.

In a similar vein, we can consider the C complementizer in an S’ to be the head of that structure too, since it is not optional either, and it is also the only non-phrasal/clausal constituent in an S. In this light, X-bar theory interprets S’ as a CP (Complementizer Phrase), in which C is the head and S its complement. This leaves the CP with an empty specifier node, but this actually plays an important role in the previously discussed phenomenon of wh-movement, which we shall revisit again later in this section. Example 1.2-19a illustrates the tree-structure of the X-bar CP. The first image in Example 1.2-19b realizes this tree structure with an actual subordinate clause “that Uche will play the guitar”, which might precede “is well known” to form a sentence like:

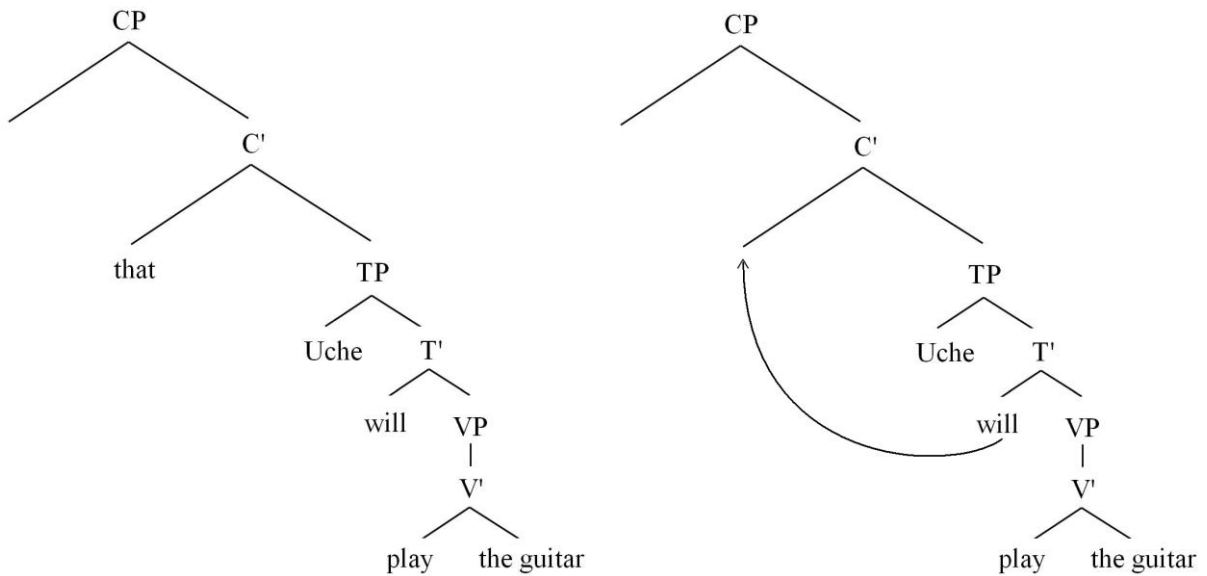
(10a) That Uche will play the guitar is well known.

Example 1.2-19. The X-bar S': (a) The general model, and (b) actual CPs with and without overt C

(a)



(b)



Notice that in both trees S has now been replaced by TP in the tree, given our above X-bar interpretation of S as TP, which is why the complement to the head complementizer “that” in (10a) is the sentence “Uche will play the guitar”. This also leads to the interesting problem of what happens in an ordinary sentence, a main clause, when it is not preceded by a complementizer. That is, does the new X-bar interpretation of CP not imply that in a main clause without a complementizer, we will essentially have a complement TP without its head C? (Which would be a violation of our grammatical principles suggested

so far, because the head is the non-optional constituent of a phrase, whereas its complement is always optional.)

To answer this conundrum, let us revisit the TP given in (9a) “Uche will play the guitar”. This is a typical main clause with no overt complementizer preceding it. However, think about the question that (9a) answers, i.e. “Will Uche play the guitar?”. This, and every, question, clearly has a connection to the sentence that answers it; presumably a semantic connection, since both question and its answer are ‘about’ the same phenomenon, in this case Uche’s playing of the guitar. As we have seen before, the introduction of movement phenomena into generative theory, like the phenomenon of T-affix lowering, was motivated by these very connections between sentences and their constituents in language – T-affix lowering was motivated by a *morphophonemic* connection between a T-affix and the verb it attaches to (so that the affix can be pronounced), and the connection between questions and their answers is what specifically motivates wh-movement, as we saw in the last chapter.

Under these considerations, generative linguists argue that the relation between (9a) and the question “Will Uche play the guitar?” is one of movement too. In particular, the head of (9a), viz. the auxiliary “will”, is said to move out of its T position into the C position occupied by the complementizer heads of CPs. This results in the word “will” now appearing in front of “Uche” in (9a) rather than after it, which yields the sentence/question “Will Uche play the guitar?”. This whole phenomenon is depicted in the second image of Example 1.2-19b. Also, this movement phenomenon is another example of head-to-head movement since it involves a word moving from the T head position to the C head position. This particular form of head-to-head movement is called “T raising”, since it involves raising the T constituent to the C position higher up in the tree. (It is also called “subject/aux inversion”, since it results in the subject “Uche” and the auxiliary “will” exchanging positions in the sentence.)

As a consequence of a movement phenomenon like T raising, generative linguists state that *all* clauses have complementizers, which is an idea that stems originally from the linguist Joan Bresnan’s doctoral dissertation (Bresnan (1972)). In subordinate clauses (CPs) the complementizer position is filled with the head complementizer of the CP as we have seen, whereas main clauses are said to have a *null*

complementizer, represented by the symbol \emptyset , which provides a spot for the T to raise to when transforming a main clause through a movement operation. With this stipulation, we now have a solution to the above problem, i.e. TPs do not seem to have overt complementizers, and so seem to be part of complementizer phrases that do not have heads. But this is not true, they *do* have complementizers, just null ones. This also means that a TP cannot be represented as an isolated entity in a tree diagram. It is always represented as occurring within a larger CP phrase, as in Example 1.2-19a, with the C position filled with the \emptyset symbol.³⁴

We have now accounted for S and S' in X-bar theory, under their new interpretations as CP and TP. This seems to account for all the PSRs, and thus represents a complete overhaul of the PSR system with the new, more elegant and detailed, X-bar system. However, there is one more revision that needs to be made, more for consistency than anything else. This has to do with the fact that in the theory we have been dealing with so far, heads are always non-phrasal/clausal and non-optional, whereas complements, adjuncts and specifiers are *always* phrasal/clausal and optional. If you revisit Examples 1.2-17-19, which depict all the XPs we have examined so far, you will see that this fact holds true of all of them – with one exception. This exception can be found in the NP tree depicted in Example 1.2-17. There, the specifier of NP is shown to be a determiner D. Now, this determiner is optional, as the parentheses show, but is often non-phrasal – seen best in the common use of the single word “the” in the D position. Which means that the specifiers of NPs, which *ex hypothesi* are supposed to be phrasal, are often non-phrasal – an inconsistency in the X-bar description of phrase structure.

To deal with this inconsistency, many generative linguists, following the work of Steven Abney (1987), argue that the D category in NPs is actually a complete phrase in itself, i.e. a DP, headed by single word determiners – which fixes the problem of the NP specifier often being non-phrasal. But for this fix

³⁴ The phenomenon of T raising is not seen in some languages like Irish. This is because these languages have complementizers (like the particle “ar” in Irish) that turn TPs into questions when added before them. Therefore, given that the complementizers in the TPs of these languages are actual and not null, T raising is neither needed to form questions, nor can it actually take place because there is no empty C position for T to raise to. In fact, raising T in these languages leads to ungrammatical sentences.

to really work, the entire NP has to be reconceived as a DP itself (or rather as a constituent within a larger structure that is the DP). To understand this, consider the left image in Example 1.2-20a. This is just a representation of the NP “the den” in the X-bar manner suggested earlier, although here the specifier of the phrase is depicted as a phrase itself, i.e. DP, to accord with our stipulation that non-heads must be phrasal or clausal. But look how this DP has just one constituent in it, the single word “the”, which is the determiner head of DP. It has no specifier phrase and no complement phrase, let alone an adjunct phrase. So, the only motivation for conceiving of this single D as a DP is so that it accords with our theory – the motivation is driven by theory rather than any data that suggests that DPs exist.

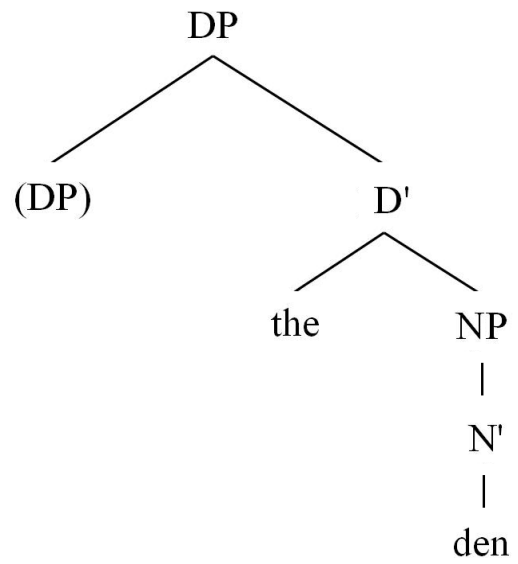
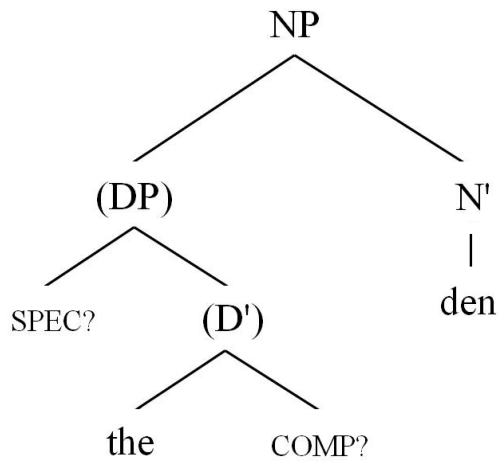
A better solution might be the right image of this example, where the NP is re-interpreted as being within a larger DP, whose head is the D. In other words D is no longer a component of NP, as the PSR for NPs stated; instead *NP* is now a component of *D*, or rather a DP headed by D. (So, we haven’t eliminated the NP category, we’ve just redefined it as being within a DP. In particular, NP is no longer the *mother* of D, with D as its specifier *daughter*, as it was in the Standard Theory – instead it is now its *sister*, both constituents now being immediately dominated by the mother node D’, and ultimately by DP.) This analysis is better because now the D does have a complement, the NP “den”, which therefore justifies the introduction of a DP category into the structure. So, re-interpreting the NP as being within a DP transforms D from being the problematic, non-phrasal, specifier of an NP into the unproblematic non-phrasal head of a DP, with the NP as its complement. This, in turn, fixes the inconsistency in X-bar theory’s description of NP structure relative to the other XPs.

Even though the prime motivation in reinterpreting NPs as being within DPs is to make X-bar theory consistent across XPs, there is data that suggests that this is the right view anyway, irrespective of whether it is consistent or not. For example, recall the constituent “Ofra’s” from (3f) earlier, which I labelled as a determiner at the time:

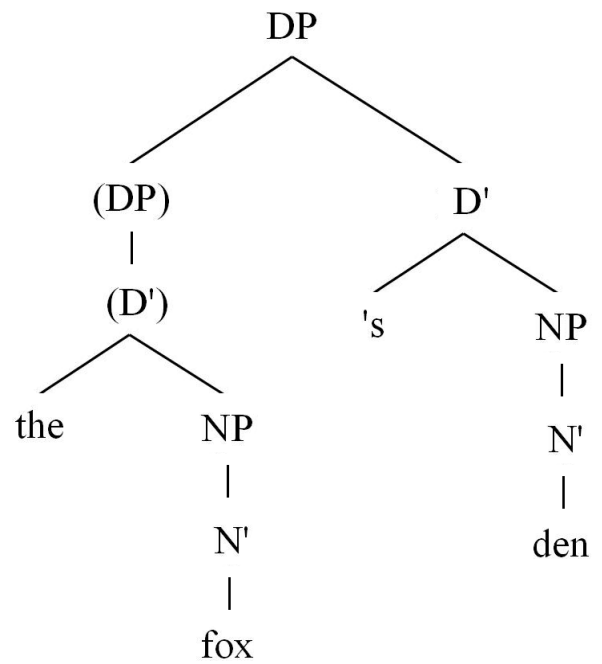
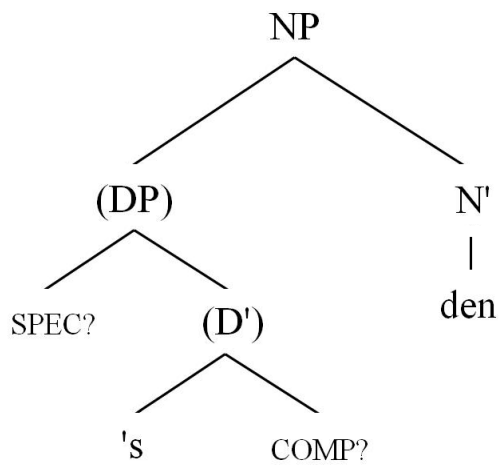
(3f) The nimble, crafty cat, with the neat whiskers, from Ofra’s shelter ran up the hill.

Example 1.2-20. The X-bar NP/DP: (a) with determiner, and (b) with construct genitive

(a)



(b)



The 's marker in "Ofra's" is known as a *construct-* or *s-*genitive, and it indicates possession of some kind, as genitives do (in this case, the fact that the shelter is possessed by Ofra). Because it indicates possession, the 's marker acts like a determiner – in (3f) it states that *the* shelter under discussion is *Ofra's* shelter. In fact, a sentence that has both an *s-*genitive and an actual determiner preceding a noun will be ungrammatical. This is because such an occurrence would be redundant; once the *s-*genitive 'determines' the noun there is no need for a determiner to do so as well, as can be seen in (11a-c). (Importantly, in 11b-c, only the 's marker is the determiner; "the fox" is a separate constituent that modifies the 's marker in some way.):

(11a) [This] den is small.

(11b) [The fox's] den is small.

(11c) *[The fox's] [this] den is small.

In the light of our preceding discussion about determiners and the NP/DP in X-bar theory, how do we describe the occurrence of 's markers in sentences? The left image of Example 1.2-20b attempts to describe it in terms of the earlier NP model, in which the 's marker is taken to be the head of a DP that specifies an NP, as its daughter. The problem here is not so much that the 's marker is the sole non-phrasal constituent of the DP, as was the case in Example 1.2-20a, because the constituent "the fox" should appear somewhere in this DP too. The problem is how we should analyze *this* constituent and where in the tree it should appear. This is because "the fox" is the same kind of structure as the larger structure under consideration in Example 1.2-20b ("the fox's den") – they are both NP/DPs. So, we could assign an analysis to the larger structure "the fox's den" by, say, placing "the fox" in the specifier position of the tree (as the daughter of DP and sister of D', since it seems to be specifying the 's marker in some way). But this only begs the question of how we should further analyze the smaller structure "the fox". After all, the "the" in "the fox" is a single word – so we could analyze this as a DP, but this would create the same problems that the first image in Example 1.2-20a did. And since there's no obvious solution to this problem within the traditional view of NPs, we're back to square one.

On the other hand, we could reconceive NPs as being within DPs as before. Under this view, shown in the right image of Example 1.2-20b, “the fox” just becomes another constituent within a larger DP. As a daughter to the larger phrase “the fox’s den”, which is also reconceived as a DP now, “the fox” now becomes the specifier of this larger DP, whose head is – no surprises here – the ’s marker. In other words, the DP interpretation of this whole structure fixes all the problems associated with its earlier interpretation as an NP. So, in sum, reconceiving the NP as a DP is not only justifiable on grounds of elegance and consistency within X-bar theory, it actually models the data better too.

But there is a problem with the DP proposal. Remember that in a sentence with an NP, the determiner that complements it is *optional* – but when a determiner does appear in a sentence, the NP that complements it is not optional. So, we can have sentences like (12a) and (12b), but not (12c):

(12a) Phurba ate the cake.

(12b) Phurba ate cake (for breakfast).

(12c) *Phurba ate the.

So, nouns seem to be more ‘basic’, more essential, for the structure of language than determiners, and the concept of a DP seems suspect in light of examples like (12c).³⁵ We will also see in chapter 2.2 how nouns and noun phrases have an important role beyond grammar too, e.g. in determining the rhythmic pronunciation of sentences. In this light, many linguists have continued to treat structures such as “the cake” or “the fox’s den” as NPs – and for the rest of the dissertation I shall follow this practice as well. But it is worth mentioning that a simple solution to the NP/DP problem exists – which is to not consider any of them to be essential categories in a grammatical theory to begin with. That is, instead of assuming the existence of NPs or DPs *a priori*, which would then force one to decide *a priori* what kind of

³⁵ Although part of the reason (12c) is ungrammatical is because the transitive verb “ate” does not get one of its two arguments – viz. the argument that a direct object like “cake” would have provided – and so it fails the Theta Criterion. So, the ungrammaticality of (12c) does not necessarily owe to the determiner not having its NP complement.

constituent “the fox’s den” is, why not allow NPs and DPs to emerge as by-products of grammar? In this view, all that one needs to assume *a priori* is the existence of a set of building blocks (i.e. the lexicon) from which grammatical structures are built, and some principles or operations with which to build these structures. No categories such as “NP” or “DP” need be assumed, so the question of deciding what kind of category “the fox’s den” is will not even arise. The Minimalist Program makes some overtures in this direction, which I will discuss later in this section.

This concludes our discussion of X-bar theory. But generative theory has continued to develop over the past 30 years, and the major grammatical concern during this time arguably has been to fix inadequacies that still persist within the X-bar description of language. X-bar theory seems to have two types of problems, both of which have to do with its failure to give an adequate account of all and only the grammatical sentences of a language, which, as we discussed in the last chapter, is one of the main goals of grammatical theory. X-bar theory does account for many grammatical phenomena in an elegant and consistent way, as we have just seen. But it also generates sentences that are *ungrammatical*, i.e. it *overgenerates*. Secondly, it *undergenerates* as well – it accounts for many, but not *all* the grammatical structures that we know languages to have.

To understand this latter point, recall our discussion of the poor job PSRs do within the Standard Theory, when it comes to describing phrase-structure differences across languages in an elegant way. X-bar theory fares better in this regard. For example, take the word order difference between English and Turkish we discussed then, in which we found that VPs are structured with the head verb before the complement in English, but after the complement in Turkish. Since X-bar theory models phrase structure in an abstract and general way, we can represent these two orderings very simply in X-bar terminology, where, as before, X is the head of a phrase, and WP its optional, phrasal, complement:

English: X (WP)

Turkish: (WP) X

This, in itself is a much simpler, elegant way of describing this word order difference than we could accomplish with PSRs. Given that PSRs are specific to parts of speech, cross-linguistic differences like the above would have to be modeled separately for VPs compared to NPs, APs, and PPs, and with a different set of PSRs in each case. In contrast, with the X-bar description above, all we have to say is that a universal grammatical principle exists in which all phrases in all languages must have a non-optional, non-phrasal head X whose branching sister is an optional, phrasal complement WP – but which sister comes first in a sentence varies from language to language. So, the universal principle is ‘parameterized’, with one parameter being that X comes first, the other that WP comes first – and that’s all. With this simple, two-line, “principles and parameters” description we can account for *all* word order differences involving head + complement pairs in *all* phrases in *all* languages.

This is no doubt a tremendous gain in descriptive power over the Standard Theory. But it has problems. Most importantly, it does not explain *how* and *why* certain phrases, and the sentences they make up, have the word orders they do in certain languages. Take, for example, one of the cases of word order variation we have already looked at, viz. T raising, in which T words like “will” raise to the C position to create a different word order (viz. that of a question) from the basic TP structure we started off with. But there is a reason, a morphophonemic one, for why this word order variation occurs – viz. that T has to raise to C for the \emptyset null complementizer *to be pronounced*, and for the resulting question to be articulated. So, just stating that different phrases and sentences in different languages have different word orders is not enough – the theory should say *how* these different orders arise (for example, through movement operations, as we have seen) and *why* they arise (for example, to meet certain morphophonemic conditions). In other words, we need a grammatical theory that explains how and why the above “principles and parameters” description of language actually occurs.

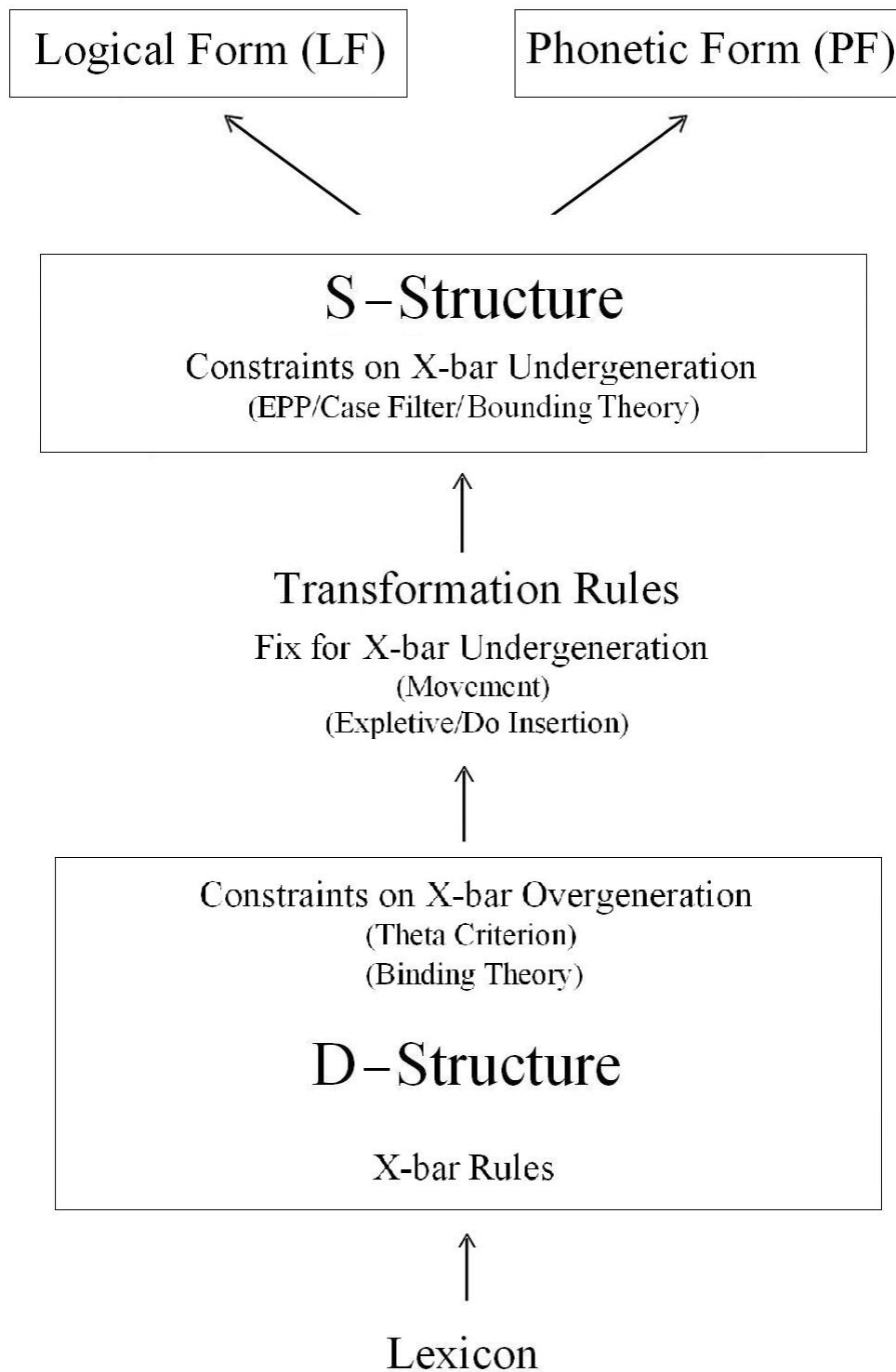
As I mentioned briefly in the last chapter, these have been the main goals of generative grammatical research in the past thirty years, and the (post X-bar) theoretical framework in which it has taken place is called, unsurprisingly, Principles and Parameters (P&P) theory. (See Baker (2001) for the most accessible

introduction to this framework.) In fact, the Minimalist Program operates within P&P theory too, which is why it does not propose an independent Minimalist *theory* of language. What Minimalism aims to do is create a research *program* (following the philosopher of science Imre Lakatos' definition of the term), i.e. a mode of inquiry with certain goals and ideals, but within the theoretical framework provided by P&P. Some of these goals, as we saw in the last chapter, have to do with giving the most economical description of C_{HL} possible, based on only the bare minimal concepts and properties necessary for grammatical theory to give such a description. That is, it aims to explain, in the most elegant way possible, why language is structured in a P&P way – an aim that has prompted it to shave off some of the excess baggage grammatical theory has acquired over the years, including many aspects of X-bar theory.

For the rest of this section, I will give an overview of some of the important technical developments in P&P theory, ending with a brief sketch of how recent Minimalist inquiries have tried to streamline this technical enterprise, to meet some of the above programmatic goals.

To start, let us revisit generative linguistics' proposed architecture of C_{HL} we saw in the last chapter. Example 1.2-21 provides this architecture, but altered significantly to account for the contributions to it from the Standard Theory and its X-bar revision, which we just explored – and with some additions from the toolkit of P&P theory. If you compare this architecture with the Standard Theory's earlier sketch of C_{HL} given in Example 1.2-15, you will notice that at the bottom of this architecture lexical items are now combined into larger structures with the help of X-bar rather than Phrase Structure rules, to account for this improvement over the PSR system. X-bar rules combine words into abstract, unarticulated syntactic structures as before, but these are now called D-structures rather than “deep” structures. This is to avoid the confusion we discussed earlier regarding the very notion of “deep structure”, which some linguists interpreted in semantic terms – specifically that the representation of a sentence at this abstract level of structure needs to be semantically interpreted before further generation can occur. As we will see in a minute, semantics does play a certain role at this level of structure even in

Example 1.2-21. An architecture of C_{HL} according to the P&P Theory



contemporary generative theory, especially in regards to the problem of X-bar theory overgenerating sentence structures. But this does not require a full semantic interpretation of sentences at this level, since generation at this level is almost pure syntax, with semantic interpretation appearing only after complete grammatical generation has taken place within the context of the level of structure called Logical Form.

So, D-structures and not deep structures are generated according to X-bar rules, but this leads to the aforementioned problem of overgeneration. For example, there is nothing to prevent the generation, within X-bar theory, of structures like (2b) and (2d), which we explored in the last section:

(2b) *Leila smiled the sandwich.

(2d) *Leila gave the sandwich.

As we noted in that section, such sentences violate the Theta Criterion, which is a semantic phenomenon that requires matching predicates with their arguments. X-bar theory needs such a semantic constraint to be applied to it, so that C_{HL} does not overgenerate by producing structures like (2b) and (2d). The Theta Criterion provides this constraint, and so becomes an intrinsic part of the generative process within the D-structure level of representation, where X-bar rules operate.

Another set of semantic criteria also help limit overgeneration, and these fall within what is known as *Binding Theory*, which is a proposal that lies at the core of an important early phase in the development of the P&P framework, called Government and Binding (or GB) Theory, suggested by Noam Chomsky in the early 1980s (in particular, in Chomsky (1981)). Just as the Theta Criterion constrains X-bar theory by targeting a specific type of constituent, viz. verbs, Binding Theory constrains X-bar theory's excesses by targeting noun phrases – specifically the NPs called *pronouns*, which includes *anaphors* (like “himself”) and *pronominals* (like “he”). (Note that the word “pronoun” is often used to refer specifically to pronominals.)

Unlike most other NPs, pronouns get their meaning from other words or phrases in a sentence, or even from the context in which the sentence is being uttered. In fact, anaphors *must* get their meaning from another NP earlier in the sentence called its *antecedent*. In (13a) below, “Xi” is the antecedent to the

anaphor “himself”, which can be illustrated by ascribing them the same subscript in the sentence, “i” in (13a) below – a process called “co-indexing”. In contrast, the “himself” in (13b) is not the person with the name “Xi”, as revealed by their different subscripts – so “Xi” is not the antecedent to “himself” in this sentence. In other words, “Xi” and “himself” are not co-indexed in (13b), and it turns out that this makes (13b) ungrammatical too:

(13a) Xi_i played the drums himself_i.

(13b) * Xi_i played the drums himself_j.

Now, recall that X-bar theory treats an NP as the specifier of a TP, and therefore the sister to the T’ level projection whose mother is TP. So, in (13a) the NP “Xi” is the specifier of the TP “Xi played the drums himself”, and the sister to the T’ projection “played the drums himself”, both of which are daughters of the TP. Note also that the T’ level projection has the T-affix “-ed” as its head, whose complement is the VP “play the drums himself”. In order for T to be pronounced, therefore, the “-ed” has to lower to the VP and attach to its head verb “play”, which is the movement phenomenon called “T-affix lowering” that we have explored before. (Note that we will have reason to revise this proposal of T-affix lowering in a bit.)

According to generative theory, the antecedent “Xi” *c-commands* the anaphor “himself”. A node “A” in a grammatical tree is said to c-command another node “B” in that tree if “every branching node dominating A also dominates B, and neither A nor B dominate the other” (Carnie (2002): 75). This basically means that a node c-commands its branching sister, and all the daughters, grand-daughters and so on, of its sister, because – by virtue of being sisters – they do not dominate each other, and there is one node (their mother) that dominates them both, and which is necessarily also a branching node because it has the two aforementioned sister nodes as its daughters. This means that the specifier in an X-bar tree c-commands the bar-level projection, since they are sisters, and every daughter, grand-daughter etc. of the bar-level projection as well. So, the TP’s NP specifier in (13a) “Xi” c-commands the bar-level projection T’ and all its daughters, grand-daughters etc. – which includes the anaphor “himself”.

What Binding Theory adds to this is the stipulation that node A *binds* node B (thus rendering a sentence grammatical) only if it *both* c-commands B and is co-indexed with it. The antecedent “Xi” c-commands the anaphor “himself” in both (13a) and (13b) – but they are co-indexed only in (13a), which is why only (13a) is grammatical. Now we see why (13c) below is grammatical, but (13d) and (13e) are not:

(13c) [Xi’s brother]_i played the drums himself_i.

(13d) *[Xi_i’s brother] played the drums himself_i.

(13e) *Himself played the drums.

In (13c), the NP “Xi’s brother” is the specifier of the sentence, thus c-commanding the anaphor “himself”, with which it is co-indexed as well. This binds the anaphor and renders the sentence grammatical. But in neither (13d) nor (13e) is the anaphor bound. In (13d), it is co-indexed with its antecedent “Xi”; but “Xi”, being the specifier of the NP “Xi’s brother” only c-commands “brother” – not the rest of the sentence – leaving the anaphor unbound. The anaphor “himself” in (13e) is not co-indexed with any NP at all, so it cannot be c-commanded by such an NP either. Therefore, the anaphor remains unbound, and the sentence ends up being ungrammatical.

Binding Theory adds another stipulation to how pronouns must be bound, in addition to the requirements that they be co-indexed with another NP that gives them their meaning, and which also c-commands them. This stipulation is that pronouns must be bound only in the correct *binding domain* for the resulting sentence to be grammatical. To understand this, notice that in (13c) the anaphor “himself” appears in the same, main, clause that its antecedent does. However, if we replace this anaphor with the *pronominal* “he”, which is co-indexed with and c-commanded by the same antecedent, the sentence is rendered ungrammatical, as shown in (13f) below. In contrast, if the anaphor of (13c) appears in a subordinate clause, rather than the main clause as it does in (13c), this renders the sentence ungrammatical too, as (13g) shows. Finally, if we replace the anaphor with the pronominal not in the main but in the subordinate clause, this time the sentence *is* grammatical, as shown by (13h):

(13f) *[Xi's brother]_i played the drums he_i.

(13g) *[Xi's brother]_i said that himself_i played the drums.

(13h) [Xi's brother]_i said that he_i played the drums.

What the above examples suggest is that for an anaphor to be correctly bound in a sentence, it has to be bound in the *same* clause as its antecedent, which also must c-command it and be co-indexed with it. So, if an anaphor appears in a different clause, e.g. a subordinate clause (as happens in (13g)), it remains unbound, even if it is correctly c-commanded and co-indexed with its antecedent. In contrast, a pronominal *cannot* be bound by an antecedent in the same clause, which is why (13f) is ungrammatical. It must appear in at least a different clause for an antecedent to bind it, as in (13h) – and often it does not even need to be bound at all for a sentence to be grammatical, e.g. in “He_i played the drums”, which is a perfectly legitimate sentence, in which “he” gets its meaning most likely from an NP in a previous sentence. In such cases, the unbound pronoun is said to be “free”, which is just another way of saying that it does not need to be bound.

From this we can conclude that (at least) the main clause is the binding *domain* of an anaphor, whereas the binding domain of a pronominal is at least a different clause from that of its antecedent, and pronouns are often free as well. Also, all other NPs are *always* free, and have no binding domain, such as the NP “Xi's brother” in (13i) below. This is why (13i) is ungrammatical – “Xi's brother” must be free, but is coindexed with, c-commanded by, and appears in a subordinate clause within the main clause that contains, the antecedent NP “he” (which is the same person as Xi's brother himself):

(13i) *He_i said that [Xi's brother]_i played the drums.³⁶

³⁶ Notice that although “he” c-commands “Xi's brother” here, “Xi's brother” does not c-command “he”. So, the pronominal “he” remains unbound, and therefore cannot be grammatical in the way (13h) is.

So, for a sentence to be grammatical, NPs appearing in it either have to be free from binding, or have to be correctly bound, which means that they have to be correctly c-commanded and co-indexed with the relevant antecedent, and also have to appear in the correct binding domain relative to this antecedent.

With these stipulations, Binding Theory adds an extra layer of (semantic) constraints, on top of the ones provided by the Theta Criterion, to the kinds of structures X-bar theory can generate. But now we must deal with the problem of X-bar theory undergenerating grammatical structures. We have examined how the main goal of generative linguistics has always been to provide an adequate description and explanation of how sentences are generated in all languages, but so far the vast majority of examples that we have discussed have been from English.³⁷ Now, generative linguists have dealt with increasingly greater numbers of languages over the years, and as I briefly mentioned earlier, P&P theory makes a specific prediction about how sentences are generated across these languages, viz. by means of grammatical principles that are universal (and hardwired into the minds of all humans), but which have particular parametric manifestations across languages. We saw an example of a universal grammatical principle in the last chapter in the *Extended Projection Principle*, which states that “all tensed clauses must have a subject”, and the parametric manifestation of this is where in a sentence this subject should appear, i.e. either before or after the main verb. (The linguist Lisa deMena Travis proposed the idea in (Travis (1984)) that word order is an example of parameterization of universal grammatical principles.)

And languages do vary in such a parametric way. If the subject can occur either before or after the main verb of a sentence, there are six possibilities that arise, Subject-Verb-Object, Subject-Object-Verb, Verb-Subject-Object, Object-Subject-Verb, Verb-Object-Subject, and Object-Verb-Subject. Of all the languages that have been studied (which number in the hundreds), these are exactly the six orderings

³⁷ A colleague once told me that examples such as the ones I have presented made her believe that generative linguistics was inherently biased against languages other than English. That this is not so is evident just from a cursory examination of the literature, where vast amounts of data from other languages are examined (and hotly debated, e.g. see Everett (2009) vs Nevins, Pesetsky and Rodrigues (2009a, 2009b)). In contrast, Schenkerian music theory, which I have been comparing to generative linguistics, might appear to be more explicitly biased towards one idiom, viz. Western Classical tonal music – not least so because of Schenker’s own prejudices about other musical idioms. But even Schenkerian music theory has been applied to much music outside the Western Classical canon, and the next chapter deals exclusively with this issue with respect to North Indian Classical music. So, the belief that generative theory, either musical or linguistic, is biased in favor of one idiom or another – the classical emic critique of such theories, as we saw in the last chapter – is inaccurate in my opinion.

that have been found (Tomlin (1986): 22), with the first two being by far the most common. (Although the existence of object-initial orderings is controversial, and restricted primarily to a handful of languages in the Amazon basin, and more than one ordering is frequently seen in a number of languages, often for different types of sentences.):

SVO (e.g. English, Russian, Mandarin)

SOV (e.g. Japanese, Turkish, Hindi)

VSO (e.g. Irish, Hebrew, Zapotec)

VOS (e.g. Malagasy, Fijian, Tzotzil)

OSV (e.g. Xavante, Apurinã, Warao)

OVS (e.g. Hixkaryana, Apalaí)

But can X-bar theory generate sentences with all such word order differences across languages? The answer is “no” – which is why X-bar theory *undergenerates* sentence structures too. Consider the following Irish (specifically Modern Irish Gaelic) sentence discussed by the linguist Andrew Carnie (Carnie (2002): 189):

(14) Phóg Máire an lucharachán.

In English, this sentence reads “Kissed Mary the leprechaun”. Irish is a language with verb-initial order, which explains the presence of “kissed” at the beginning of the sentence. As Carnie also says, this order is “the basic order of about 9 percent of the world’s languages ... [so] the failure of X-bar theory to account for 9 percent of the world’s languages is a significant one!” (Carnie (2002): 199-202).

Why does X-bar theory fail to account for a sentence such as (14)? Well, in our discussion on X-bar theory’s description of the verb phrase, verbs were understood as being the heads of VPs, whose complements are either NPs or subordinate clauses. The reason that these constituents are the complements to V is because they seem to belong together, as a unit – so separating them in a sentence leads to ungrammaticality. But in (14), the V “kissed” and the NP “the leprechaun” *are* separated, by the

NP “Mary”, which is the subject of the sentence – yet the sentence is grammatical, and this is true of any paradigmatic sentence in a language with VSO order. And X-bar theory is unable to account for this.

So, how can we fix this problem with X-bar theory? Let us look again at the sketch of C_{HL} according to P&P theory, given in Example 1.2-21. There you will notice that after X-bar theory’s overgeneration has been constrained by the Theta Criterion and Binding Theory, *transformation rules* are applied to fix X-bar theory’s undergeneration. Recall from Example 1.2-15 that such transformational rules were invoked even by the Standard Theory to account for movement operations that PSRs could not accomplish by themselves. X-bar theory cannot accomplish such operations either; it generates phrase structures with constituents in head, specifier, complement and adjunct positions, but has no provision for moving these constituents around, in order to realize such transformational operations as wh-movement or T raising. So, transformation rules have to be invoked to move constituents around.

Also, these rules are applied *to* D-structure, since D-structure is already generated by X-bar rules (and previously PSRs), after being appropriately constrained by the Theta Criterion and Binding Theory. Therefore, through the application of transformation rules to D-structure, word order ‘problems’, such as X-bar theory’s inability to generate (14) can be fixed, by moving the constituents of a sentence from positions that fit X-bar theory better to their observed S-structure forms. In other words, by applying transformational rules to D-structure, many more, actual sentences can be generated, viz. the S-structure in Example 1.2-21, what we previously called “surface” structure, and of which (14) is an example. As a result, not only can the different word orders within a language be generated, such as those of questions and answers, the different word orders of sentences across languages can be generated too. In this way, X-bar theory’s undergeneration can be fixed.³⁸

How movement transformations allow us to describe a structure such as (14) I shall describe shortly, along with some other important transformation phenomena. However, such a description must

³⁸ The fact that the order in which words appear in a sentence can vary even within a language, let alone between languages, makes word order’s utility as a descriptive tool rather limited – and the study of word order differences in different languages (i.e. word order *typology*) a problematic subject. It certainly applies more to the study of E-languages, rather than I-languages where word order properties are not conceptually necessary for grammatical theory, and can emerge as a by-product of more fundamental properties and operations, such as movement.

also include a discussion of the motivation behind such transformations. Recall how I described the P&P theory as a proposal that explains *how* and *why* sentences have the word orders they do in certain languages. As per the above discussion, movement transformations are how sentences end up having the specific word orders they do in a given language. So now the theory has to explain why these movements occur. Saying that they occur to fix problems in X-bar theory, so that C_{HL} can generate sentences that X-bar theory cannot by itself, is not a good enough explanation – in fact that would make movements an arbitrary theoretical device invented by linguists to fix a theoretical problem created by linguists, none of this necessarily having any basis in how languages really work. So, let us briefly look at the motivation behind why movement transformations are really invoked by P&P. After this I shall present some examples of actual transformations, including the one that yields the structure of (14), and I will also discuss how the motivation for invoking transformations is justified by what these transformations actually accomplish.

To do this, let us revisit Example 1.2-21 one last time. Notice there that two phenomena, viz. Expletive and Do insertion, are also listed as ways of fixing X-bar theory's undergeneration. The reason these insertion operations are also listed in the example is because movement transformations cannot always be invoked to help generate an S-structure that X-bar rules cannot generate by themselves, which is when insertion transformations come into play. But also notice that both movement and insertion transformations are merely *fixes* for X-bar theory's undergeneration. They are not independent constraints on what kinds of S-structures are allowable in a language, akin to the Binding and Theta criteria that govern what kinds of D-structures are permissible in a language, and which therefore help constrain X-bar theory's overgeneration, as we saw earlier. So, the real reason for invoking transformations is not to fix a theoretical problem within generative theory, but to satisfy some actual constraints on the kinds of S-structures permissible in a language. In the following discussion about the nature of and motivation for transformations, I will describe three such constraints, also listed in Example 1.2-21, viz. the Extended Projection Principle (or "EPP") we discussed earlier, and also the Case Filter, and the constraint on Bounding (not to be confused with the *Binding* constraint on pronouns). Like the Theta and Binding

constraints these three constraints are actual constraints on what kinds of structures are permissible in a language, although unlike Theta and Binding criteria they apply to S-structure – so they govern, for example, how the surface forms of questions and answers are generated from D-structure, through movements and insertions, after the basic D-structure has already been generated by X-bar rules, and constrained by Theta and Binding criteria. Therefore, the fact that movement and insertion transformations fix X-bar theory’s undergeneration of S-structures is essentially just a by-product of their meeting these constraints.

And these constraints are real facts about how languages work. For example, linguists, and not just those working in the generative tradition,³⁹ have long accepted that grammatical case is a feature seen across languages, and that for a sentence in any of these languages to be grammatical, it must be adequately case marked. This is basically all that the Case Filter constraint mentioned above does – it requires that an S-structure generated from D-structure be adequately case-marked or else the generative process will crash. All that movement transformations do is to move words around within a sentence so that they get case marked or meet some other, real, constraint on grammaticality, which the X-bar rules cannot meet by themselves. In the process many more, grammatical, S-structures can be generated by C_{HL} – and this fixes X-bar theory’s undergeneration, but only as a by-product of an independent constraint-satisfying transformational process. Insertion transformations act in a similar manner; they also exist to help satisfy actual constraints on the S-structures permissible in languages. For example, Expletive insertion specifically meets the EPP constraint. The EPP constraint requires that all tensed clauses have a subject.⁴⁰ Since information about subjecthood is stored in the lexicon, this information projects upwards

³⁹ For example, see Fillmore (1968). The linguist Charles Fillmore is known for his work in “Frame Semantics”, which is a proposal that falls within the broad, anti-Chomskyan linguistic tradition known as “cognitive linguistics”, other well-known proponents of which are George Lakoff and Paul Postal. (Although the cited article was written within the Standard Theory of generative linguistics.)

⁴⁰ Although this principle is subject to parametric variation, as we saw in the last chapter. So, in a language like English, the subject must be overtly pronounced, but in Italian the subject position is often occupied by an unpronounced, covert subject, variously referred to as a “null subject” or “pro” (Carnie (2002): 273-274). “Pro” is short for “pronoun”. Subjects are always NPs, as are pronouns, and the null subject specifically acts like a pronoun, since it gets its meaning from the broader context of the sentence – just like a pronoun, as we saw earlier in our discussion of Binding Theory. (This is because pro cannot express its meaning itself, being unpronounced.) The existence of covert subjects in Italian leads to a grammatical Italian sentence like “Parla” (“(He) speaks”), which

all the way to the level of S-structure in the generation of a grammatical sentence, which is why the EPP is the *extended projection* principle. Often transformation operations will move a word or phrase into the subject position, thus satisfying the EPP. But sometimes this is not possible, and that is when the operation of Expletive insertion kicks in, as we shall see shortly.

In the above manner, we can see how research within the P&P approach shows us both how and why sentences have the word orders they do in certain languages, viz. by proposing certain constraints that govern how C_{HL} generates sentences, which result in D-structures being transformed via various transformations to generate S-structures that meet these constraints – and which ultimately leads to grammatical sentences arriving at the final word orders they have in various languages. This sketch of the architecture of C_{HL} has much greater generative power than X-bar theory taken alone. But it is still rather inelegant, given its complicated set of constraints and transformation rules. As discussed before, the Minimalist Program has tried to streamline this complex architecture further, which I shall discuss as promised, at the end of this section. But let us spend a little time now looking at some actual transformations and the phenomena that constrain them. This will help us understand how sentences like (14) are actually generated, given X-bar theory’s failure to account for them, and it will also reveal some important properties of transformations – and this will actually shed some light on aspects of *musical* structure too, which I shall describe in the next section.

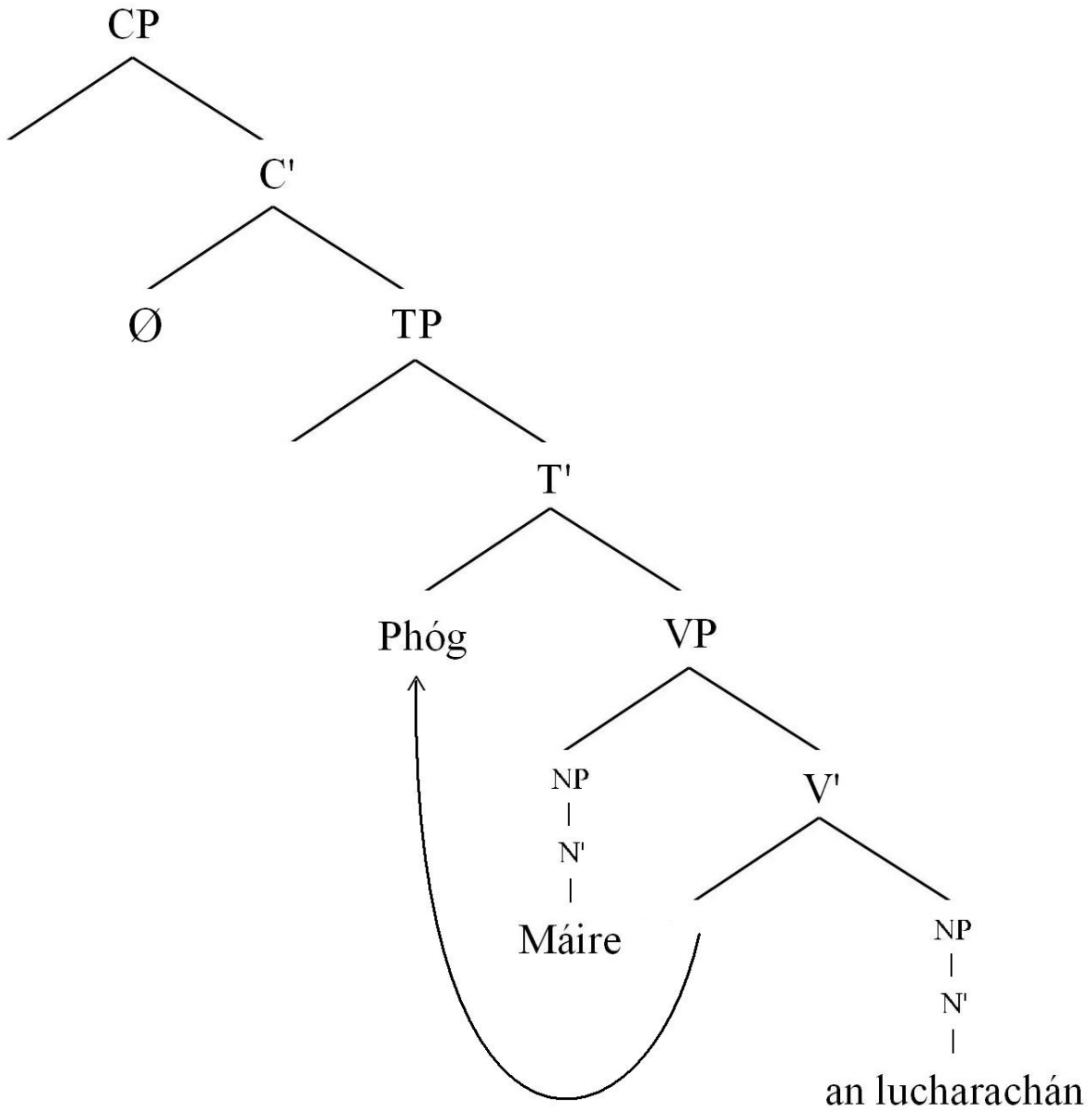
Recall that the problem with (14) was that the purported specifier of the sentence, the NP subject “Máire” was intervening between the head “phóg” and its complement NP “an lucharachán”, which X-bar theory does not allow. A particular view of sentence structure was proposed in the 1980s to deal with this and

seems to lack any subject at all (which would violate the EPP), but whose subject is really just covert. This is evidenced by similar Italian sentences in which an *overt* subject is sometimes used, such as “E parla” (Graffi (2001): 456), the position of the overt subject “E” in this sentence being replaced by “pro” in “Parla”. Such replacements cannot happen in a language like English, where “*speaks” (without a subject like “he”) is ungrammatical. So, the presence or absence of null subjects in languages is parameterized, by what is called the “pro-drop” or “null subject” parameter, which is ‘switched off’ in English (null subjects not being permissible in this language), but ‘on’ in Italian.

other related phenomena, called the “VP-internal subject hypothesis” (Zagona (1982), Kitagawa (1986), Koopman and Sportiche (1991)). According to this view, the NP subject of a TP such as (14) is not generated in the specifier position of TP, as we saw in Example 1.2-18a, but rather as the specifier of the VP that is the complement to the head T of this TP. (That is, it is generated in a VP-internal position.) The NP is then *moved* to its final S-structure position, in a transformation called, unsurprisingly, “NP movement”. This allows the S-structures of both sentences like (14) to be generated unproblematically, and those of paradigmatic English sentences to be generated by moving the NP to the specifier of TP position as in Example 1.2-18a – but only after the NP starts off in a VP-internal position.

To understand this better, consider Example 1.2-22, which displays the S-structure of (14). Here we see the subject of the sentence “Máire” inside the VP, in the position of specifier, rather than two branch-pairs higher as the specifier of TP, and sister of T’, as was the case in Example 1.2-18a. The main advantage of this VP-internal position of the subject is that it leaves many positions open above and to its left in the tree for the verb “phóg” to move to, which is exactly what we see happening in Example 1.2-22. Here the verb moves from its initial position as head of V’ and VP (where it was positioned in the D-structure of the sentence) to its S-structure position to the left of the NP “Máire” (as shown by the arrow) in the position designated for the head T of the TP. Such a move was not possible in the earlier model of sentence structure we discussed because there the subject NP “Máire” would have been generated directly in the specifier of TP position. This would leave only two positions to its left for the verb “phóg” to move to, but the verb cannot move to either of these positions, and this would prevent the S-structure of (14) from being generated. If you look at Example 1.2-22 again, you will see that the first of these two positions is that of the specifier of CP. I briefly mentioned earlier how this position plays an important role in wh-movement; basically this is the position that wh-words (or more accurately, wh-phrases) are moved to in wh-movement, i.e. to the front of a sentence (recall the concept of “wh-fronting” from the last chapter in this context), and this results in the generation of a question. So this position is reserved for wh-phrases, and so the verb “phóg” cannot move here. The only remaining position is to the right of this

Example 1.2-22. VSO order in Irish and the VP-Internal Subject Hypothesis



CP specifier, i.e. the position of the head of CP – but this position is already occupied by the null complementizer, as the example shows, meaning that the verb cannot move here either.

So, the VP-internal subject hypothesis gives us a way of actually generating the S-structure of (14) in a way the earlier, traditional X-bar model does not. And the VP-internal subject hypothesis has other advantages too, even though it remains a controversial proposal (e.g. see McCloskey (1997) for a

critical review). One advantage it has lies in the way it positions the NP subject of a sentence *local* to the verb within the VP – in fact the verb is the head of the V' projection that is the NP subject's branching sister. Now the locality of constituents has been an important point of focus in recent generative theory. Joining constituents that are close to each other is an efficient way of generating S-structures, and the Minimalist Program therefore prizes such phenomena, given its interest in describing the workings of C_{HL} efficiently. But the importance of local constituents in generative theory is not merely programmatic; constituents that are far away from each other often cannot be joined properly at all, and this often leads to ungrammaticality. (We saw an instance of this in the inability of pronouns to be bound outside their binding domain, which leads to ungrammaticality, and we shall see another example of this when we explore the, similarly named but different, Bounding constraint on S-structure generation.)

NP subjects and verbs need to be joined together, since the NP often provides the verb with an important argument – usually its agent, as we know from our discussion of the Theta Criterion. In the light of the preceding argument about locality, it would make sense then to stipulate that the NP subject and verb of a sentence be placed in positions local to each other. In fact, this locality condition is a requirement for grammaticality. Consider the English version of (14) in this regard:

(15a) Mary kissed the leprechaun.

In this sentence, the agent “Mary” appears in the same clause as the predicate “kissed”. However, placing these two constituents in separate clauses, i.e. in positions that are non-local to each other, leads to serious ungrammaticality:

(15b) *Mary said that kissed the leprechaun.

The VP-internal subject hypothesis, however, *does* place the agent in the same clause as its predicate; in fact, it places it in the same phrase as well, i.e. in the same VP. So, it obeys the locality condition on assigning theta roles. This, in consequence, gives the hypothesis a major advantage over its rivals.

Now, the ultimate test of such a hypothesis of course lies in whether it can actually generate acceptable S-structures, and the VP-internal subject hypothesis can do so only if we allow for some further movement transformations. So, for an actual S-structure like (14) to be generated we have to move the verb from its position within the VP in D-structure, as posited by the hypothesis, to the T position as Example 1.2-22 shows, and we have to allow a further NP movement of the subject NP from its VP internal position to the specifier of TP position to generate the S-structure of the paradigmatic English sentence. And as we discussed just a few pages ago, all such movements need independent justification – just invoking them to fix problems with how a theory generates S-structures will not do. I will return to the motivation for invoking NP movement for the subject NP, from its VP internal position, in a moment; but the motivation for moving the verb to T is quite straightforward. It is morphophonemic, just like the head-to-head movement of T-affix lowering we explored earlier. In that phenomenon an affix to a verb that expresses tense (such as “-ed”) was said to lower to the verb itself, in order for it to be pronounced. And the movement of raising the verb to T – a movement called, appropriately, “V raising” or just “raising” – is just the same head-to-head movement phenomenon, since by raising the verb head of VP to the T head of TP, the tense feature is joined with the verb and thus made pronounceable.⁴¹

So raising, in conjunction with the VP-internal subject hypothesis, helps explain how the structure of (14) arises. It also helps explain unusual structures in other languages, in which verbs also

⁴¹ This leaves us with the problem of why we need to propose two different transformations, i.e. T-affix lowering and V raising, if they are both doing the same thing – an inelegance that is at odds with the programmatic goals of generative theory. One solution to this is to say that V raising also occurs in languages like English rather than T-affix lowering, it is just that V raising here is *covert* (Carnie (2002): 321-323), i.e. it is unpronounced – or rather the verb raises to T after it has been pronounced (i.e. after its S-structure has been generated and mapped to the articulatory system through the Phonetic Form interface). This makes the matter of whether a language has overt or covert movement a parametric issue, similar to the parameterization of overt versus covert subjects, in languages like English versus Italian, which was discussed in the previous footnote. That such covert movement happens in languages is evident from the case of wh-movement in Mandarin. As we have discussed at length now, movement transformations are motivated by various constraints on S-structure, and wh-movement is constrained by Bounding Theory, which I shall discuss after our current discussion on V raising and NP movement. Now, the linguist Cheng-Teh James Huang (2006) has discussed how the fronting of wh-phrases does not seem to happen in Mandarin, which leads to fully grammatical questions such as “Ni kanjian-le shei?”, which reads in English as “You see who?” – giving the appearance that wh-movement does not take place in Mandarin. (Such a language, in which the wh-phrase remains in its original position at the end of a sentence, is called a “wh-in-situ” language.) But if wh-movement really does not take place in wh-in-situ languages, then Bounding violations would not happen in them either, since it is improper wh-movement that leads to such violations. But Huang argues that Bounding violations *can* happen in Mandarin too – thus proving that a transformation like wh-movement must occur in it as well.

raise to meet morphophonemic requirements. Consider the following sentences discussed by Andrew Carnie, the first from the West African language Vata (based on data from Koopman (1984)), and the second from French:

(16a) A li saka. (“We eat rice”) (Carnie (2002): 198)

(16b) Je mange souvent des pommes. (“I eat often (the) apples”) (Carnie (2002): 189)

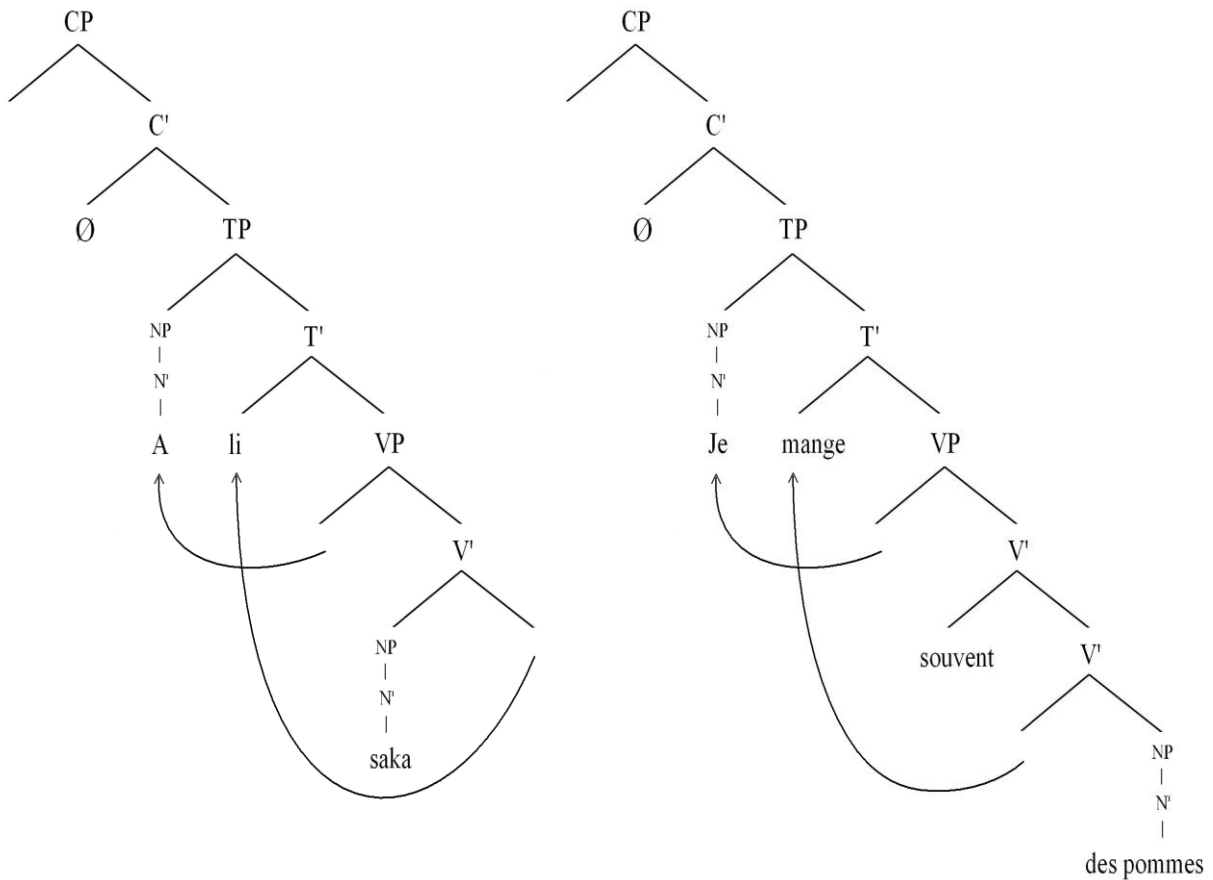
(16a) might seem good enough – but Vata is an SOV language like Japanese or Turkish, so the transitive verb “li” would normally appear at the end of the sentence, as in “A la saka li” (“We have rice eaten”). Therefore, (16a) would normally be as ungrammatical in Vata as “We have rice eaten” is in English.

However, notice that (16a) does not have an auxiliary between “A” and “saka” in the way “A la saka li” does with the auxiliary “la” (i.e. “have”). The explanation for (16a) therefore, is that when there is no auxiliary to fill the T position in the tree structure of a sentence, the verb raises to the T position – i.e. the movement of raising occurs – to ‘support’ T morphophonemically (i.e. it makes T pronounceable), in a way the auxiliary normally would have done. Example 1.2-23a illustrates this by showing how the verb “li” raises from its VP-internal position, as head of the VP and branching sister of its complement NP “saka”, to the T position, to make T pronounceable. Also, since Vata is an SOV language, unlike Irish, the subject NP “A” cannot remain in its VP-internal position, as it did in Irish, and as is stipulated by the VP-internal subject hypothesis. So, “A” raises too, through the transformation of NP movement, to its S-structure position as the specifier of TP.

Turning to (16b), which is a sentence in French – a language with SVO order like English – notice that the word “souvent” intervenes between the verb “mange” and its NP complement “des pommes”. This is not allowable in X-bar theory because “souvent” is an adjunct⁴² to the verb “mange”,

⁴² Andrew Carnie considers the AP “souvent” (i.e. “often”) to be an adjunct (Carnie (2002): 190), although in our discussion of X-bar VPs earlier, we treated words like “often” to be AP *specifiers* of VP, according to Fromkin, Rodman and Hyams (2009): 146. This earlier view of course does not work with the VP-internal subject hypothesis, since if VPs have AP specifiers, they cannot simultaneously have NP specifiers too, as the VP-internal subject

Example 1.2-23. Word order parameters and Head-to-head movement: (a) Vata (b) French



and therefore cannot intervene between it and its NP or S' complement. Recall the English sentence of (6b) in this regard, which is rendered ungrammatical by the intervention of the adjunct PPs “with her pocket money” and “from Ofra’s shelter” between the verb “adopted” and its NP complement “nimble crafty cats with neat whiskers”:

(6b) *The girl always quickly adopted with her pocket money from Ofra’s shelter nimble crafty cats with neat whiskers joyfully.

hypothesis requires. This is why Carnie, a supporter of the VP-internal subject hypothesis, treats APs as adjuncts rather than specifiers of VP. But in any case, this entire controversy is irrelevant to the above description of (16b) because neither a specifier nor an adjunct is allowed to intervene between a head and its complement in X-bar theory, so the problem with (16b) persists regardless of how one categorizes “souvent”.

But notice that the problem here is similar to that concerning (14). A constituent is intervening between two other constituents that cannot be separated according to X-bar theory (in the case of (14) the NP specifier of the VP “Máire” was intervening between the head verb “phóg” and its complement “an lucharachán”). So, the explanation for (16b), just as for (14) – and similar to (16a) – might involve raising. Example 1.2-23b shows how this is exactly the case. “Mange” is generated in its VP-internal position, as the head of the VP, but then raises to the T position, for the same morphophonemic reason suggested above – i.e. to make T pronounceable. In the process, “souvent” and “des pommes” are left adjacent to each other, thus generating the S-structure of (16b).

Notice again that the subject NP “Je” also raises – it undergoes NP movement, that is – from its VP-internal position, as stipulated by the VP-internal subject hypothesis, to its S-structure position as specifier of TP. We have already seen a couple of instances of NP movement in the preceding examples, so it is now time to explore the motivation behind invoking this type of movement transformation. The motivation for this transformation has much to do with the notion of locality of constituents, whose importance we have already discussed a little bit earlier, and which was also exemplified by the way NPs behave within C_{HL} . (Recall the importance of NPs like anaphors being close to (i.e. in the same binding domain as) their antecedents for them to be properly bound, according to Binding Theory. Also recall that NPs must appear in the same clause as a predicate, to provide the predicate with the right number of arguments required by the Theta Criterion – a condition whose fulfillment gives the VP-internal subject hypothesis an advantage over rival hypotheses of sentence structure too.)

Now, Binding and Theta criteria govern the behavior of NPs at the level of D-structure, but NP *movement* is governed by a different constraint at the level of S-structure, and one that also depends on NPs being close to another constituent. This constraint is the Case Filter. The Case Filter constraint is not the same as the Binding and Theta constraints, but they are all similar in that they both govern the generation of sentence structures, and both privilege locality relations between constituents. As a result, the Case Filter constraint was introduced at the same stage of development of generative theory as the Binding constraint, particularly in Noam Chomsky’s *Lectures on Government and Binding* (Chomsky

(1981)). But unlike the Binding and Theta constraints, the Case Filter does not deal with semantic information. Instead, it deals with information relating to grammatical case, which itself figures in grammatical relations such as those between subject and object. Consider (15a) again:

(15a) Mary kissed the leprechaun.

We know from our previous discussion of this sentence that “Mary” is the subject of the sentence, whereas “the leprechaun” is its direct object. To put things in a slightly oversimplified manner, subjects normally get nominative case, while direct objects get accusative case. This is particularly evident in a language like Japanese, where case is overtly pronounced via a nominative case affix “-ga” that is attached to subjects and an accusative affix “-o” that is attached to direct objects. However, in a language like English, case is not always overtly pronounced via such affixes; if we make “Mary” the object rather than the subject of (15a), the word remains unaltered morphologically, even though it now receives accusative case:

(15c) The leprechaun kissed Mary.

In light of this, Chomsky proposed in his theory of the Case Filter that all languages have case, except that in some languages, like English, it is covert. Therefore, all languages have, not case in the traditional, overt morphological, sense, but an abstract Case, which can either be overt or covert depending on the language. More importantly, all sentences in a language need to express this Case organization in the generation of S-structure or the structure will not map to the phonological systems (i.e. it will not be pronounceable) and the generative process will crash. The reason why Case has to be expressed for a correct mapping to phonology is because, as Chomsky says himself:

“In surface structure, verbal constructions differ from nominal and adjectival constructions in form. I assume that the reasons derive from Case theory. The crucial idea is that *every noun with a phonetic content must have Case* [my emphasis]... Assuming that Case is assigned to NPs by virtue of the configurations in which they appear and percolates to their heads, [the preceding, crucial idea] follows from the *Case Filter*, which I will assume to be a filter in the PF-component.” (Chomsky (1981): 49)

In other words, for an NP to be pronounced as such, i.e. for it to be recognized overtly as being different from verbal, adjectival and other constructions, its Case organization has to figure in the generation of its overt, pronounced form, since this is what distinguishes NPs grammatically from other constructions. And this Case organization can only figure in the generation of an NP if the NP is in the “correct configuration”, as Chomsky says. According to Case theory, this means that an NP that expresses nominative Case, i.e. a subject, must be local to another constituent that assigns this Case, specifically constituents that occupy the T position in a grammatical tree, e.g. auxiliary verbs. Similarly, NPs that express accusative Case, i.e. direct objects, must be local to transitive verbs, which are the constituents that assign *this* Case.

To understand why this locality requirement exists, recall that a constituent’s lexical structure contains a variety of information about the constituent, such as information about its Theta grid, its grammatical category etc., as we saw earlier – and also information about its Case organization. So, Case is a *lexical feature* of constituents. For two constituents to be joined together in the generation of a grammatical structure, these features have to *agree*, i.e. they have to be matched up or “checked”, so that the C_{HL} knows that it is okay to join them. Right at the beginning of this chapter I gave an example of this in the sentences “cats drink milk” and “a cat drinks milk”, where the plural feature “s” at the end of “cats” is checked with the plural feature in “drink”, to allow the generation of the first sentence, whereas the absence of this feature in “cat” allows it to agree with the *singular* form of “drink”, i.e. “drinks”, to generate the second sentence.⁴³

So, for Case to be properly expressed in a sentence, the Case feature of an NP must check with the Case feature of the constituent that assigns that type of Case – the hypothesis being that such feature-

⁴³ In more recent, Minimalist, approaches to the notion of feature checking, a distinction is made between “strong” and “weak” features (Chomsky (1995b): 197-198). Strong features must be checked during the generation of S-structure or else the computation will crash. In this light, Case features in English nouns are strong features that must be checked in S-structure generation, or else the structure will not map to phonology and the computation will crash. In contrast, weak features do not have to be checked during S-structure generation, e.g. tense features in English verbs or wh features in wh-in-situ languages like Mandarin (see footnote 41). This means that the presence of strong and weak features in the lexicon varies from language to language, i.e. which features are strong or weak in a language is parameterized. Of course weak features are still checked, albeit *after* the mapping to phonology has occurred, i.e. at LF rather than PF. In other words, only weak features can remain *uninterpreted* when the mapping to phonology occurs – which therefore leads to the phenomenon of covert (i.e. unpronounced) movement.

checking can only happen when the NP and the Case-assigning constituent are local to each other. Therefore, to receive nominative Case, a subject NP has to be local to an auxiliary verb, and to receive accusative Case a direct object has to be local to a transitive verb. If this constraint is satisfied, the generated structure will pass the Case Filter, and its phonetic content will become pronounceable – in other words, the generation of the sentence will be successful and an actual, pronounceable utterance will be realized.

The above Case Filter proposal achieves a number of things. First, it reveals how the ultimate role of the C_{HL} is to generate structures that meet external conditions of meaning and phonology, so that generated structures are meaningful and pronounceable. With the help of the Case Filter, C_{HL} generates structures that are pronounceable, which therefore maps the grammatical system to phonology successfully. The mapping of grammar to semantics and phonology gets a special emphasis in the Minimalist Program through its focus on the LF and PF levels of grammatical representation, so the Case Filter shows us how this important mapping might be achieved. Another thing the Case Filter achieves is that it reinforces the importance of locality of constituents, especially NPs, in the generation of sentence structures. This gives us another way of understanding why heads and complements cannot be separated in a sentence (at least at the level of D-structure, prior to movement transformations). Recall from sentence (15a) how the NP “the leprechaun” is the complement of the verb “kissed”:

(15a) [Mary]_{SPEC} [kissed]_{HEAD} [the leprechaun]_{COMP}.

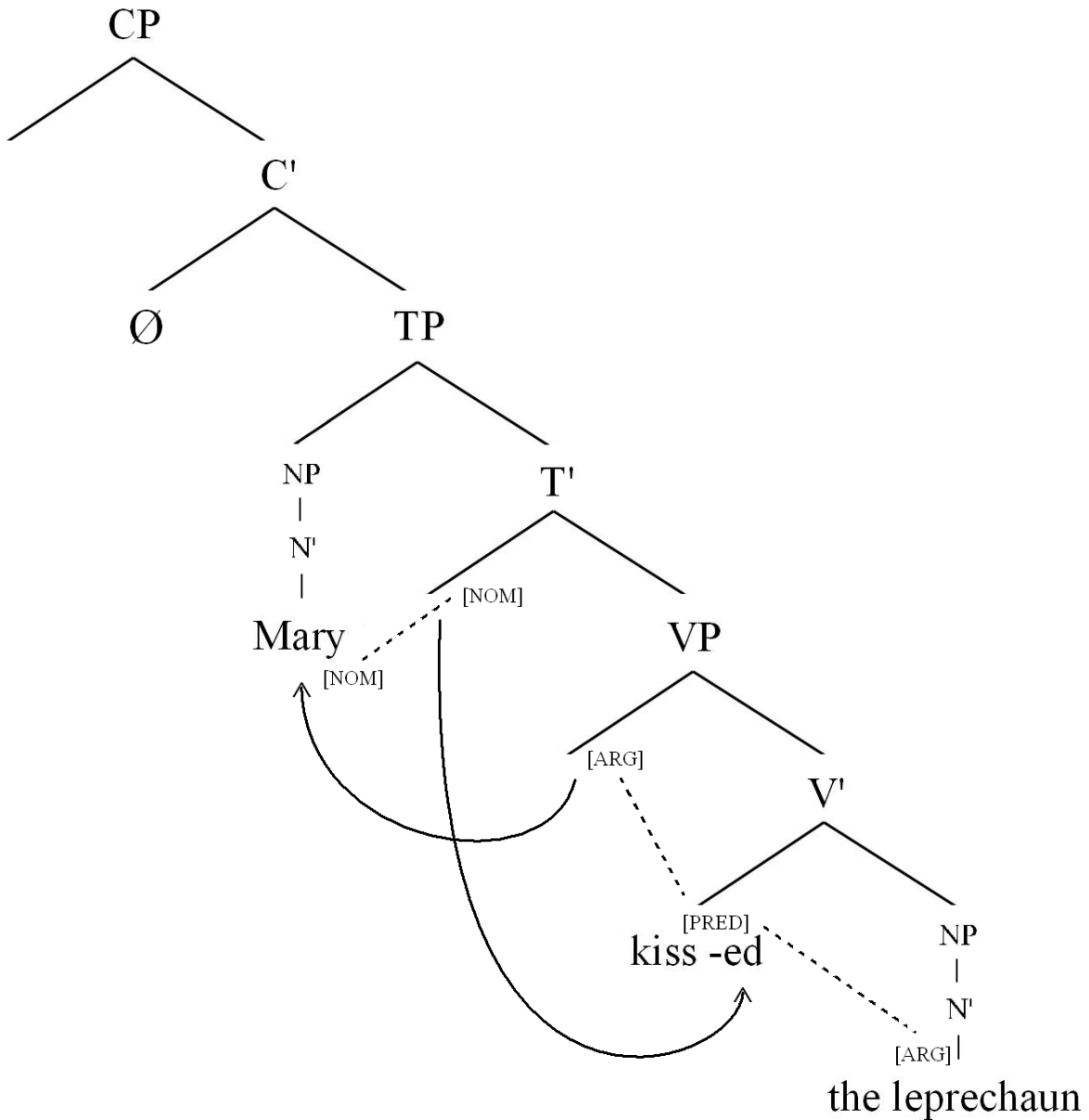
Because of this structure, no other constituent is allowed to intervene between “kissed” and “the leprechaun” – in fact, the fact that the subject NP “Mary” does intervene between them, in the Irish version of the sentence in (14), was a problem, to solve which we had to invoke the movement of V raising. We also saw how (6b) was rendered ungrammatical because of the way the head verb “adopted” and its complement NP “crafty cats with neat whiskers” were separated in it. But why can the head and its complement not be separated? Well, for one, the complement NP provides the verb with one of its

arguments, if it happens to be a transitive or ditransitive verb – and this helps satisfy the Theta Criterion. But the verb also helps out the complement NP in turn, by checking its accusative Case feature, since the NP is also the direct object of the sentence. Accusative Case is assigned by transitive verbs to direct objects as we saw earlier – so by checking this feature, the verb helps its complement NP satisfy the Case Filter. And since the verb and its complement NP have to be local to each other for this feature-checking to occur, we now see why separating the head verb and its complement NP, prior to movement, leads to ungrammaticality.

This tells us something about what the Case Filter achieves when no movement is required, i.e. when all that is needed to generate a successful S-structure is that a head and its complement are placed in adjacent positions in D-structure. But one last achievement of the Case Filter proposal lies in how it also helps explain phenomena where movement *is* involved – which, as I mentioned earlier, happens specifically during NP movement. We already saw how the Case Filter is satisfied when a transitive verb feature-checks a direct object for accusative Case.

But what about the subject of the sentence? The subject of a sentence, at least in English, receives nominative Case, and this is assigned by the occupier of T in a sentence. So, for the Case Filter to be satisfied, the subject NP has to be local to T, so that it can be feature-checked for nominative Case. This is one of the reasons the subject NP of a sentence was generated in the position of specifier of TP in earlier versions of generative theory, as we saw in Example 1.2-18, which is a position that is adjacent to T. But if we accept the VP-internal subject hypothesis, the subject NP is generated inside VP, as a specifier of VP and not TP. How is the subject NP checked for Case in this situation? This is where NP movement is involved. As Example 1.2-24 shows, the subject NP “Mary” in (15a) *raises* from its D-structure position as specifier of VP to its S-structure position as specifier of TP, which is an example, as we saw earlier, of NP movement. Through this movement transformation, “Mary” is brought to a position adjacent to T, which can therefore check it for the nominative Case feature (marked in the example with [NOM] and a

Example 1.2-24. Case marking, the Theta Criterion, and movement in an English sentence



dotted line) and thus satisfy the Case Filter.⁴⁴ Notice that the T constituent here, the T-affix “-ed”, itself lowers to the position of V, for morphophonemic reasons as we saw earlier (although this movement is

⁴⁴ If you revisit Example 1.2-22, you will notice that the subject NP “Máire” there does *not* raise to the position of specifier of TP. Instead it remains in its VP-internal position, to allow the sentence to have the correct VSO order in Irish. How does the subject receive nominative Case then, without which it would violate the Case Filter? Well, notice that the subject is still local to T, even in its VP-internal configuration. It is merely to the right of T, rather

properly a form of V raising rather than T-affix lowering, as I discussed in footnote 41). Finally, notice how the predicate of the sentence, the transitive verb “kiss” (marked with [PRED]), is able to receive its two arguments (i.e. [ARG]) because of their adjacent positions to the predicate as stipulated by the VP-internal subject hypothesis. This, in turn, satisfies the Theta Criterion.

In the above manner we see that NP movement is motivated by a need for the C_{HL} to satisfy the Case Filter constraint.⁴⁵ But NP movement simultaneously satisfies another constraint, viz. that of the EPP. We noted earlier that the EPP requires that all tensed sentences across languages have a subject, either overt or covert. By raising to the position of specifier of TP, the NP “Mary” in Example 1.2-24 also

than the left, which is where it would be as specifier of TP. So, it can still receive nominative Case, *by being c-commanded by T*. This is how it still manages to satisfy the Case Filter, despite staying in its VP-internal position.

⁴⁵ By satisfying the Case Filter, NP movement also allows the moved NP to be pronounced. As we have seen, this happens because the surface form of an NP is distinguished from those of verbal and adjectival constructions by the Case marking it takes, which makes an NP *sound* different from these other constructions, especially in languages with overt, morphological case. So, by getting Case, an NP acquires its surface, pronounceable form. This connection between Case and phonology is especially clear in a grammatical phenomenon known as “control”. To understand this phenomenon, consider the sentence “John is likely to leave” (Carnie (2002): 226). Here the NP “John” is the agent of the predicate verb “leave”. So, as per the locality condition on the Theta Criterion, “John” must be in the same clause as the predicate that assigns its agent theta role, at least initially, prior to movement. In consequence, the D-structure of this sentence is “is likely John to leave”, where “John” appears in the same clause as the predicate “leave” (i.e. “John to leave”). But notice that “leave” is in its infinitive form in this clause, i.e. “to leave”, which results, according to Case theory, in its not being able to assign Case. Therefore, “John” must *raise* – through NP movement – to its position at the front of the sentence, adjacent to the T “is”, in order to receive Case in the way subject NPs do, which generates the S-structure “John is likely to leave”. This raising is also made possible by the fact that the other predicate in the sentence “is likely” does not take an agent as one of its arguments, so the position to its left is open for “John” to raise to. But consider the sentence “John is reluctant to leave”, which appears to be rather similar to “John is likely to leave”. Unlike “is likely”, the predicate “is reluctant” actually takes an agent argument (or more specifically, an “experiencer”). So, “John” must be present adjacent to it in D-structure itself, to satisfy the Theta Criterion, rather than raising to this position after D-structure to get Case. But this means that the other predicate in *this* sentence, i.e. “leave”, will *not* have John adjacent to it in D-structure to give it its agent argument, as it did in “John is likely to leave”, and this would be a violation of the Theta Criterion. So, to fix this problem, generative theory places a special null NP within the clause of the predicate “leave” at D-structure to help it satisfy the Theta Criterion too. This special NP is called “PRO” (not to be confused with little “pro”, which is an NP used in languages with optional subjects). This results in the D-structure “John is reluctant PRO to leave”. Now, “PRO” clearly gets its meaning from the antecedent NP “John” (“PRO” and “John” refer to the same person, since it is John who is both reluctant, and about the possibility of *his* leaving), meaning that “PRO” acts like a pronoun – hence its name. This also means that “PRO” is controlled by “John”, which is why the appearance of “PRO” in the above sentence is an example of the phenomenon of *control* – which occurs specifically to provide an infinitival clause with a subject NP to meet a constraint such as the Theta Criterion. Now, PRO does not move to the front of the sentence in the way “John” did in “John is likely to leave” (although see a recent proposal that treats control as a movement phenomenon too, stated in Boeckx, Hornstein and Nunes (2010)). Which means that PRO cannot get Case, given the infinitival nature of the predicate “leave” here. Does this not violate the Case Filter? No – and precisely because PRO has no phonetic content, and therefore does not need to be Case marked – and which is why it is not pronounced in the sentence “John is reluctant to leave”. This makes PRO a special, *Caseless* NP – and the phenomenon of “control” another example of how Case marking goes hand in hand with phonology, which is what makes Case-marked NPs pronounceable.

provides it with a subject, thus satisfying the EPP. So, NP movement is invoked to satisfy constraints such as the Case Filter and the EPP, without which a sentence would be ungrammatical – although in the process this allows many more S-structures to be generated, which thus fixes X-bar theory’s undergeneration too. In this manner, we see, again, how developments in P&P theory explain how and why sentences have the word orders they do in languages – by proposing constraints on the generation of such sentences, and movement transformations to meet these constraints.

Sometimes a sentence will not have enough NPs to effect an NP movement though. For example, certain verbs (like “rain”, “pour”, “thunder” and other weather verbs) do not assign Theta roles. This means that a sentence with such a verb as a predicate might not have any NPs in it, since no NPs are needed to provide the predicate with its arguments. But without any NPs such a sentence will not have a subject either, and this will violate the EPP. In other words, without any NPs an NP movement cannot be effected to satisfy the EPP, in the way “Mary” did in (15a) above. Recall that it is exactly such a situation, where no movement transformation is possible to satisfy a constraint, that the other transformation of “insertion” kicks in. Also, recall that the specific insertion transformation invoked to satisfy the EPP constraint is that of “Expletive insertion”. An “expletive” is a pronoun like “it” in (17a-b), which does not refer to anything and does not get its meaning from anything either, unlike any other NP we have seen so far:

(17a) It poured yesterday.

(17b) It thundered loudly last night.

By inserting the expletive “it” in the above sentences, they acquire subjects and thus satisfy the EPP. I mentioned another insertion transformation earlier, viz. “Do insertion”. Just like Expletive insertion, Do insertion occurs to meet a constraint on sentence generation when no movement transformation can be invoked to satisfy it. It specifically occurs in the situation where the head-to-head movement of T raising cannot occur. In our earlier discussion of this transformation, we learned that T is often raised to fill the C complementizer position in languages with null complementizers, in order to form questions. In other

words, we can say that T raises to check a Q question feature that the null complementizer is not able to support. (In contrast, languages that have overt complementizers, like Irish, already have a C constituent to support or check the Q feature, and so in these languages T raising does not occur.) But for the T raising movement to occur, a sentence has to have an overt T to move in the first place, such as an auxiliary like “will” or “is”. We saw an example of this in the case of (9a), when we discussed T raising:

(9a) Uche will play the guitar. (cf. “Will Uche play the guitar?”)

Without such an overt T, T raising is not possible, which means that the Q feature will go unchecked, and the sentence (or rather, question) will be ungrammatical:

(18a) *Uche play the guitar?

This is where Do insertion kicks in. “Do” (and its variants “did”, “does” etc.) acts as a meaningless substitute for T, just as “it” acts as a meaningless substitute for an NP. So, in cases where T raising is not possible for want of an overt T, “do”, or a variant thereof, is inserted at C – which then checks the Q feature and renders the question grammatical:

(18b) Did Uche play the guitar? or

(18c) Does Uche play the guitar?

Finally, the idea that “do” and its variants act as substitutes for T can be verified from the fact that when Do insertion occurs, V raising cannot – because the tense feature has already been raised to C and does not require support from a raised V. But it still requires some support, albeit at the C rather than the T position now, since it is not an overt auxiliary. This support is provided by the inserted “do”. This is why in questions that show Do insertion the verb cannot take tense inflections:

(18d) *Did Uche played the guitar? (cf. 18b)

(18e) *Does Uche playing the guitar? (cf. 18c)

There is one last constraint on S-structure generation that we have not yet discussed, viz. the constraint that arises from Bounding Theory. Providing restrictions on the kinds of S-structures that can be generated specifically by *wh-movement* is one of the main contributions of Bounding Theory, so we will explore this constraint through an examination of wh-movement. It so happens that wh-movement was the very example I used, early on in the last chapter, to illustrate the concept of a transformation, and Chomsky's transformational approach to grammar too, both of which we have spent much time exploring in this chapter. So it is fitting that we should end our discussion of transformations and P&P theory with one last look at this movement phenomenon. The same pair of sentences with which I introduced wh-movement in the last chapter can serve again as examples of this phenomenon here:

(1a) Jürgen read a book.

(1b) What did Jürgen read?

First of all, notice how (1a) does not have an overt auxiliary. So, the formation of a question out of (1a) here cannot involve overt head-to-head T raising. Therefore, Do insertion must take place instead:

(1c) Did Jürgen read a book?

Now the object of (1c), i.e. the object of Jürgen's reading, is of course the NP "a book". As should hopefully be clear now, wh-movement involves this object being replaced with the wh-phrase "what", which is then fronted, i.e. it is moved to the front of the sentence, to the left of "did" to get the S-structure of (1b). In accordance with our discussion over the past pages, this fronting can be considered to happen in order to check a feature too, viz. a "wh" feature, which allows the S-structure of wh-questions to be generated (from a D-structure they share with their answers). This checking can only occur if wh-phrases move to the one position available for such checking, which is the specifier of CP position, as we have briefly discussed before. If you revisit Example 1.2-24 you will notice that wh-feature checking, and therefore wh-movement, has to happen in this, leftmost, position in the tree because every other position in the tree is taken. The position to the right of this, the head position of C is normally occupied by the

null complementizer, as the example shows, although in the case of (1b) and (1c) Do insertion puts the word “did” here, making this position unavailable for wh-movement. Next, the specifier of TP position has to be occupied by the subject NP (“Jürgen” in (1a-c)), which raises from the specifier of VP position to get nominative Case, so this position is unavailable for wh-movement too. Finally, the head positions of T and V are also occupied, from covert V raising (which we understood earlier as T-affix lowering, as the example shows), preventing wh-movement to these positions as well. The remaining position, at the bottom right corner of the tree is where the object NP that undergoes wh-movement moves from, so wh-movement obviously cannot occur *to* this position. So, the wh-phrase has to move to the specifier of CP position, right at the front of the sentence, resulting in the wh-movement-related phenomenon of wh-fronting. This also makes sense given that wh-movement targets the object NP, because such a constituent can only move to a non-head position (being phrasal in structure), which is what the specifier of CP position is.⁴⁶

Now, CP is the X-bar way of representing a subordinate clause. So, the fact that wh-movement involves the wh-phrase moving to the specifier of CP position implies that wh-movement can take place within subordinate clauses too. For example, compare “Ulrike believes [_{CP} Jürgen read a book]” with the following two questions, both of which show wh-movement involving the subordinate clause – to the front of it in (1d), and out of it to the front of the main clause itself in (1e):

(1d) Ulrike believes [_{CP} who read a book]?⁴⁷

⁴⁶ The object NP also provides the predicate with one of its arguments, and this predicate, in turn, assigns the NP with accusative Case. So, it might seem that by moving to the front of the sentence, i.e. by moving out of the locality of the predicate, the object NP can no longer participate in these two phenomena – which would violate both the Theta Criterion and the Case Filter. This is not true for the Theta Criterion part of the issue because wh-movement, like head-to-head and NP movement, takes place after D-structure has been generated, at which point the Theta Criterion has already been satisfied. The problem might seem to be more serious for the Case Filter part of the issue because this is a constraint on S-structure, which is generated *by*, not after, movement transformations. But it is a problem only if wh-movement takes place before Case checking occurs – if we stipulate that wh-movement happens *after* Case checking then there is no problem at all.

⁴⁷ If (1d) sounds ungrammatical to you, try stressing “who”, and raising the pitch of your voice towards the end while pronouncing the sentence, as is normal in asking questions. Also note that wh-movement has really occurred in this sentence, even though “who” appears to occupy the same position in the sentence as “Jürgen” did in the original ‘unmoved’ CP “Jürgen read a book”. This is because “Jürgen” appears as the specifier of TP, within the CP – as subject NPs do – but wh-movement moves this NP to the specifier of CP itself, after replacing it with the wh-phrase “who”, as just discussed. So, “who” and “Jürgen” do not occupy the same position in S-structure. However,

(1e) What does Ulrike believe [_{CP} Jürgen read]?

Wh-movement targets the subject NP “Jürgen” of the CP “Jürgen read a book” in (1d). In contrast, it targets the object NP “a book” in (1e). (Note that in (1e) Do insertion must also occur, as in (1c) above, because the entire main clause has now been transformed into a question. This explains the second word “does” in (1e). It also explains why the predicate “believes” in “Ulrike believes Jürgen read a book” is robbed of its T-affix “-s” in (1e) – because this affix has now raised to C, where Do insertion makes it pronounceable (since the inserted “does” is just “do” + “-s”), which leaves (1e) with the tenseless predicate “believe”).

So, wh-movement can target both subject NPs and object NPs, and it can happen from a main clause to the front of that clause, as in (1b), from a subordinate clause to the front of that clause, as in (1d), and from a subordinate clause to the front of a main clause that contains it, as in (1e). But if we try to combine these movements, say the one in (1d) with the one in (1e), we run into trouble:

(1f) *What does Ulrike believe who read?

Now we know that wh-movement occurs in order for wh-feature checking to occur. But this has to occur in accordance with restrictions provided by Bounding Theory, or else ungrammaticality will result – which is exactly what happens in (1f). The specific restriction Bounding Theory places on wh-movement is that it requires two nodes or positions in a sentence tree that are involved in a movement transformation to be separated by no more than one “bounding node” – a “bounding node” being a position in a tree (an NP or a TP in the case of wh-movement) that limits how many and what kinds of movement transformations can take place across it.^{48,49} To this extent, bounding nodes place a locality constraint on

they both seem to appear in the same place, to the right of the verb “believes”, because the CP that they are both in is the complement of this verb – which makes the CP take the position immediately to the right of the verb too.

⁴⁸ This specific restriction on wh-movement is also known as “subjacency”, from an earlier proposal by Noam Chomsky (Chomsky (1973): 261-262), which was articulated more fully within Bounding Theory in the 1980s.

⁴⁹ A bounding node is also referred to as a “barrier”, a name that arises from the title of a text by Noam Chomsky in which he proposes his version of Bounding Theory (viz. Chomsky (1986a)).

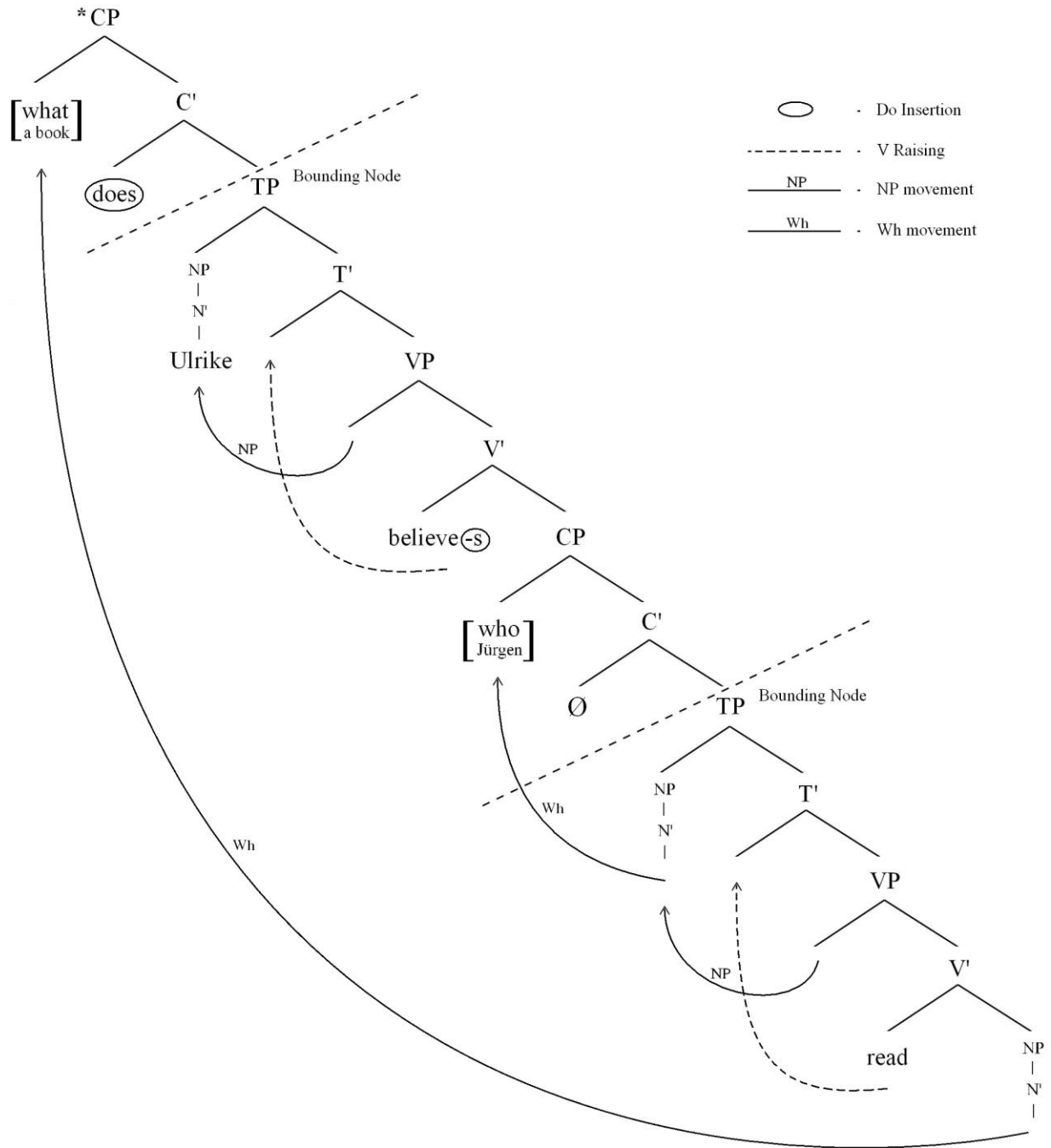
sentence generation too, of the kind we have explored earlier in the context of Binding Theory, the Theta Criterion and the Case Filter. I shall return to this important feature later.

To see how the above Bounding constraint leads to the ungrammaticality of (1f), consider Example 1.2-25. This example gives the ill-formed S-structure of (1f). This is also the last example of linguistic transformations that we will look at, so it might not be surprising that it contains examples of almost all of the transformations we have seen so far, in addition to the examples of wh-movement currently under consideration. So, it might be worth exploring the structure of this sentence in some detail.

Let us start with the leftmost constituent in the sentence, the wh-phrase “what”. As suggested earlier, this word is really the object NP “a book” in the subordinate clause “Jürgen read a book”, which has undergone wh-movement to the front of the sentence, as shown by the long arrow connecting the two ends of the sentence. To the right of “what” is “does”, which is positioned here by Do insertion as the circle around it indicates. “Does” is of course the “do” of Do insertion with an -s T-affix, the latter of which it gets from “believe-s” further down the sentence, as shown by the circle around the “-s”. Next to “does” is the main clause’s subject NP “Ulrike”, which occupies the specifier of TP position after raising from its VP-internal specifier position through NP movement, as the arrow with the NP index indicates. The head T position in between these two specifier positions is occupied by the verb “believe”, after covert raising from its VP-internal head verb position, as the dotted arrow shows.

This brings us to the embedded tree of the subordinate clause. In the leftmost, specifier position of this CP, which is to the right of the verb “believe” (given that the CP is the complement of this verb), we find the wh-phrase “who”, which represents this CP’s subject NP “Jürgen”. This word appears in this position after wh-movement from the position occupied by “Jürgen” earlier, viz. the position of specifier of TP within the subordinate clause. But as the arrow *to* this position shows, “Jürgen” actually arrives at this position after NP movement, from its initial, VP-internal, D-structure position within the subordinate clause, where “Jürgen” was the VP’s specifier. Meanwhile, the head C position of the subordinate clause

Example 1.2-25. Wh-movement and Bounding in English



does not show Do insertion, and so reveals a null complementizer instead. Finally, the verb “read” undergoes covert V raising from the position of head verb within the subordinate clause to the position of head T.

What this complicated structure shows is that all the branching positions within the sentence tree are occupied. This means that after all the NP movements, V raisings and Do insertions have occurred, there are only two positions left for wh-movement to manifest itself – viz. at the position of specifier of the *main* clause, right at the top of the tree, or at the position of specifier of the *subordinate* clause (where we see the wh-phrase “who” positioned in Example 1.2-25). (There is also the position filled by \emptyset , which does not figure in any transformation operation in the sentence, but it cannot figure in wh-movement either because it is occupied by the null complementizer.) That only the two CP specifier positions are available for wh-movement might not seem problematic, since the CP specifier position is to where wh-movement normally occurs – so it might seem that such a sentence can support at least two wh-movements. But consider how a second wh-movement might manifest itself in the sentence, after a first wh-movement has taken place (which by itself is okay, as we have seen).

Let us say the first wh-movement is the one that targets the subject NP “Jürgen”, which moves the wh-phrase “who” to the specifier of the subordinate clause. This means that the second wh-movement, which targets the object NP “a book” would have to move the wh-phrase “what” all the way to the front of the sentence, to the specifier of the main clause, the only position now available for this second wh-movement. This would make the original, rightmost, position of “the book”, and the CP specifier position at the front of the sentence it has undergone wh-movement to, the two nodes in (1f)’s tree that participate in this second wh-movement transformation. However, as the example shows, there are two TP bounding nodes between these positions – which violates the Bounding constraint we stipulated above, and this is what makes (1f) ungrammatical.

It might seem that a solution to this would be to do the two wh-movements the other way around. That is, one might move “what” to the position occupied by “who” first – where it just crosses the lower of the two TPs, in accordance with the Bounding constraint – and then to its final position at the front of

the sentence. If the second wh-movement occurs only at this point, moving “who” into the subordinate CP position just vacated by “what”, we might end up crossing the two TP bounding nodes one at a time. This manner of generating an S-structure is called “cyclic”, since it involves generating the two CPs of this sentence, first the subordinate one, then the main one – one at a time, in two separate cycles. Cyclicity is actually required to ensure that two bounding nodes are not crossed within a single generative operation – i.e. it is required to satisfy Bounding constraints on S-structure generation. (In fact, this is what makes (1e) grammatical. This sentence involves a wh-movement that crosses two TPs – but it does so cyclically, which helps avoid violating the Bounding constraint on wh-movement.) But even cyclic generation will not prevent (1f) from being ungrammatical. This is because even if “what” moves to the subordinate clause’s specifier position first, and then to the final, main clause specifier position, it leaves behind what is called a “trace” of itself at the earlier position, which blocks “who” from moving here subsequently. So, for (1f) even to be generated, “who” *has* to move first, leaving “what” with no option but to proceed in acyclic fashion directly to the front of the sentence, producing the ill-formed S-structure of (1f):⁵⁰

(1f) *What_{obj} does Ulrike believe who_{subj} *t*_{subj} read *t*_{obj}? (here *t* represents the trace of a constituent that has moved, which is indexed to the trace via a shared subscript)

Another way of putting this is to say that after we have generated (1d), by moving “who” to the subordinate clause specifier position, we cannot move “what” to the front of the sentence in a second wh-

⁵⁰ An alternative position within generative theory states that “what” leaves an entire copy of itself in the earlier position from which it has moved, rather than a trace. Evidence for this copy theory of movement comes from the fact that in some languages, e.g. Afrikaans and vernacular German, a moved wh-phrase may be pronounced in *both* the earlier and the final, moved positions. Of course, in a language like English, both wh-phrases will not be pronounced, since the resulting sentence, as in (1f), would be deemed ungrammatical:

(1f) *What does Ulrike believe who who read what?

So, which “what” and which “who” is actually pronounced, when S-structure maps to the articulatory system (via the PF interface), becomes a parametric matter, which varies across languages. In English, the highest, i.e. leftmost, copy is pronounced. In languages like Mandarin the lowest copy is pronounced, which gives the impression that wh-movement does not occur in these languages (called “wh-in-situ” languages). However, wh-movement does occur in wh-in-situ languages too, albeit covertly, as discussed in footnote 41 – and covert wh-movement can even occur in a language like English, in “echo questions” like “Who saw what”. (For a discussion of echo questions, see Santorini and Kroch (2007), accessible at: <http://www.ling.upenn.edu/~beatrice/syntax-textbook/ch13.html#notes-wh-in-situ.>)

movement. This means that once “who” has moved in the first wh-movement, the resulting subordinate clause in (1d) “who read a book” does not allow “what” to move out of it to the front of the sentence (after replacing “a book”, of course). This makes the CP “who read a book” an example of a structure called an “island”, first described by the linguist John Ross in his doctoral dissertation (Ross (1967)). An island is a place one cannot easily leave without a boat or a plane; grammatical islands are similar in that they are structures that one cannot easily move out of in a movement transformation, without violating a constraint of some kind. In the case of a wh-island like “who read a book”, the wh-phrase “what” cannot move out of it in a wh-movement without violating a Bounding constraint.⁵¹

This ends our discussion of P&P theory. It also brings us to my long-promised discussion of the Minimalist Program now, and how the MP has built upon the already substantial achievements of generative theory in the past several decades – much of which we have explored in this section.

As I have mentioned before, one of the main goals of Minimalism has been to increase the explanatory power of generative theory even further than what X-bar and P&P theory accomplished, while streamlining the theoretical baggage inherent in them to the bare minimum required for descriptive and explanatory adequacy. Part of this goal is programmatic – it is motivated by a desire to see linguistic theory explain language in terms of the most basic, elegant, and efficient principles, in a similar vein to the great scientific theories of the universe. (This has also been the basis for criticisms by anti-Minimalists, who think that this programmatic goal makes the MP unfalsifiable and unnecessarily idealistic (e.g. see Lappin, Levine and Johnson (2000), and the responses by Piattelli-Palmarini (2000) and Uriagereka (2000)). It is for this specific reason that Minimalists describe the MP as a “research

⁵¹ There are actually some kinds of sentences, also identified by John Ross, in which wh-movement across more than one bounding node can occur – that is, in these sentences, one *can* legitimately move out of a wh-island without violating the Bounding constraint we have been discussing above. An example of such a sentence is “[Which book]_i did Ulrike forget how_j Jürgen read *t*_i *t*_j?”, which is answered by “Ulrike forgot Jürgen read [Gunther’s book]_i [with his new glasses]_j”. To account for these sentences, another constraint is needed, which these sentences satisfy – and which therefore allows them to be grammatical, even though they violate the earlier Bounding constraint we have been discussing. This new constraint is called the “Empty Category Principle” (or the “ECP”). I will not discuss the ECP here in the interests of space.

program” too and not an actual theory, since the theoretical basis for the MP is just the P&P framework (which in itself does make truly theoretical claims and testable hypotheses).

But programmatic goals aside there are several aspects of older generative theory that are cumbersome and empirically problematic – so the MP’s goal of streamlining this older theory is driven at least in part by an actual, empirical requirement to make generative theory deal with the facts better. Much of this has to do with a Minimalist observation about language that we examined in the last chapter, which is that language seems to be unique in the way that it is *underspecified* and *economical* in its structure. For example, it is because C_{HL} is underspecified that the parameters of an I-language have to be set, before one can be said to have acquired that language. Therefore, the MP seeks to ascertain how and why the P&P architecture of C_{HL} shows such underspecification and economy. As the linguist Cedric Boeckx says:

“Minimalism seeks to understand why the general conditions of grammar are of the sort the P&P system [has] roughly established. In particular, minimalism proposes investigating this question by asking how much of the P&P architecture follows from general properties of optimal, computationally efficient systems. Minimalism’s *why*-questions are strongly reminiscent of general questions asked in modern theoretical physics: one may have a good theory, even the right theory, of what the universe is; still, one can worry about such issues as *why* the theory invokes the dimensions it appears to have ... or even *why* the universe it describes exists at all.” (Boeckx (2006): 8, see also 85-87)

There are three parts to this issue in my opinion, or rather three aspects of C_{HL} , to which the MP has made significant contributions, and so these are the three issues I will deal with over the next few pages. First is the issue of how the generation of D-structure is described in generative theory, second is the issue of how S-structure generation is described (i.e. via transformations), and third is the issue of how all of this is constrained so that the output of C_{HL} can be mapped adequately to phonology and semantics.

First, regarding the generation of D-structure, we have seen that the main achievement of generative theory here has been X-bar theory. X-bar theory was motivated by a desire to simplify Phrase Structure Rules, both to account for similarities between different PSRs, and to account for the patterned, “Specifier/Head/Complement” fine structure of phrases that PSRs did not account for either. (This

motivation behind X-bar theory should illustrate how the drive towards simplicity and elegance in generative linguistics long precedes the MP.)

Now, X-bar theory had some shortcomings, as we have seen, which work in transformations, Binding Theory etc. has helped address. But X-bar theory can be improved upon even in the case of D-structures that it *does* generate successfully. For example, we have seen how X-bar rules can generate NPs successfully – but then there is the issue of whether NPs should really be thought of in terms of DPs, which is an issue that cannot be fully resolved within X-bar theory itself (which is why several linguists have continued to avoid treating NPs as DPs, a strategy I pursued myself in this section). The problem here is that X-bar theory assumes (albeit based on a large body of evidence) that phrases in all languages have a “Specifier/Head/Complement” structure – which means that a phrase (i.e. an XP) will be assumed to have this structure before we even generate it. This means in turn that we will also have to determine what constituents occupy the different positions in X-bar theory’s Specifier/Head/Complement representational schema – which is precisely what led to the problem of whether an NP is really a DP. For a research program that strives towards elegance and simplicity in its descriptions, such a complicated *a priori* representational system as the one X-bar theory proposes becomes a shortcoming, especially given its attendant problems, even if it explains a significant amount of data, and presents an improvement over the earlier PSR-based Standard Theory.

In fact, if simplicity and elegance is a goal, all that a grammatical theory should propose is a simple set of computational operations (preferably just one) to join constituents into more complex representations whose sole purpose is to provide a mapping to phonology and semantics – without which they would be incomprehensible and unpronounceable. Anything more than this would be *conceptually unnecessary*. This is what leads to the Minimalist Program’s revision of X-bar theory.

In the matter of describing D-structure, the MP does not assume X-bar theory’s representational schema. Instead, it takes a *derivational* approach to generating grammatical structures. In this approach, no preexisting structures are assumed – all that is assumed is in fact the lexicon, and just *one* computational operation called Merge, which joins lexical items into more complex structures. (So, not

even X-bar theory's basic set of rules to generate X' and XP structures is assumed here. This means, importantly, and as was mentioned in the last chapter, that *grammar is no longer being treated as a "rule-based" system*, as it previously was – and as it still often is in more popular contexts.)⁵² Merge joins lexical items to form sets of two items at a time, and in this manner derives more complex grammatical structures. Given the two-member sets it forms through its derivations, Merge automatically generates binary-branching structures too, which is in accordance with earlier beliefs about how linguistic structures are organized. And this is *all* that is needed to generate the D-structure of a sentence.

So this Minimalist revision of X-bar theory assumes no specifiers, heads or complements. It just derives a minimal structure arising from the workings of Merge on the lexicon, called "Bare Phrase Structure" (Chomsky (1995c): 245-249). But specifiers, heads, and complements *can* arise from bare phrase structure though, as emergent properties. To understand this, consider the various possibilities for how Merge might join lexical items. Logically speaking, three possibilities might arise (there is actually a fourth, which I shall discuss later):

- (a) Merge joins a word with another word (creating a phrase).
- (b) Merge joins a word with a phrase (the latter created by (a))
- (c) Merge joins a phrase with another phrase (both created by (a))

If Merge joins a word and a phrase, as per (b), the word will be the head of the resulting set of items because a phrase cannot be a head, as we have seen before – which means that the phrase will then be the complement to the head word. In this way, the head-complement structure within X-bar trees can arise as a by-product of a Merge-based derivational operation. For this reason (c) is actually impossible too,

⁵² In other words, language users do not have to know any rules to generate grammatical sentences. They just need to learn the lexicon of their native languages, which Merge automatically combines into phrases and sentences. Of course, Merge operates in a systematic way too, i.e. according to abstract principles of structure generation that Noam Chomsky calls "third factor principles" (Chomsky (2005): 6-11), which the language user needs to (innately) know as well. But such principles are different from grammatical rules, of the kind we see in X-bar theory, and even of the kind we see in language classes, where acquiring grammar usually involves learning 'rules' about which words to use or not to use in a grammatical sentence. (See Uriagereka (1998): 193 for a longer discussion of this issue.)

because if Merge were to join a phrase with another phrase, the resulting, complex, phrase will have no head – which is impossible by definition.

We have (a) above left to discuss, and this has a bearing on how X-bar theory’s “specifier” category arises in its Minimalist revision. As we have discussed before, lexical information from a word projects up a grammatical tree when the word is added to the tree – which makes feature-checking for Case etc. possible, among other things. So, when two words are joined by Merge, as in (a), lexical information of some sort will project up the tree. But in logical terms lexical information from only word can project.⁵³ It should be obvious now that *this* word will be the head word of the pair, the other its complement. Now, since the two words have merged, if the head word merges with another word or phrase, it will really be its bar-level projection that is doing the merger – the bar-level projection being the result of the head’s merger with its complement. So, any word or phrase that merges with this bar-level projection will automatically be what X-bar theory called a specifier – since specifiers are branching sisters to bar-level projections. In this manner, all X-bar representations can be derived from the workings of a single Merge operation, without us having to assume any of them beforehand.

So, we can see now how the MP reduces X-bar theory to only the minimal number of objects and processes logically and conceptually required to generate a D-structure – and it also eliminates some of

⁵³ Since the two merged words form a set, there are three set-theoretic possibilities for how their lexical information might project. Information from only one of them might project, or information from both of them might project – in which case it will either be the union (in set-theoretic terms, i.e. \cup) of their information, or the intersection \cap of their information. However, the two latter possibilities are both untenable, so only information from one word can logically project. The union of their information cannot project because if some of their information is mutually contradictory, then the union of their information will be of an indeterminate nature – and if this projects it would lead to an indeterminate language, some sort of ‘quantum language’ (cf. Uriagereka (1998): 178). (This might happen if one of the words being merged is a noun and the other a verb, which would generate a phrase like “babies cry”. The fact that one word is a noun will be represented in its lexical information as a noun feature, say +N. But the fact that the other word is a verb will be represented in *its* lexicon as a verb feature, or more specifically a ‘not-noun’ feature, i.e. –N. So, the union of their information would have both +N and –N features – which means that if this information projects, the resulting phrase will not be a noun phrase or a verb phrase but an indeterminate ‘noun and not-noun’ phrase. The intersection of their information fares no better. If two words with mutually contradictory information are merged, then the intersection of their information (i.e. what information they have in common) will not be indeterminate, as in the case of the union of their information, but it could be null (if they have no features in common) – which means that *no* information would project. This would imply that if a noun and verb are merged, they might not be able to form a legitimate phrase at all – which is obviously wrong because we can form a phrase like “babies cry”.

the problems with X-bar theory in the process. But its streamlining does not stop there. We have only discussed the MP's attitude to D-structure. What about its description of how S-structure arises? Well, we have seen that to correct for X-bar theory's undergeneration, many proposals regarding transformations were proposed, especially during X-bar theory's evolution into P&P theory – and these affect how S-structures are generated. But if you review all the transformations we have explored in this section, you will notice that they all have one thing in common; they involve placing a constituent in a certain position or node in a sentence tree, in order to check a certain lexical feature, such as that of Case. Moreover, how and where a constituent is placed in a transformation always seems to depend on a constraint on transformations that involves a locality condition, i.e. a constituent must be placed in a position that is local to the very constituent that checks its features. This is why subject NPs must move to the specifier of TP position for Case checking, and wh-phrases must move to the specifier of CP position for wh-feature checking (although wh-phrases must also move to the CP specifier position *cyclically* to avoid Bounding violations, and this adds another localizing constraint on this type of movement).

What this suggests is that all transformations are really just instances of a general operation, whose job is basically to 'move something somewhere' in accordance with locality constraints. Generative linguists actually realized this long before the ideas of the MP were developed. This is what Noam Chomsky had to say about this more general operation all the way back in 1980, which came to be known as "Move α " or simply "Move":

"The fundamental idea is that surface structures are formed through the interaction of at least two distinct types of rules: base rules, which generate abstract phrase structure representations and transformational rules, which move elements and otherwise rearrange structures to give surface structure ... the transformational mapping to S-structure can be reduced to (possibly multiple applications of) a single general rule: 'Move Alpha'." (Chomsky (1980a): 144-145)

So, all transformation operations can be reduced to a single operation "Move". The insight that the MP brought to this idea is that "Move" is actually an instance of Merge – which means that even this general operation can be reduced to the workings of a single Merge operation across D- and S-structure! Recall the discussion from a few pages ago about the possible ways in which Merge can join two items. We

examined three possibilities there, viz. Merge joining a word with another word, a word with a phrase, and a phrase with another phrase (with only the first two being tenable). Notice that all these mergers involve an item being merged with a *different* item to form a larger structure. But recall how I also mentioned that there is a fourth possibility for how Merge might join two items – and that possibility is of an item joining *with itself*. If you think about it, this is exactly what happens during a movement transformation – a word (or a phrasal projection of a word, such as an NP) is merged with the same structure that it originally occurs in, albeit in a different position in that structure (hence the appearance of movement). So, the transformations that yield S-structure, which were previously thought to be instances of Move, can now be seen as just instances of Merge working *internally* – i.e. working in a way that merges an item with another item that is internal to it. In contrast, the Merge operations that generate D-structure are instances of an *external* Merge, since they involve merging an item with another one that is distinct from it, i.e. one that is external to it. So, just one Merge operation can now be seen to be the basis for both D-structure and S-structure generation – and this is why the MP does away with the distinction between D- and S-structure.

Now, there is still a third aspect of C_{HL} , in addition to its generation of D- and S-structures, whose description has been fruitfully and significantly revised by the MP. As I mentioned earlier, this is the issue of how C_{HL} 's workings are constrained so that its outputs can be comprehended by the semantic (or “conceptual-intentional”) system, and articulated by the phonological (or “sensorimotor”) system. After all, the only two conceptually necessary statements a generative theory needs to make are about the operation through which C_{HL} joins lexical items into more complex representations, and about how these representations map to semantics and phonology. The Merge proposal takes care of the first of these two statements, so now we have to deal with the second one.

The first step in this direction is to include in a grammatical theory only those structural representations that are conceptually necessary for the mapping of grammar to semantics and phonology. The various representations of X-bar theory were not seen as being such representations, which is why the MP streamlined them out of grammatical theory, through its Merge-based derivational approach. Instead,

the MP suggests that only the final output of C_{HL} , S-structure, is necessary to realize the mapping between grammar and semantics/phonology, but only if it takes either a purely semantic or a purely phonological form. That is, the only structural representation that grammatical theory needs to necessarily propose to realize the mapping of grammar to the *semantic* system, is an S-structure that is devoid of all but semantic information, and similarly, the only structural representation that grammatical theory needs to necessarily propose to realize the mapping of grammar to the phonological system, is an S-structure that is devoid of all but phonetic information. As we saw in the last chapter, the MP realizes this by proposing the levels of structural representations called Logical Form (LF), which is the level at which a purely semantic representation of an S-structure is generated, and Phonetic Form (PF), which is the level at which a purely phonological representation of an S-structure is generated. So, LF and PF representations are all that generative grammatical theory needs to propose to realize the conceptually necessary mapping between C_{HL} , and the conceptual-intentional and sensorimotor systems, without which generated sentences would be incomprehensible or unpronounceable.

But we also need a way to get Merge to generate *legitimate* LF and PF representations. That is, grammatical theory necessarily needs to propose a constraint on what Merge generates too, so that legitimate LF and PF structures result from the workings of C_{HL} . Now, we have seen how generative theory has proposed a variety of constraints, both on D-structure and on S-structure generation, which ensures that a resulting sentence will be comprehensible and pronounceable. For example, we explored the Case Filter constraint, whose main purpose is to ensure that NPs are adequately checked for Case, and therefore *pronounceable*. Since we have simplified the workings of C_{HL} to a single Merge operation, operating across D- and S-structure, doing away with the distinction between D- and S-structure in the process, can we simplify the variety of constraints on sentence generation across D- and S-structure to a single constraint on Merge too? This would not only be in accordance with the MP's goals of simplicity and elegance, but would also help us reduce grammatical theory's conceptually unnecessary baggage in the department of constraints.

The MP's answer to the above is that, yes, we can simplify the set of constraints imposed on sentence generation, primarily because all the constraints we have discussed – just like all the transformations we have discussed – seem to have much in common, and therefore seem to be instances of a single, more general, constraint on sentence generation. To understand this, think about how all the constraints on both D- and S-structure generation that we have explored seem to make the two same requirements of all sentence generation operations. First, they all require that a lexical feature be checked in the generation of a structure, and secondly that constituents be placed in certain local configurations during generation, so that this checking can take place. (In fact, all the transformations we looked at are able to meet these two requirements, which is why they appear to be instances of a single, more general, grammatical operation too.) So, the Binding constraint requires that NPs be placed in the appropriate local configuration (i.e. binding domain) so that their semantic features can be checked against the relevant antecedent – to ensure that they refer to the same person, place or thing. The Theta constraint requires that NPs be placed in the appropriate local configuration (in this case, the same clause as the relevant predicate) so that their argument features can be checked against this predicate. The Case Filter requires that NPs be placed in the appropriate local configuration (in this case, adjacent to the relevant auxiliary or transitive verb) so that their Case features can be checked against this verb. The Bounding constraint requires that wh-phrases be placed in the appropriate local configuration (in this case, in the specifier of CP position) so that their wh-features can be checked at the edge of each CP (which is where the specifier of CP position is) to ensure that wh-structures are generated cyclically, in order to avoid Bounding violations.⁵⁴

The inference we might derive from the above facts is that there seems to be a single, general, constraint on Merge-based sentence generation of which all the above constraints are instances. This is

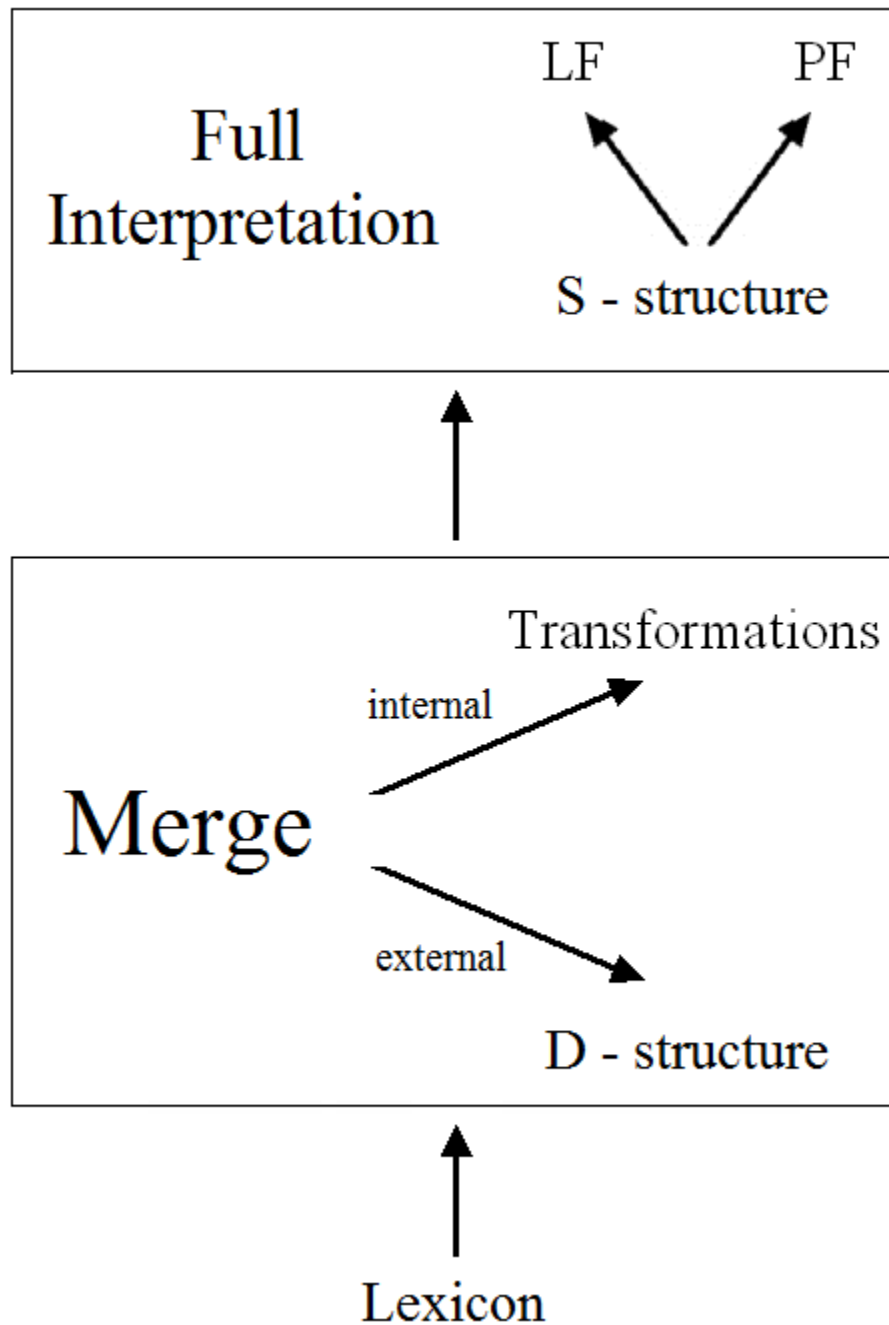
⁵⁴ I did not discuss the EPP constraint above because why this constraint is required by grammar is still a bit of a mystery. However, there is a traditionally-accepted argument, owing to the linguist David Pesetsky's work (see Pesetsky (1982)), which states that sentences are extended projections of T (which is why we took a sentence to be a TP in the X-bar model). So, for an NP to be recognized as the subject of a sentence it has to be in a local configuration to T – hence the positioning of subject NPs in the specifier of TP position in an X-bar tree to satisfy the EPP.

because all of the above constraints essentially require one thing of Merge, viz. that it joins constituents whose features can be checked in the appropriate local configurations, so that legitimate S- (and therefore LF and PF) structures result from such merger, which ensures, in turn, that a generated structure is comprehensible and pronounceable. Minimalists call this single, general, constraint on the workings of Merge the constraint of “Full Interpretation”. The term “interpretation” here comes from the fact that the constraint’s job is to ensure a successful mapping between C_{HL} , and the semantic and phonological systems. So, we could say that it makes the outputs of C_{HL} *interpretable* by the semantic and phonological systems. If a structure is ‘fully interpreted’ by both systems, the semantic system will be able to ‘semantically interpret’ it (i.e. comprehend it), and the phonological system will be able to ‘phonetically interpret’ it (i.e. find the correct pronunciation for it). In this way, the MP streamlines not only the actual workings of C_{HL} , but also the constraints that govern these workings.

Note though, that just because Full Interpretation makes the outputs of C_{HL} interpretable, it does not ensure that they will necessarily be interpreted too. (Examples of this are sentences that are grammatical, yet semantically ambiguous, and which are therefore hard to interpret. We will look at some sentences like this in the next section.) Interpretation by the semantic and phonological systems is ultimately a matter of *performance*, and not *competence* (see section 1.1.1 from the previous chapter to review this distinction). So, there could be extra-grammatical performance factors that prevent an output of C_{HL} from being interpreted, even when it is grammatical. All that (grammatical) competence requires is that C_{HL} be able to generate outputs that are interpretable, i.e. structures that can receive Full Interpretation at the levels of PF and LF.

Example 1.2-26 gives a revised sketch of the Minimalist architecture of C_{HL} discussed in the last chapter. Here we see the Minimalist goal of describing the human faculty of language in simple and elegant terms at work, in all its glory. If you compare this sketch with the previous sketches from the 1960s, 70s and 80s (shown in Examples 1.2-15 and 21), you will notice what a drastic simplification of

Example 1.2-26. An architecture of C_{HL} according to the Minimalist Program (II)



those models this one is. The complicated set of rules in them that generate D- and S-structure are reduced to the external and internal workings of a single Merge operation, all of which is constrained by a single constraint on interpretability.

What is even more relevant about this model for our current, musical, purposes is that by eliminating much of the software from earlier versions of generative theory, the MP also makes the above sketch of C_{HL} applicable to other human faculties, including music. Previous versions of generative theory made use of concepts and phenomena that are clearly language-specific, such as the language-specific Phrase Structure or X-bar rules for structure generation, the language-specific structures like NPs and VPs that are generated by these rules, and the language-specific constraints on the generation of said NPs and VPs, such as those provided by Case, Binding and Bounding theory. By eliminating many of these concepts and phenomena from its architecture, the MP provides a view of C_{HL} that is much more amenable to being applied to other faculties, since all that it requires of a faculty is that it have a set of inputs provided by a lexicon, which Merge can join into multi-elements sets, and a set of external constraints on the generation of these sets, provided by systems like those of semantics and phonology. Noam Chomsky makes this point himself, when he says:

“In the earlier framework, not only rules but also UG principles were expressed in terms of grammatical constructions ... all inherently specific to language, without even remote counterparts in other biological systems. Within the P&P framework [and presumably its Minimalist extensions], the basic computational ingredients are considerably more abstract ... and it becomes quite reasonable to seek principled explanation in terms that may apply well beyond language, as well as related properties in other systems.” (Chomsky (2005): 9)

As I have argued earlier in this chapter, music seems to have, albeit controversially, something akin to a lexicon, whose constituents can be joined into larger structures by the set-theoretic workings of Merge. And as I will argue extensively in the second half of this dissertation, musical meaning and rhythm seem to provide the kinds of constraints on interpretability that semantics and phonology in language impose on the workings of Merge within C_{HL} too.

So, armed with the tools and techniques that generative linguistics has provided us over the past 50 years in its description of C_{HL} , as we have now explored extensively, let us now move on to an examination of the computational system of human music (i.e. C_{HM}) itself, proposed earlier in this dissertation as the locus of human musicality.

1.2.4. The Relation of C_{HM} to C_{HL}

Rather than providing a comprehensive theory of C_{HM} , which I have argued as being impossible at this early stage of research in cognitive music theory, I will proceed with this examination by exploring a number of issues within music theory about how musical sentences are generated by musical grammar. Part of the reason behind this particular approach, in addition to the one just mentioned, is to demonstrate how ideas from linguistics can shed light on some important issues within music theory itself. Ultimately though, this examination will also hopefully illustrate the deep overlap, if not identity, between the human computational systems of music and language – musical and linguistic grammar – which is the basis for this dissertation.

i. Keiler on *Stufen* and binary-branching musical trees

As was the case with our discussion of C_{HL} in the last section, the first issue that might be worth exploring for music is whether music has constituents like language, which make up the structure of phrases and sentences. These would be the entities that are involved in, and which result from, Merge-based operations – so for a Minimalist Program for music to work, a theory of musical grammar should be able to include a description of constituency as well. Part of this description will also have to deal with the notion of a lexicon, of course, because that is what constituents are formed from. And this is a tricky notion within music, as I discussed in the first section of this chapter. But this issue is still worth exploring a little bit more, mainly because it will help with our discussions in some of the subsequent sections of the chapter.

The notion of a musical constituent within a theory of musical grammar was discussed insightfully by Allan Keiler, in Keiler (1977). In this paper, he related the notion of a musical constituent to Schenker's concept of the scale-step, or *Stufe*. Recall from the discussion in section 1.2.1 my proposal that chords make up a lexicon for music, albeit not in the triadic 'vertical' sense as normally understood in Western music but rather in a more melodic, "structural motive" sense. In this sense, what gives a melodic 'chord' its grammatical function is the *Stufe* it represents, which I discussed in the context of a Roman numeral analysis of a passage from Beethoven's Kreutzer sonata. This means that even though musical grammar is realized melodically, its foundation is ultimately harmonic, although harmonic in the abstract, Schenkerian *Stufe* sense.

This point, about musical grammar having a harmonic basis, is essentially 'Keiler-ian'. In some of his earliest work on this topic, Keiler reviewed the Schenkerian idea that musical structure arises through melodic means, but then argued that Schenker is inconsistent about whether a passage thus generated has a harmonic, *Stufe*-based status or not. His main claim was that Schenker assigns a *Stufe* to a sonority or string of sonorities only when they are modified with other sonorities, but, inconsistently, not to the modifying sonorities themselves. Keiler says specifically that the musical examples in Schenker's writing in which such assignments can be seen:

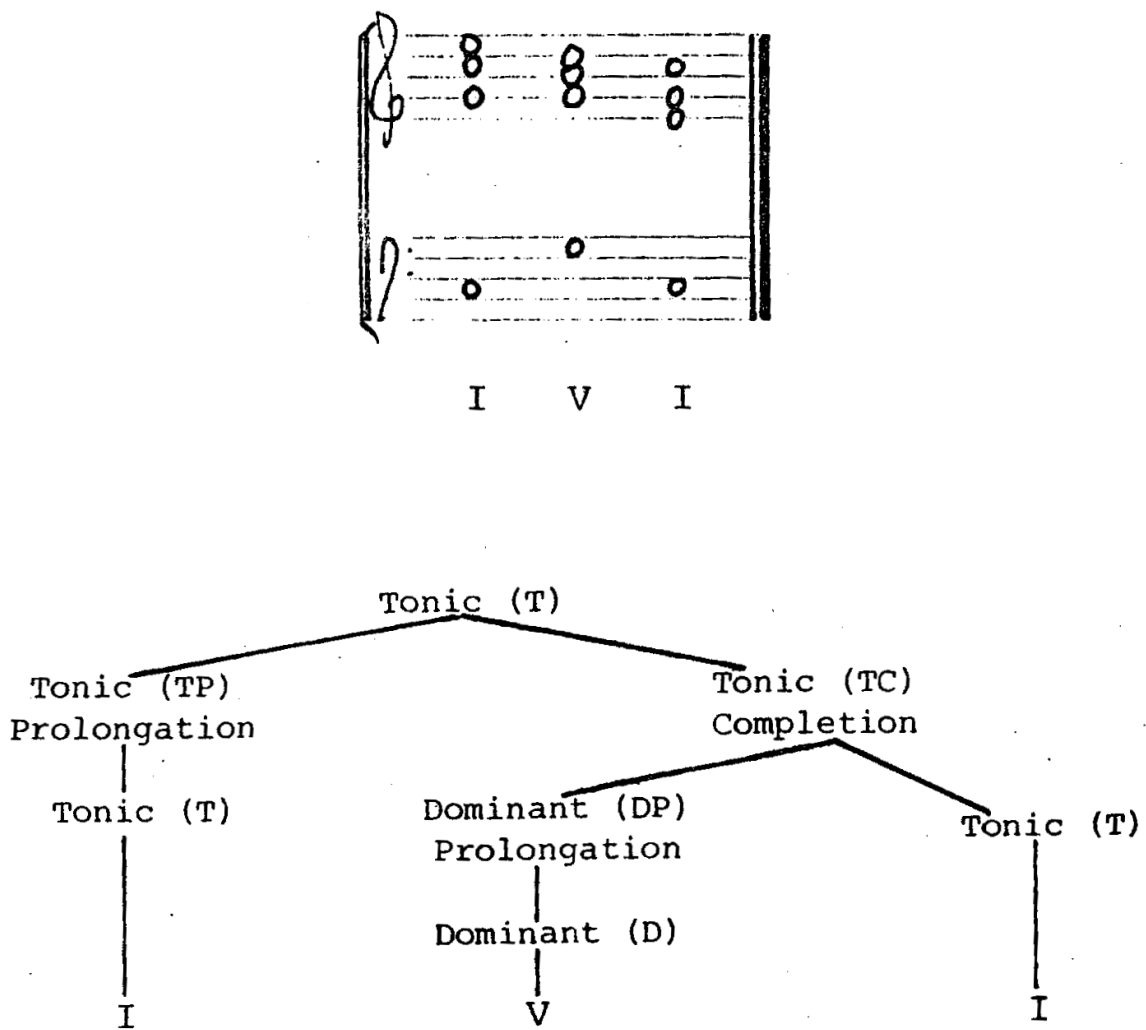
"...demonstrate what seems to me to be a conceptual arbitrariness about the application of the concept [of the *Stufe*] to specific derivations. Since any passing sonority on any hierarchical level of elaboration may be potentially the source (or goal) of further prolongation on some lower hierarchical level, it is ad hoc to insist that *Stufen* exist only on the level of the *Ursatz*, or that certain voice-leading procedures relate directly to harmonic relationships, and others do not. The concept of the *Stufe* is by nature applicable to every level of prolongation of a piece until the surface form is reached." (Keiler (1977): 12)

To draw an analogy here with the linguistic discussion we had in the previous section, it seems that Keiler is saying that Schenker's inconsistency is akin to labeling certain words or strings of words as grammatical constituents because they happen to be heads (such as a noun) or phrasal projections of heads (e.g. a noun phrase), but not to the words that modify these heads to generate larger structures – which in language *are* accorded constituent status, e.g. as complements, specifiers and adjuncts. In fact, to correct for this inconsistency, Keiler proposed a revision to Schenker's conception of the *Stufe* that

directly borrows ideas of constituency from generative linguistics. Example 1.2-27 (taken from Keiler (1977): 16) illustrates this revision.

In the upper half of the image, we see a representation of Schenker's *Ursatz*, the foundational I-V-I chord progression from which, according to Schenker, every tonal musical surface (i.e. every musical piece written by the 18th- and 19th-century masters Schenker admired) can be generated, through

Example 1.2-27. Keiler's binary-branching tree representation of the *Ursatz*



successive levels of voice-leading elaboration. (We saw an example of such a generation in the last chapter, in my description of Bellini's *Casta diva* aria in Example 1.1-7.) The Roman numerals "I" and "V" in the *Ursatz* representation in Example 1.2-27, are, respectively, the labels for the scale-degree 1 and 5 *Stufen*. As Keiler claims, Schenker inconsistently applied these labels only to the harmonies that make up the *Ursatz*, or other harmonies close to the harmonic backbone of a piece, before they have been modified by successive levels of voice-leading elaboration. In the lower half of the image, we see Keiler's linguistics-inspired reinterpretation of the *Ursatz*, in which the initial I *Stufe* is treated as a terminal node T in a tree representation of the *Ursatz*, which projects itself as the head of a I-phrase that Keiler calls TP or "Tonic Prolongation". The branching sister of TP is another I-phrase called TC or "Tonic Completion", whose head is the final structural, or cadential, tonic T of a piece. The intermediate, dominant, harmony D of Schenker's *Ursatz* heads its own phrasal projection, called DP or "Dominant Prolongation", which is treated as the final tonic's phrasal complement, whose mother node in the tree is the final tonic's phrasal projection TC.

As is clear from the example, Keiler treats Schenkerian *Stufen* as grammatical constituents of musical trees, and specifically as *binary-branching* constituents of a tree. This already shows an overlap in the way Keiler depicts tonal structure and how generative linguistics describes the structure of sentences in languages. Moreover, Keiler treats *Stufen* as the heads of larger phrasal projections, just as nouns and verbs act within sentences. This is why the final TC constituent is the phrasal projection of the final T, which is a constituent in itself, specifically the head of TC. These phrasal projections would then become larger constituents, formed by the prolongation of a *Stufe* by some other sonority – e.g. TC, which is itself formed by the prolongation of the *Stufe* T by DP. (Notice how through this model, Keiler also seems to be suggesting that chords – in the *Stufe* sense – constitute the lexicon of tonal music, which is an idea I have broached a few times in this dissertation now.)

Keeping with the above stipulation that such a constituency analysis be consistent, Keiler proceeds to analyze passages from a few pieces by Handel and Bach using the categories shown in Example 1.2-27, in which he labels not only the T, D, TP, TC, and DP constituents, but every other

harmony as a constituent too, with the help of Roman numerals – including those harmonies whose main job is to prolong structurally ‘deeper’ harmonies closer to the *Ursatz* (see Keiler (1977): 17-26). In so doing, Keiler not only treats harmonies as grammatical constituents, but also demonstrates how a harmonic constituent analysis of a piece reveals a hierarchical grammatical structure in tonal music akin to language, and one that helps bring a certain consistency to Schenker’s abstract *Stufe*-based approach to describing phrase structure in tonal music.

Keiler’s revision of Schenker’s *Stufe*-based approach to tonal phrase structure suggests an intriguing way of modeling the grammar of tonal music that both depicts constituency in musical structure (specifically a harmonic one, albeit in the Schenkerian sense), and also a hierarchical organization among musical constituents. Other scholars have sensed the suggestiveness of this way of modeling tonal grammar too over the years, even if they have not been directly influenced by, or even aware of, Keiler’s work. A good example is the recent model proposed by Martin Rohrmeier (2011), which uses similar harmonic constituency categories as Keiler proposes, but which are then combined into larger constituents through a series of phrase structure rules – the end result being a hierarchical model of tonal phrases, akin to the one shown in Example 1.2-27, albeit with a larger number of constituents, and therefore much more harmonic and grammatical detail.

There are a couple of things worth noticing about Keiler’s characterization of tonal grammar though. First, the labels that constituents have in his trees are not really *Stufe* labels – he does not label the nodes of his trees with scale-degree numbers such as 1 or 5 with caret symbols on top, which is the way scale-degrees are usually represented. Instead, Keiler uses a mixture of Roman numerals like I and V, and alphabets like T and D. Although Roman numerals are often used to depict Schenkerian *Stufen* in graphic analyses of musical passages, including by Schenker himself, alphabet symbols like T and D are not – especially since they are more often associated with harmonic functions, in the sense proposed by the 19th-century music theorist Hugo Riemann. So, Keiler’s labeling system reveals an inconsistent mixture of *Stufe* theory and function theory, despite his insistence on Schenkerian ideas in the founding of his

model. Unfortunately, this mixture of harmonic theories is a problem, since Schenkerian *Stufen* and Riemannian functions are not the same. *Stufen* represent grammatical constituents, and particularly heads as argued above, which is why they can form the terminal nodes of a tree, which are then joined together to form larger structures. However, harmonic functions seem to have more of a semantic role in grammar, they represent how one *interprets* grammatical tree relationships at S-structure. This is a point made by the linguists Jonah Katz and David Pesetsky in their critique of Rohrmeier's above model, which also makes use of harmonic function 'constituents' in generating musical phrases (Katz and Pesetsky (2011): 57-64). I cited Katz and Pesetsky's model in section 1.2.2 above, in our discussion of whether music has a logical structure, where Katz and Pesetsky's argument that harmonic functions are semantic allowed us to claim that such functions make up the logical structure of a musical passage – akin to how Fregean subjects and predicates are argued to make up the semantic, logical structure of linguistic constructions. (In this sense, the representation of a musical passage in terms of harmonic functions (such as “Tonic – Subdominant – Dominant – Tonic”) is essentially the semantic, Logical Form (LF) representation of the passage's S-structure; a point I will discuss in detail in chapter 2.1.)

I will return to the debate between Rohrmeier, and Katz and Pesetsky later in this subsection, where I will propose a more explicitly *Stufe*-based approach to generating musical phrases. But let us examine another shortcoming of Keiler's model first. Since his model starts off as a linguistics-inspired reinterpretation of Schenker's *Ursatz*, his trees share one important feature of the *Ursatz*, viz. they are *representations* of musical phrase structure, akin to how X-bar trees, with their “specifier/head/complement” structure also provide representations for linguistic phrase structure. (Rohrmeier's model, which has a lot in common with Keiler's approach, is not so explicitly representational, but it does generate phrases by means of a number of proposed musical PSRs. This means that each PSR generates a node in a grammatical tree, and the variety of nodes thus generated by Rohrmeier's PSRs give the appearance of a representational tree too when taken together – which need to be filled in *a posteriori* by PSR-based phrase generating operations.)

Both Schenker's *Ursatz* representation and X-bar theory's "specifier/head/complement" representation are problematic, as we have discussed at various points in this chapter. Particularly, Schenker's *a priori* assumption that all tonal passages should reveal an *Ursatz* structure leads to the problems of defining what a musical sentence is, especially if the passage revealing the *Ursatz* can be thought of as a poetic structure, such as a complete movement in a piece that is hundreds of measures long. Similarly, X-bar trees run into trouble when we cannot decide *a priori* how to fill in a representational category in a tree, as we saw in the issue of whether NPs are really DPs.

Not surprisingly, Keiler runs into similar issues with his representational model too. First of all, given that his representation of musical phrase structure is *Ursatz* based, it is not clear how Keiler defines a musical phrase or sentence. Even if we ignore this problem, there is a second issue of what the head of such a phrase is in Keiler's tree representation of it. A musical phrase, at least in the Western Classical tonal idiom, is certainly a projection of T (i.e. Tonic, not Tense as in linguistic TPs), as is implicit in Schenkerian theory, and also in Lerdahl and Jackendoff's (albeit, un-Schenkerian) model of tonal grammar. But in Keiler's model, there are two projections of T that make up a tonal phrase, viz. the initial T that projects TP, and the final T that projects TC. Which one is the head of the complete phrase? In other words, are musical phrases projections of an initial tonic, or a final, cadential tonic?

The final cadential tonic is usually accepted as the head of a musical tree, following the conventional belief that a tonal ending is more stable, and therefore hierarchically more superior, than the chord that represents a tonal beginning – tonal stability obviously being the main criterion for determining hierarchical superiority in a tree structure. Moreover, the final tonic, and the authentic cadence it ends, are supposed to play the most significant role in reinforcing the key of the phrase in which the cadence occurs. Allan Keiler discusses how this is central to Schenker's conception of phrase structure too – a point we shall return to in the next subsection, where we will also explore David Pesetsky's utilization of this idea to stress the importance of cadences in modeling musical grammar. However, the belief in the hierarchical superiority of the final tonic has not been subjected to further scrutiny. For example, Lerdahl and Jackendoff merely assert that "the ending of a piece is usually more stable than its beginning"

(Lerdahl and Jackendoff (1983): 137), without justifying this point any further. This is a problem for a representational model of tonal grammar, because without an adequate justification for why a tonal phrase should be the *final* tonic's phrasal projection there is no way to determine what the phrase's actual representation is, and what label it should be given in a tree diagram of this representation (i.e. at the position of the root node at the top of the tree, which represents the entire phrase).

This problem is evident in Keiler's tree representations of tonal structure too. In his first use of tree diagrams to depict musical structure (in Keiler (1977)), he does not make either the initial or the final tonic the head of a tree, labeling the head merely as "T" (for "tonic") – as you can see for yourself in Example 1.2-27. In a subsequent paper he initially adopts this strategy again, but then switches to labeling the *initial* tonic chord as the head of the entire tree structure (Keiler (1983-84), beginning with his figure 8). So, Keiler is inconsistent about how he labels the root node of a tonal phrase's complete tree. This is not a problem in and of itself – but it is for a representational model of musical grammar, where the root node's label should be given in the tree, and not something indeterminate. (This is similar to the problem within X-bar theory about the indeterminacy over whether an NP/DP is a projection of N or D.)

A final problem with Keiler's representational model lies in how it describes the phrase structure of actual musical passages. A representational model assumes that the various nodes of a tree, given *a priori* by the representation, can be filled in exhaustively by the constituents of an actual phrase. But what if an actual phrase lacks a constituent that would normally fill a node in the tree? This means the tree would have an empty node, and the representational model that proposes that specific tree for phrase structure will then have to justify the presence of such empty nodes in the tree, which on the basis of the structure of actual phrases seem to be *prima facie* non-existent. In other words, why propose a certain representation of phrase structure, when actual phrases seems to lack the categories proposed by the representation? This is the issue that arises in the way Keiler analyzes the phrase structure of a Bach chorale passage (Keiler (1977): 19-23). The chorale is in F major, and is marked by a descending bass line at the beginning with the notes F-E-E-flat-D, which are harmonized by the chords I, V⁶, V²/IV and IV⁶. This is followed by the bass notes C and F, which realize a C-major to F-major, V-I authentic

cadence – which by itself can easily be fitted into Keiler’s model as a TC phrase, in which the final F-major chord T is modified by a DP, whose head is the dominant C-major chord. But how do we represent the chords that realize the descending bass line? The first bass note F clearly represents another T, this time the head T of the initial TP phrase. Since the following E bass note realizes a first inversion, V^6 , C-major harmony, we can label this as a D too. But in Keiler’s model, D heads a phrasal projection DP that subsequently modifies T, as we just saw in the case of the authentic cadence that ends this passage. This means that a tonic chord of some kind, which heads a TC phrase, should also follow the V^6 of the descending bass line – but instead we get a V^2/IV chord. This V^2/IV chord might seem to be a tonic F-major chord (because it has the notes F, A, and C in it), but it really acts as a dominant to IV, i.e. B-flat major, because of the E-flat that is also in this chord, in the bass. Moreover, a first inversion B-flat-major chord, i.e. IV^6 , follows this V^2/IV chord too anyway, which suggests that the V^2/IV and the subsequent IV^6 chord both form a TC phrase in themselves – which leaves the preceding V^6 without a subsequent tonic chord to modify, and thus fill in a TC node in Keiler’s tree.

In other words, Keiler’s representational model requires a dominant chord to be followed by a tonic chord (to generate a TC phrase) – which means that a ‘Keiler-ian’ tree will have an empty tonic node in examples like the above Bach passage, where a dominant chord is not followed by the expected tonic chord, but by a V^2/IV chord instead. To fit Keiler’s model, the descending bass line should have been something like F, E, F, E-flat, D – with the second F being realized as the missing tonic chord. However, the bass line obviously does not have this F, and does not realize a tonic chord at that point. So, the question is, why propose an (empty) tonic category in a model of phrase structure, when actual musical passages like the Bach one lack this tonic? Note that this is not a problem for generative models of musical grammar per se, but only for *representational* ones, which assume *a priori* that phrases will have certain tree architectures, filled out by specific constituents at specific nodes in the tree. Given that Keiler’s model is representational, he is forced to explain his way out of the above problem, which he does by saying that the V^6 chord in the Bach passage is not a true dominant chord, but more of a voice-leading, sonority, which effects a passing, voice-leading motion, inherent in the descending bass line. As

a result, the V^6 does not have to be followed by the expected tonic. But since even voice-leading sonorities have to be treated as *Stufen*, in the interests of consistency – something Keiler criticized Schenker for not doing – the V^6 is placed at a node representing a dominant D *Stufe* anyway, even though this D ends up modifying an empty tonic category subsequently. Moreover, Keiler claims that the chord that actually does follow the V^6 , viz. the V^2/IV , has both tonic properties (in its pitch structure), and voice-leading properties (in the fact that its bass note E-flat continues the descending, passing motion of the bass line). In consequence, Keiler allows it to occupy the empty tonic category and the following dominant of IV^6 category simultaneously, which he does by blending into one the two branches that represent these categories in the phrase's tree – which seems like an unwarranted modification of the tree just to fix its inability to adequately represent the actual structure of the passage. (In fact, rather than blending the two branches together, separate branches would have been warranted here, one with the tonic chord that is missing from the passage, followed by one for the V^2/IV chord, because the former chord *prepares* the latter – a standard phenomenon in dominant seventh chord preparations.)

As we saw in our discussion of C_{HL} in the last section, a derivational approach to phrase structure was adopted by the Minimalist Program to account for some of the shortcomings of the earlier representational approach of X-bar theory. Might such an approach be appropriate here too, in the context of *Stufe*-based musical phrase generation? In fact, I believe Schenker intuitively sensed the advantage of this approach himself (without explicitly proposing one), which is why he was reluctant to propose a *Stufe*-based representational structure for an entire musical phrase, all the way up to the surface – although it is for this very reason that Keiler, ironically, finds Schenker's way of assigning *Stufen* (i.e. only to constituents in, or close to, the *Ursatz*) as being inconsistent. Consider Keiler's own words about Schenker's reluctance to propose a representational model of tonal passages in this regard:

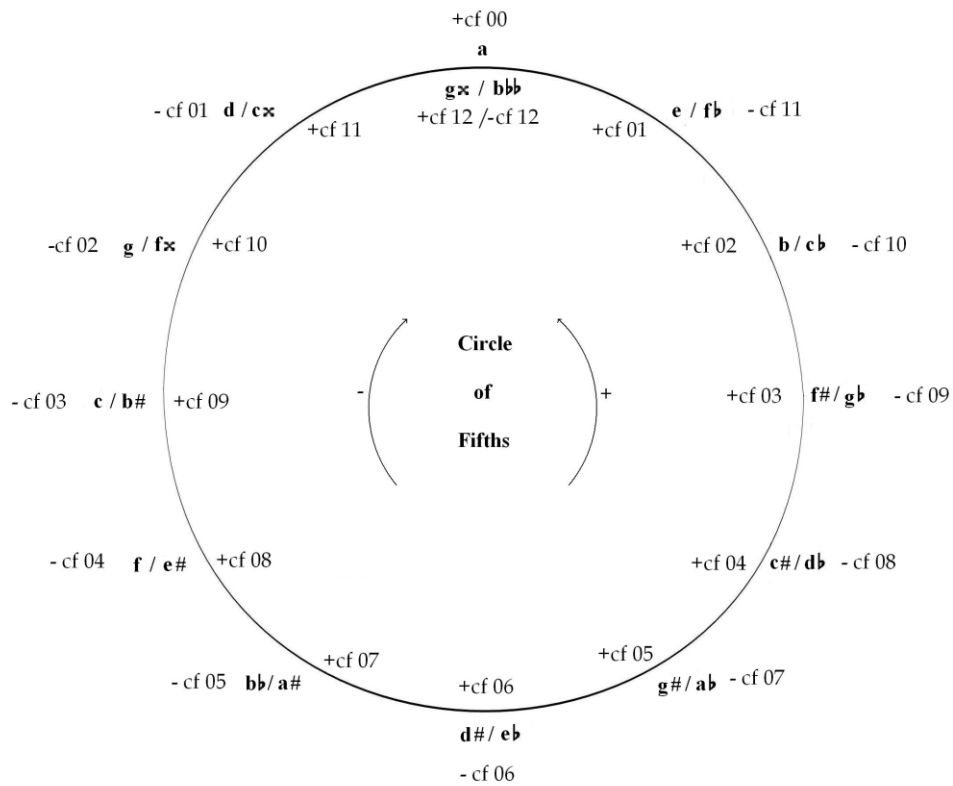
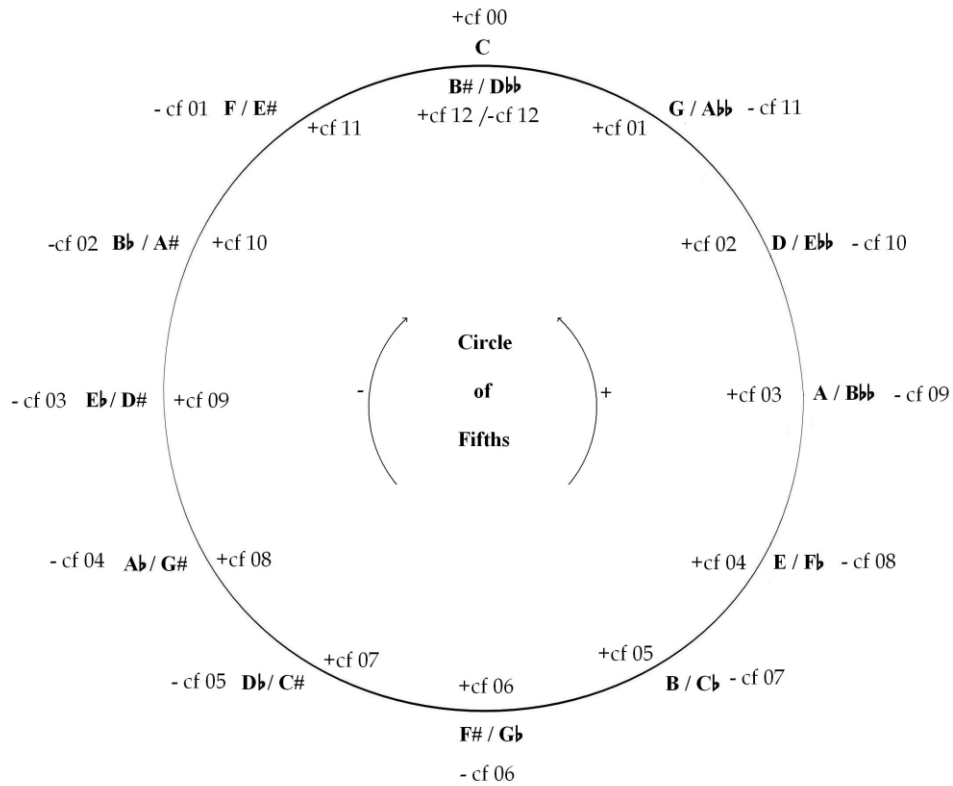
“He realized that it would be impossible to create procedural definitions that would apply appropriately under all possible conditions, since that would in effect require that all possible applications and realizations of an abstract concept [viz. that of the *Stufe*] be built into its definition.” (Keiler (1977): 12)

So, a derivational (and possibly Minimalist) approach might actually be the most Schenkerian approach to describing musical phrase structure too, at least in spirit. How might such an approach work? Well, we can adopt Keiler's attitude that *Stufen* are the grammatical constituents of a musical phrase. But rather than proposing an *a priori* representation of what such a phrase should look like, we can just propose a Merge-based approach instead, in which various *Stufen* are merged into two-member sets, with one member (the one whose lexical information projects further up the tree) being taken as the head member of the set. Moreover, this would not assume the presence of any representational categories, such as Keiler's TP and TC, although the binary-branching nature of those categories will be realized through the set-theoretic workings of Merge. This would be more akin to a Bare Phrase Structure approach to musical phrases too.

Of course, as the previous paragraph implies, this approach requires a definition of what a lexical item is in music, and what amounts to the lexical information in such items. I will just assume from the discussion earlier in this chapter that chords, and particularly *Stufen*, constitute a lexicon for tonal music. What kind of 'lexical' information do *Stufen* have then, which allows them to merge, and which they further project into the more complex constituents of a musical phrase? Well, for one, this information is not of the tonic-dominant sort, since that lies in the interpretive side of things, as suggested before. (That is, *Stufen* do not project "tonic" features or "dominant" features.) However, by definition, they project *scale-degree features* (since "*Stufe*" is, after all, German for scale-degree).

A scale-degree feature represents a *Stufe*'s characteristics within an abstract musical pitch space. When *Stufen* are close to each other in such a pitch space, their scale-degree features can be said to agree, in the same way a noun and a verb's features agree when they are in a certain close configuration – which therefore allows them to merge into a larger structure. So proximity in a given pitch space can be considered a locality constraint on phrase generation, of the kind we explored for language in the previous section. To understand this better, consider Example 1.2-28, which depicts the scale-degree features *Stufen* have within an abstract pitch space called the "circle of fifths". This is a very rudimentary

Example 1.2-28. Major/minor triadic relationships represented as “Circle of Fifths” features



(although important) pitch space in tonal music – so a more sophisticated description of the ‘lexical’ features possessed by *Stufen* would have to look at how *Stufen* relate to each other in other kinds of pitch spaces too.⁵⁵

In the Western Classical tonal idiom, *Stufen* are realized by triads, as we have discussed before, so Example 1.2-28 presents two different circles of fifths here, one for *Stufen* realized by major triads (shown in the upper circle), and one for *Stufen* realized by minor triads (shown in the lower circle). For ease of explanation, Example 1.2-28 represents *Stufen* as actual pitch classes, so the upper circle of fifths in the example takes C to be scale degree 1, and the lower circle takes A to be scale degree 1 – although it is important to remember that *Stufen* are abstract entities that can be realized by a variety of pitches. You will notice that every major triad in the upper circle has a corresponding minor triad in the lower circle with the same “cf” value (i.e. “circle of fifths” value – an idea I will explain very soon). For example, C major and A minor share the same cf value of +cf 00. This just reflects the fact that the scale degrees represented in the upper and lower circles are the same – they just happen to be realized by different, major and minor, pitch class triads, specifically triads in a *relative* relation with each other (C major being the relative major of A minor, and A minor being the relative minor of C major). *Relative* major and minor triads therefore share the same scale degree properties within their respective, i.e. major and minor, circles of fifths, which results from their sharing the same scales and pitch information. This means that the two circles in Example 1.2-28 can be superimposed on each other. (Given this isomorphism between the two circles, I will just focus on the upper, major, circle in the subsequent discussion.)

As might be obvious now, pitch classes representing *Stufen* a fifth apart are adjacent to each other in the circle of fifths, and chords built on them can therefore combine to form a variety of musical phrases in Western tonal music (in light of the locality constraint on phrase generation just mentioned). In the upper circle of Example 1.2-28, C and G are a fifth apart, and therefore adjacent to each other in the

⁵⁵ In fact, Fred Lerdahl’s *Tonal Pitch Space* (Lerdahl (2001)) does exactly that, which is why I suggested early in this chapter that a text like that could be the basis for a deeper theory of what the musical lexicon is. (Although it should be noted that Lerdahl does not describe pitch-space relationships in his text in terms of lexical features, primarily because a linguistic focus is not the emphasis of that text.)

circle, so chords built on them can combine to form a larger musical phrase too – in particular, a G - C harmonic phrase, in which the harmonic functions of the chords are interpreted as a V - I progression, which is one of the most fundamental functional relationships in tonal harmony (in fact, it is just Keiler's TC constituent). (Note again here how harmonic functions are not the same as scale degrees, since the latter give rise, grammatically, to larger musical phrases, which are then interpreted, semantically, in harmonic functional terms.) Like C and G, F# and C# are also a fifth apart, so chords built on them can also form a larger phrase when merged. However, C and F# are not adjacent in the circle, since they are not a fifth apart, and so chords built on them would not normally merge to form a larger phrase.

There is a problem though, with this description of the characteristics of *Stufen* in the circle of fifths. It presents these characteristics as properties within an actual circle, of the kind visualized in Example 1.2-28. But *Stufen* are abstract entities, and so do not inhabit actual circles anywhere. The circles in Example 1.2-28 are just visual representations of an abstract relationship between *Stufen*, presented thus for clarity of exposition. So we need to be able to describe a *Stufen*'s characteristics within the circle of fifths in more abstract terms, akin to the abstract features possessed by lexical items in language. In other words, we need to be able to describe a *Stufen*'s properties in terms of abstract features that can be checked when *Stufen* are in a certain local configuration (which we might represent visually as adjacency in the circle of fifths), which then allows grammatical musical phrases to be generated from them. This is akin to the lexical features possessed by words that are also checked when words are in a certain local configuration, which then allows grammatical linguistic structures to be generated from such words.

This is where I would like to invoke the aforementioned notion of a scale-degree feature – an abstract property possessed by *Stufen* that can be checked when *Stufen* are in a certain close configuration, which can be represented visually in a given musical pitch space. I would like to talk specifically about the scale-degree feature called the “cf-feature” or simply “cf”, which is the particular scale-degree feature that is checked when *Stufen* are in a certain local configuration within the circle of fifths pitch space, and which can take values of the kind I briefly mentioned above. So, the idea here is that if *Stufen* have certain ‘checkable’ cf-features, then they can be legitimately merged. For example, the

pitch classes C and G in the upper circle of Example 1.2-28 (which are being taken to represent the scale degrees 1 and 5 here) have cf-feature values of +cf 00 and +cf 01 respectively, which have a difference in value of 1 – the smallest *real number* difference possible between two cf values. (I will explain the + prefix on these cf values in a moment.) This similarity in values is what we were previously representing, visually, as adjacency in the circle of fifths. So, the similarity of their cf values implies that the *Stufen* represented by the pitch classes C and G are in a certain local configuration in an abstract pitch space, which allows chords built on C and G to merge. In this manner, F# and C# have cf values of +cf 06 and +cf 07 respectively, which also differ by a real number value of 1 – and hence, chords built on these two pitch classes can merge to form larger structures too. However, C and F# have cf values of +cf 00 and +cf 06 respectively, as we just saw, which actually puts them at opposite ends of the circle, so chords built on them will not normally merge to form a larger phrase.

But having similar cf values is not enough for *Stufen* to merge. *Stufen* have to merge in the right order as well. This is similar to the way words must merge in the right order for the phrase resulting from their merger to be correctly *linearized*. For example, in English, an object NP comes *after* the predicate verb when the two are merged to form a VP. To understand this in the case of music, consider the fact that the *Stufen* represented by the pitch classes C and G can merge in four different orders:

- (a) G then C, with C as the head of the resulting set (as in a V-I progression in C major).
- (b) C then G, with C as the head of the resulting set (as in a I-V progression in C major).
- (c) G then C, with G as the head of the resulting set (as in a I-IV progression in G major).
- (d) C then G, with G as the head of the resulting set (as in a IV-I progression in G major).

I will illustrate in the next subsection how (b) and (c) are really the result of movement transformations, meaning that neither of these sets are formed by merging ‘lexical’ chords to form D-structures. Instead, they are actually S-structure transformations of certain other D-structures – (a) and (d) respectively, to be precise. Now, the ‘plagal’ progression represented by (d) is actually a legitimate D-structure in another idiom, viz. Rock music (as I will show in a later subsection too) – but I will claim that it is actually

illegitimate in the Western Classical tonal idiom, especially since it doesn't seem to fit with the rest of Western Classical tonal harmony.⁵⁶ Therefore, only (a) is a legitimate D-structure, or part of a legitimate D-structure, in the Western Classical tonal idiom. (It actually represents the canonical “descending fifths” progression of Western tonality.)

So, when chords built on the *Stufen* represented by the pitch classes C and G merge, they have to merge in a way such that the G-chord comes first. We can discuss this phenomenon visually, as the requirement that *Stufen* merge in the right *direction* in the circle of fifths, i.e. *from G to C*. But if you look at the top circle in Example 1.2-28 again, you will notice that there are two ways in which this directional requirement can be satisfied – either by moving anti-clockwise from G to the adjacent C position, or clockwise all the way around the circle to the C position. However, the latter option would also make G and C non-adjacent, so there should be a way to disallow this ‘directionality’ in the merger of G and C.⁵⁷

Actually, this directionality is automatically disallowed if you examine the fine structure of the circles in Example 1.2-28. If you examine this fine structure, you will notice that multiple pitch classes realize each position in a circle. For example, the position occupied by C at the top of the circle is also

⁵⁶ This is because appearances of this progression really arise from either (a) voice leading, or (b) borrowings from another idiom into tonal harmony. Regarding (a), this applies especially to progressions involving an apparently IV sonority in *inversion*. Two common examples are the neighboring or pedal 6/4 use of IV (e.g. I - IV^{6/4} - I) – which is really an example of neighboring 6/4 voice leading (Aldwell and Schachter (2011): 350) – and the use of IV⁶ to expand tonic harmony through a descending bass motion, as seen in the progression I - IV⁶ - I⁶. Regarding (b), first consider that a chord built on scale degree 4 is normally interpreted as having subdominant or predominant function – meaning that it normally proceeds to a dominant-functioning harmony. But in (d) above, IV (which represents the same predominant function) is proceeding to I, or tonic-functioning harmony. This is certainly not how the other common predominant-functioning harmony, i.e. the chord built on scale degree 2 (the supertonic), functions in Western tonality – this chord always proceeds to dominant-functioning harmony, and never to the tonic (at least in functional harmonic, as opposed to voice leading, progressions). So, IV-I as a functional progression is quite odd. Moreover, the supertonic was the predominant-functioning harmony of choice in the earlier years of tonal harmony, particularly in the hands of masters like Haydn, Mozart and Beethoven, and it arises from an earlier contrapuntal use in the Renaissance too (e.g. see Gauldin (1995): 138-139). So, the way the supertonic functions in harmonic progressions should be taken as the yardstick for how predominant-functioning sonorities should behave in Western tonality – making the IV-I progression of (d) doubly odd. Finally, the use of IV as a common harmonic function really occurred in the 19th century, in the hands of the Romantic composers, many of whom were interested in folk music and musical nationalism. Which means that their use of IV could very well have been borrowed from these other idioms – which are, after all, the same vernacular idioms from which Rock music evolved.

⁵⁷ In fact, in this, clockwise, direction the pitch class adjacent to C is F. So, if we allow the merger of *Stufen* from this direction, we have to allow the combination {F, C} given that it involves two adjacent *Stufen*. However, {F, C} can be interpreted either as a IV-I functional constituent in C major, or a I-V constituent in F major – and these are both constituents that arise from movement transformations as I suggested above, and so cannot be allowed as legitimate D-structures. So, the merger of *Stufen* from a clockwise direction in the circle of fifths cannot be allowed, at all, at least in the generation of phrases in the Western tonal idiom.

occupied by B# and D-double-flat, and the position to the right of this is not only occupied by G, but also by A-double-flat. This is obviously a result of the phenomenon known as “enharmonic equivalence”, which can be seen in equal-tempered systems like the modern piano, where the notes C, B# and D-double-flat – and G and A-double-flat – are represented by the same key on the keyboard. We also see the phenomenon of “mod-12 octave equivalence” here, where after every twelve steps of the circle we are assumed to return to the same pitch class, albeit an octave higher or lower depending on which direction of the circle one travels in. This is why B# and D-double-flat, whose cf values are +12 and -12 respectively, are placed in the same position in the circle as C.

But even though many of the pitch classes in the circle share the same positions in the circle, their enharmonic or octave equivalence is *not* assumed. The reason I have positioned many pitch classes in the same spot in the circle is mainly to make the circle easier to read. This makes two pitch classes positioned in the same spot apparently equivalent too. But notice that no two pitch classes in the circle share the same cf value, which means that they *could* have been assigned a unique position on the circle, if visual clarity was not an issue – which implies that their enharmonic or octave equivalence is not necessarily assumed in the circle. In this light, there is actually only one path you could travel along the circle to merge G with C, and that is the correct, anti-clockwise one. If you were to travel clockwise from G, you would actually never meet the pitch class C, if we do not assume enharmonic or octave equivalence – which we do not have to anyway, as just argued. If we do not assume enharmonic or octave equivalence, traveling clockwise from G, via D, then A, then E and onwards – would actually take us to the pitch class B#. So, the incorrect, clockwise merger of G and C would never happen.

However, the model should work even if we do assume enharmonic or octave equivalence. In other words, it would not be a great model of musical grammar if assuming enharmonic or octave equivalence makes the model generate ungrammatical structures. So, the correct direction of merger of *Stufen* in the generation of musical phrase structures has to be accounted for by C_{HM} explicitly. Ideally, this should happen via a scale-degree feature too, since the direction in which we move in the circle of fifths is just a visual representation of a more abstract property of *Stufen* and of musical grammar. I have

discussed how the correct direction of merger for G and C involves anticlockwise motion in Example 1.2-28. But this is just the particular visual representation I have chosen for this particular example – I could have represented this as a clockwise motion too, with G in the opposite position in the circle, i.e. in the position currently occupied by F. So, the way we have been talking about directionality depends on the arbitrary frame of reference chosen. This is of course of no use to two *Stufen* that need to find a way to check their features, independently of how they are represented in a circle on a page in a dissertation project. This is why the order in which *Stufen* merge should be incorporated in the feature structure of *Stufen* as well.

One way to deal with this could be to require that when two *Stufen* merge, in addition to having close cf values, the merger should be from the higher cf value to the lower one. This would ensure that G merges with C, rather than the other way around, because G has the higher cf value of 01, compared to C's value of cf 00. But the pitch class F has an absolute cf value of 01 too – which means that the above stipulation would license the combination {F, C}, even though this is illegitimate, as I argued earlier.

This is where the + and - prefixes to a *Stufen*'s cf value come to the rescue. Even though G and F share the same absolute, or natural number, cf value of 01, F's *integer* cf value is actually -cf 01, which makes it smaller in value than C. This means that if F and C merge, the direction will be *from C to F*, which is the correct order (since this also represents a V-I functional progression, in the way a G-C progression would). So, the + and - prefixes ensure that merger will be from higher to lower cf value, something the numerical component of a cf value cannot do by itself. In this way, the complete cf value, with both its + or - prefix and its numerical suffix, ensures, 'lexically', both the merger of *Stufen* that are adjacent on the circle of fifths, and in an anti-clockwise direction of motion along the (or rather along this particular) circle of fifths.

In Example 1.2-28, the curving arrows in the middle of the two circles represent the directionality component of a musical Merge operation. As we can see in the upper circle, to merge two chords built on scale degrees a fifth apart, like C and G, we have to move anti-clockwise in the circle – as shown by the

curving arrow in the middle-right of this circle with the + sign next to it. The curving arrow in the middle-left of the image, with the - sign next to it, will generate phrases in clockwise fashion, which would be ungrammatical within Western tonality. However, I suggested earlier that this order *does* generate grammatical phrases in the idiom of Rock music – a point I will develop later in subsection 1.2.4.iii.

Note that I have been representing *Stufen* in Example 1.2-28 in terms of actual pitch classes, primarily for ease of reading the example. But I should point out again that *Stufen* are not pitch classes themselves, but are only manifested by the latter. As a result, the scale-degree features I have been discussing so far are *not* pitch class features – the ‘lexical’ information contained within the abstract structure of a *Stufe* is not pitch class information. The reason for this might be obvious by now – *Stufen* are grammatical constituents, as per Allan Keiler’s proposal above; but pitch classes are abstract, *acoustical* structures, which represent the similarity between individual pitches that share the same fundamental in the overtone series. There have been attempts to develop theories of musical phrase structure on acoustical grounds (such as Parncutt (1989)), but just taking pitch classes as constituents will not help our Minimalist generative grammatical cause here. For one, if we try to merge the pitch classes C and G (as opposed to the *Stufen* they represent), how do we know whether that they can be merged to begin with, without some sort of adjacency information to license this merger (which is provided by a *Stufe*’s circle of fifths feature structure)? Moreover, how would we know what order the two pitch class constituents should be merged in, and what would be the head of the set that results from this merger, without some sort of directionality information to license these aspects of the merger (and which are provided by the + and - prefixes in a *Stufe*’s feature structure)?

Example 1.2-29 reinforces the above point. Here we see the scale degree information of Example 1.2-28 again, but presented this time with reference to actual *Stufen*, rather than through pitch classes that realize these *Stufen* in the way Example 1.2-28 did. The top row of the example just lays out the 12 *Stufen* that make up the scale degrees of the chromatic scale in Western tonality, represented here in the conventional way with numbers and caret symbols. The middle and bottom rows lay out the cf values

Example 1.2-29. Scale degrees represented as “Circle of Fifths” features

$\hat{1}$	$\hat{b2}$	$\hat{2}$	$\hat{b3}$	$\hat{3}$	$\hat{4}$	$\hat{5}$	$\hat{b6}$	$\hat{6}$	$\hat{b7}$	$\hat{7}$
+cf 00	+cf 07	+cf 02	+cf 09	+cf 04	+cf 11	+cf 06	+cf 08	+cf 03	+cf 10	+cf 05
-cf 12	-cf 05	-cf 10	-cf 03	-cf 08	-cf 01	-cf 06	-cf 04	-cf 09	-cf 02	-cf 07

associated with these *Stufen*, as depicted in the previous example too. If you look at the middle of Example 1.2-29, in the column for scale degree 5, you will notice that it has cf values of +cf 01 and -cf 11. Looking back at Example 1.2-28, you will also notice that these were the values given to the pitch classes G and A-double-flat respectively. The fact that both values are represented in the column of scale degree 5 in Example 1.2-29 implies that G and A-double-flat are being treated as enharmonically equivalent, with both pitches realizing the scale degree 5 *Stufe*. But, again, enharmonic equivalence is assumed here just for the sake of visual clarity. So, Example 1.2-29 could have ascribed the two above cf values to two different *Stufe* – specifically +cf 01 for the scale degree 5 *Stufe*, and -cf 11 for the scale degree double-flat 6 *Stufe*. I did not include a double-flat 6 *Stufe* in the example for simplicity and visual clarity, but it is important to note that we could have done so if we did not want to assume enharmonic equivalence – which the example does not commit us to anyway. It is also important to note that though the +cf 01 value ascribed to the scale degree 5 *Stufe* here was realized by the G pitch class in the previous example, it could be realized by any pitch class whatsoever, since pitch classes are just realizations of abstract *Stufen*, and pitch class information is not scale degree information.⁵⁸ So, we could ascribe the +cf 01 value to the pitch class D too, in which case *this* pitch class would be the realization of the scale degree 5 *Stufe*, which would happen, for example, in a piece in G major or minor.

There is a final property of *Stufen* worth noting here. I just said that, in a G major context, the pitch class D would be the specific realization of the scale degree 5 *Stufe*. But just because D realizes this particular *Stufe* in G major, this does not automatically mean that a D chord has *dominant function* in G major. First of all dominant function can be realized by other chords, such as the leading-tone F# diminished triad in G major. But more importantly, Schenkerian *Stufe* are *not* harmonic functions – if a D chord does have dominant function in G major, it is because this is how it was *semantically* interpreted at S-structure, not because it realized the *grammatical* constituent that is the scale degree 5 *Stufe*. (I have

⁵⁸ Of course, the sharp and flat prefixes attached to the scale degree numbers in Example 1.2-29 could be realized by pitch classes that do not necessarily share these prefixes. So, if one chose E-natural as scale degree 1 (to discuss a piece in, say, E minor) then the flat-2 scale degree would actually be realized by F-natural and not F-flat, and the natural-2 scale degree by F-sharp and not F-natural.

made this point earlier, and will develop it a bit more later.) What function a chord has, as opposed to what *Stufe* it realizes, is therefore a semantic phenomenon. And a particularly idiom-specific one too – if the pitch class D were to realize a scale degree 2 *Stufe*, the “supertonic” *Stufe*, it would do so across idioms, as long as the pitch space we are working in is centered around the pitch class C (in which D is the supertonic). However, what harmonic function this supertonic *Stufe*, realized by the D, has, varies across idioms. In Western Classical tonal music, the supertonic is normally interpreted at S-structure as having “predominant” function, primarily because it is normally followed by a dominant-functioning harmony. But in the idiom of Rock music, a chord that realizes a scale degree 4 *Stufe* is what normally follows the supertonic. So supertonic sonorities in this idiom should be interpreted as having “pre-subdominant” function – a fact I will examine more in the following subsection 1.2.4.iii of this chapter. So, to summarize:

Pitch class \neq *Stufe* \neq Harmonic function

Keeping the above characterization of *Stufen* in mind, let us now see how this might help us generate an actual musical S-structure. That is, let us try to develop a Merge-based, derivational approach to generating musical phrases that takes *Stufen* as the constituents to be merged.

I will pursue this goal by attempting to generate a phrase from the hymn “Tebe poem” by the 18th century Ukrainian composer Dmitri Bortniansky. The reason for choosing a phrase from this piece lies in the fact that both Martin Rohrmeier (2011), and Jonah Katz and David Pesetsky (2011), have discussed it in their respective generative models of musical grammar – so we can compare the model I am about to propose to theirs subsequently.

Example 1.2-30 depicts my Merge-based derivation of a phrase from *Tebe poem*. The derivational tree of the phrase takes up the majority of the example. This tree has a number of nodes, which have been numbered from 1 to 9 for convenience of description. The stave system at the very bottom of the example presents the score for this phrase, which begins and ends in root position C-major

Example 1.2-30. Merge-based derivation of a phrase from Bortniansky's *Tebe poem*

The diagram illustrates the merge-based derivation of a musical phrase. The root node is $+cf00$ ⑨, which branches into $(TP) +cf00$ ⑧ and $+cf00$ ⑦. Node ⑧ further branches into $+cf01$ ⑤ and $+cf00$ ③. Node ⑤ branches into $-cf01$ ① and $+cf01$ ②. Node ⑦ branches into $+cf00$ ③ and $(DP) +cf01$ ④. Node ④ branches into $+cf01$ and $+cf00$. Node ③ branches into $+cf00$ and $+cf06$. Node ① branches into $+cf00$ and $-cf01$. Node ② branches into $+cf02$ and $+cf01$.

Below the tree is a musical score in 4/4 time. The top staff shows a melodic line with notes G⁴, F^{#4}, E⁴, D⁴, C⁴, B³, A³, G³, F³, E³, D³, C³. The bottom staff shows a piano accompaniment with chords. Chord symbols are: C, C⁵, F, D⁵, G, E⁵, Am, F^{#o7}, G⁴⁻³, C. Fingerings are indicated: 3 for G⁴, 2 for F^{#4}, 1 for E⁴. The piano part has a steady eighth-note accompaniment in the right hand and a more complex accompaniment in the left hand.

harmony. The staff system above that presents a mild reduction of this score, essentially to help us focus on the basic *Stufe* structure of this phrase, from which the phrase will be subsequently derived. This is accomplished by reducing all repeated chords in the score at the bottom (like the one that is played three times on the first two beats of the first bar) to one whole-note triad or seventh chord in the reduction. Each whole-note sonority in the reduction thus represents a *Stufe*, which, in turn, is manifested by the bracketed chords underneath it, in the bottom staff system. By representing the bracketed chords in the reduction as whole notes *without* barlines we can also focus on just their harmonic structure without worrying about the rhythmic aspects of the phrase.

You will notice that some of the chords within a bracket have notes that do not belong to the *Stufe*. For example, the chord on the third beat of the first bar has an F-natural note in its top voice, which does not belong to the C-major scale degree 1 *Stufe* these bracketed chords represent. But this F-natural is of course a passing tone, an example of an elaboration or “figuration” of a, in this case top, voice (Aldwell and Schachter (2011): 371-388). Such figurations are non-chord tones, which therefore have no grammatical status within a harmonic progression, and do not reflect how *Stufen* combine to form phrases. So, all such figurations in the score have been ignored, and thus omitted, from the reduction above.

Another thing the reduction helps us accomplish is that it allows us to examine the voice leading between the harmonies in this phrase. This is important because the voice leading of the phrase reveals some of its rather important structural features, which could – and should – play a role in a description of how the phrase is grammatically generated. So before we proceed to derive the phrase in a Merge-based way, with the help of cf features, we should discuss some of these voice-leading aspects of the phrase first.

Probably the most important voice-leading aspect of the phrase is its fundamental melodic line, i.e. its *Urfinie*, and the fundamental structure this is part of, i.e. the phrase’s *Ursatz*. If you look at the reduction closely, you will notice that the top voice E4 of the first chord, the root position C-major triad, leads to the top voice C4 of the final C-major triad, via the D4 top voice of the penultimate G-major triad,

as represented by the long horizontal beam in the image that connects these pitches (i.e. scale degrees 3, 2 and 1). This, then, is the *Urlinie* of the phrase, with the *Ursatz* of the phrase being made up of the initial C-major, the penultimate G-major, and the final C-major chords that harmonize this line. What the voice leading of the phrase also shows is that the initial scale degree 3 of the *Urlinie*, called the *Kopfton* as discussed before, is sustained in the *Urlinie* from the initial C-major triad to the A-minor triad in measure 6, as represented by the long slur that connects these two pitches. This suggests that the A-minor triad of measure 6 is a *right-branching prolongation* of the initial C-major triad, which helps to prolong the *Kopfton* in the *Urlinie* before its descent to scale degree 1 in the last two measures. This C-major to A-minor, I-vi, progression is actually quite a common progression in tonal music, and a standard way of prolonging initial tonic harmony and scale degree 3 in the *Urlinie* (e.g. see Aldwell and Schachter (2011): 195).

What this suggests is that the initial C-major triad's grammatical connection is not to the first inversion C^7 (i.e. $C^{6/5}$) seventh chord that follows it, but rather to the non-adjacent A-minor triad that follows it at the long distance of six measures. This is why the initial C-major triad's first branching sister in the tree above the reduction in Example 1.2-30 is the A-minor triad, which you can see high up in the tree at its second highest node, viz. node 8. Also notice that this node therefore represents the prolongation of the initial tonic before the final descent of the *Kopfton* – in other words, it represents the constituent that would be interpreted at S-structure as Keiler's "Tonic Prolongation", hence the (TP) label next to it.

Following the A-minor triad are four sonorities – an F# diminished seventh chord ($F\#^{o7}$), what appears to be a second inversion C-major triad, a root position G-major triad and the final C-major triad. The second of these two sonorities, the apparent second inversion C-triad, is just a leftward voice-leading elaboration of the subsequent G-major triad, called a "cadential 6-4". (We explored this sonority in the previous chapter, when analyzing a passage from Mozart's *Sinfonia Concertante*, K. 364.) In this light, the cadential 6-4 is not a harmonic entity unto itself, and therefore is not associated with a *Stufe* by itself either. If anything, it is part of the subsequent G-major triad's *Stufe*, which it elaborates at a level prior to

when *Stufen* enter into grammatical operations that generate phrases, akin to the morphological alteration of a word with certain affixes prior to the word's being merged with other words to form phrases and sentences. (This implies a certain connection between voice leading in music and morphological operations in language, which I believe is worth exploring further – but which shall not be taken up here in the interests of time.) It for this reason that the cadential 6-4, rather than being a true second inversion C-major triad, is taken here as part of the following G-major triad itself, as represented by the $G^{8/6/4-7/5/3}$ label these two sonorities are given – and which is the conventional label for cadential 6-4s in music theory anyway (e.g. see Aldwell and Schachter (2011): 181). It is for this reason that both these sonorities are given the joint cf value of +cf 01 in Example 1.2-30 too.

Prior to the cadential 6-4 is the diminished seventh $F\#^{07}$ chord. Unlike the cadential 6-4, this chord is not just a voice leading elaboration of the G-major triad, since it can progress, harmonically, to other chords – such as D-flat major – and is not tied to G major in the way the above cadential 6-4 is. However, since the *Tebe poem* phrase under consideration is in C major, at S-structure this $F\#^{07}$ chord would be interpreted as having a *specific* harmonic function, viz. as an applied (or secondary) leading-tone seventh chord to the dominant of C major, i.e. a vii^{07} of V. But this interpretation can only arise if the grammar makes the $F\#^{07}$ chord a left-branching sister to the dominant-functioning G-major triad (or to the whole [cadential 6-4 – G-major triad] complex). So, grammatically, the $F\#^{07}$ and the subsequent, elaborated G-major harmony must merge to generate a phrase, in which the G-major triad is the head and the $F\#^{07}$ its left-branching complement. Interestingly, this phrase would be interpreted at S-structure as a dominant prolongation, or Keiler's DP – as shown by the label at node 4 where the $F\#^{07}$ and G-major chords merge. Moreover, the head of this DP, the G-major triad, and the following root position C-major triad, are part of the *Ursatz* of this phrase – in fact they merge to form the second half of the *Ursatz*, the part that comes after the *Kopftön* begins its descent to the final structural tonic in the *Urlinie*. This part is nothing but the constituent that would be interpreted at S-structure as Keiler's "Tonic Completion" – and you can see this at node 6 in the tree, with its (TC) label too. The head of this TC constituent will have to be the final C-major triad – since this node arises from the merger of the C-major *triad* with a DP *phrasal*

constituent, and we know that heads cannot be phrasal. This explains why the TC constituent is called that, and not a Dominant Completion constituent.

In this manner, all the representational aspects of Keiler's model can arise, as by-products, within a generative framework that does not assume them *a priori*. This already suggests that the model we are pursuing here has the derivational flavor of a Merge-based model, as opposed to a representational X-bar-type model of grammar. But before we see how this model actually works, by using of *Stufen* features, we should look at one final voice-leading aspect of the *Tebe poem* phrase, which has to do with all the remaining sonorities in the phrase, all of which occur between the initial C-major triad, and the A-minor triad in measure 6 that right-prolongs this C-major triad to yield the Tonic Prolongation part of the phrase's S-structure.

Three of these sonorities are first inversion seventh chords, i.e. the $C^{6/5}$, the $D^{6/5}$, and the $E^{6/5}$. These chords are all independent harmonic entities, like the $F\#^{07}$ chord we just discussed – but like that chord, they all have a voice-leading component to them too because of the dissonant intervals of a seventh and a tritone they all contain. These dissonances have to be resolved in a subsequent chord according to certain regulations on dissonance treatment, which itself comes under the purview of voice leading. So, in order to resolve these dissonances correctly, the voices of the $C^{6/5}$ must move in a way that will resolve the chord to a *root* position F-major or minor triad. Similarly, the voices of the $D^{6/5}$ will resolve to a root position G-major or minor triad, and of the $E^{6/5}$ to a root position A-major or minor triad. This is exactly what happens in the *Tebe poem* phrase, which suggests that the above three seventh chords are all branching sisters to the following triads to which they resolve. In this way, they all form phrases with these triads, in which the triad is head and the seventh chord its branching sister.

The tree diagram of Example 1.2-30 depicts all these structural relationships, especially if you look at nodes 1, 2, and 3 at the bottom of the tree. Moreover, these little phrases [$C^{6/5}$ -F], [$D^{6/5}$ -G], and [$E^{6/5}$ -Am] all form a long left-branching prolongation of the A-minor triad, as the long slur in the bass clef of the reduction shows – the A-minor triad itself being a right-branching prolongation of the initial C-major triad, as previously discussed. This also means that there is no actual, harmonic, progression from,

say, the initial C-major triad to the following $C^{6/5}$ chord, since the initial C-major triad's actual progression is to the A-minor triad six bars later. (Therefore, the $C^{6/5}$ chord's top voice G4 does not come from the *Kopfton* E4 top voice of the preceding C-major chord, but rather from an inner (tenor) voice – as the diagonal line shows. This just shows again how the $C^{6/5}$ chord's top voice is not part of the *Urlinie*, but arises from a voice-leading operation from the C-major triad, whose *Kopfton* E4 actually leads to the E4 of the A-minor triad.) All of this just reinforces how important voice leading factors can be when describing a musical phrase's tonal structure. And all of this will have to be factored into our Merge-based derivation of this phrase for a Minimalist model of musical grammar to work.

In the light of the above voice leading considerations, what will a Merge-based derivation of the *Tebe poem* phrase look like? Well, first we have to establish the *Stufe* structure of the phrase, since these will be the inputs that Merge will combine into larger constituents, and ultimately the whole phrase. We have already discussed how the whole-note chords of the reduction in Example 1.2-30 represent the *Stufe* structure of this passage – with the exception of the cadential 6-4 sonority, which is subsumed under the following G-major triad's *Stufe*. Now all *Stufen* have scale degree features, as I have argued before, and one of these features is the cf feature, whose values have been ascribed to the *Stufen* in the *Tebe poem* phrase in Example 1.2-30 according to the metric developed in Examples 1.2-28 and 1.2-29. So, the initial C-major triad has been ascribed the arbitrary cf value of +cf 00, the following $C^{6/5}$ the same +cf 00 value (since it is a chord built on the same scale degree), the F-major triad after this the cf value of -cf 01, and so on.

Now, one could question whether a seventh chord should be given the same cf value as its triadic form, as is the case with the C-major and $C^{6/5}$ chords above. After all, their grammatical roles are very different – the C-major chord is part of the *Ursatz*, whereas the $C^{6/5}$ locally left-prolongs the following F-major triad. (In functional terms, we could say that the C-major triad is interpreted as tonic at S-structure, whereas the $C^{6/5}$ is interpreted as an applied or secondary dominant, i.e. a $V^{6/5}$ of IV.) The answer to this problem could lie in the fact that seventh chords are partly voice-leading phenomena, given the

dissonances in them that require appropriate voice-leading treatment. In this light – and given my earlier argument that voice-leading operations that transform a triad into its seventh chord form are akin to morphological operations in language – it could be that a seventh chord actually ends up having a different feature structure than its triadic form because of these pre-syntactic, ‘morphological’ voice-leading operations. This would explain the different syntactic behavior of a seventh chord compared to its triadic form. But without a proper consideration of the role of voice leading in determining a chord’s syntactic behavior, this answer is purely speculative – so I will continue to assume for convenience’s sake that triads and their seventh chord forms have identical feature structures, including in their cf values.

We can now proceed to a discussion of how the tree structure of Example 1.2-30 is derived, on the basis of our earlier discussion of how cf features ‘agree’. Starting with the lowest set of branches, and reading from left to right, we can see that the $C^{6/5}$ chord will merge with the following F-major triad because they have adjacent cf values of +00 and -01, and the order of merger will be *from* $C^{6/5}$ *to* F-major – which is the order these two chords appear in, in the surface of the piece. Also, the latter of the two chords will be the head of the resulting phrase as well, which means that it is the F-major triad that will project its ‘lexical information’ higher up in the tree. This is why node 1, where the $C^{6/5}$ and F-major branches join, projects the F-major chord’s cf value of -cf 01. In a similar vein, the $D^{6/5}$ chord merges with the following G-major triad to form a phrase at node 2 whose head is the G-major triad, and the $E^{6/5}$ chord merges with the following A-minor triad to form a phrase at node 3 whose head is the A-minor triad.

This last merger might raise some eyebrows because the $E^{6/5}$ and A-minor chords have non-adjacent cf values of +04 and +00 respectively. But the only reason for this discrepancy is because the $E^{6/5}$ chord is merging with an *A-minor* triad. If this minor triad were replaced with an A-major triad, which would be assigned the arbitrary cf value of +cf 03, the merger between $E^{6/5}$ and A-major would not be a problem at all, since these two chords *will* have adjacent cf values, of +cf 04 and +cf 03 respectively. Now as I have mentioned earlier, voice-leading considerations require an $E^{6/5}$ chord to resolve to either a root-position A-major triad or A-minor triad. In this light, we could invoke an exception clause that

allows two chords with cf values a fourth apart, such as E^{6/5} and A-minor, to merge if such a merger satisfies independent voice-leading requirements. Alternatively, we could say that two chords with the same root, such as A-major and A-minor, have the same cf value. This would give the A-minor chord a cf value of +cf 03 like the A-major chord, which will allow the E^{6/5} to successfully merge with it. Of course, this would also mean that C-major and A-minor will no longer have the same cf value of +cf 00 – and the two circles of fifths in Example 1.2-28 would no longer be commensurate, since we would have to twist the lower, minor-chord, circle by 90 degrees clockwise to be able to superimpose it on the upper, major-chord, circle. The problem with this is that the Tonic Prolongation part of the *Tebe poem* phrase, which involved the C-major to A-minor, I-vi, progression, will now be harder to generate, since it makes sense to think that two chords with the same or adjacent cf values can merge, but harder to see how two chords with cf values a third apart, like C-major's +cf 00 and A-minor's +cf 03, can possibly merge.

However, this problem might be more apparent than real, because there is more to the relationship between scale degrees than the ones they have in the circle of fifths. I mentioned at the outset of this section that a thorough exploration of the musical 'lexicon' requires an exploration of pitch spaces other than the circle of fifths. And one of these other pitch spaces is exactly the space that relates chords such as C-major and A-minor, i.e. the pitch space of chords whose roots span the interval of a third. So, even if C-major and A-minor end up being distant in a circle of fifths space, i.e. in terms of their cf values, they could still be adjacent to each other in a thirds-based pitch space. In fact, C-major and A-minor *are* closely related chords for voice-leading reasons – you can transform one into the other by just moving one of their voices by a step (e.g. you can move the fifth of C-major, G, up a step to A to get the A-minor triad, which is a voice leading operation called "5-6 motion"). In light of this we can even stipulate that chords that have cf values a third apart *are* adjacent in a "circle of thirds", if not in the circle of fifths. In this way, we can license the merger of A-minor with both the E^{6/5} that precedes it locally (in circle of fifths terms), and also the initial C-major whose *Kopfton* it prolongs (in circle of thirds terms).

The above argument will help us with a knotty problem when we continue on to the other *Stufe* mergers that occur in the generation of the *Tebe poem* phrase. First consider node 5, where the F-major

and G-major chords, previously merged with the $C^{6/5}$ and $D^{6/5}$ chords at nodes 1 and 2, now merge together themselves to form a more complex phrase. The F-major and G-major chords have non-adjacent cf values of -cf 01 and +cf 01 respectively. This might seem to be a problem for their further merger with each other – but such a merger is clearly possible, since the F-G progression in C major is usually taken to be a progression from a predominant (or subdominant) harmony to a dominant one, when interpreted in functional harmonic terms, which is a common and important progression in tonal harmony. Given that functional harmony licenses the merger of F-major and G-major, in a way their cf values do not seem to, could a functional harmonic approach to modeling grammar, as opposed to a *Stufe*-based one, be a solution here? As we saw earlier, this was exactly Allan Keiler’s approach, and this is the approach taken by Martin Rohrmeier too, in his aforementioned model of tonal grammar. So, we could say that the F-major and G-major chords have functional harmonic features, such as those of predominant and dominant (as opposed to scale degree features) – and this is what licenses the merger of F-major and G-major in the *Tebe poem* phrase.

Unfortunately, this will not do. This is because prior to their merger at node 5, both the F-major and G-major chords have already merged, at nodes 1 and 2 respectively. And at node 1, the F-major chord had *tonic* function, which allowed it to merge with its *dominant seventh*-functioning chord (i.e. the $C^{6/5}$). Similarly, at node 2, the G-major chord had tonic function too, which allowed it to merge with its dominant seventh $D^{6/5}$. So, for F-major and G-major to merge at node 5, they would have to suddenly switch their tonic function to predominant and dominant function – and there is no obvious reason for such a switch. Moreover, they cannot both continue to have tonic function, since this makes their merger impossible – the phrase resulting from their merger would be in both F-major and G-major simultaneously.

If instead, we take harmonic function to be a *semantic* feature of chords, which is interpreted at S-structure (and specifically at LF) *after* the *syntactic* merger of chords is complete, then the problem is solved – the changing nature of harmonic functions, which depends on which two chords are being merged, will not get in the way of their merger. This is why I have been pursuing a *Stufe*-based approach

to tonal grammar, where harmonic functions are taken to be semantic phenomena, interpreted at S-structure, after the *Stufe*-based grammatical merger of chords has been completed. As mentioned before, this is similar to the approach taken by Jonah Katz and David Pesetsky in their model of tonal grammar, where a harmonic-function based approach is rejected for similar reasons as the ones I have just stated. But instead of adopting a *Stufe*-based approach like I have, their model is based on simple pitch class information. This approach is problematic too, and I will review some of these problems after completing the description of my *Stufe*-based model.

So the problem of how F-major and G-major can be merged in a *Stufe*-based model, as opposed to a harmonic function based one, persists, given that F-major and G-major's cf values are non-adjacent. One solution to this problem is to propose that the F-major triad progresses to a *D-minor* triad – covertly, since we do not hear this D-minor triad in S-structure – forming the set {F-major, D-minor} with the D-minor triad as its head. This set then merges successfully with the following G-major triad, because the former projects its D-minor head's +cf 02 value, which is adjacent to the G-major triad's +cf 01 value. This is what we see at node 5, a phrase whose head is the G-major triad. Such a move would make complete sense in traditional music-theoretic terms too, since D-minor, as the chord built on scale degree 2, is adjacent to G-major in the circle of fifths, and is a common predominant functioning harmony – in fact, it was the predominant functioning harmony of choice for Mozart, Haydn, Beethoven and others, well into the 19th century. So, taking F-major to progress covertly to D-minor would allow F-major and G-major to successfully merge subsequently, yielding a phrase whose head would be G-major – which is exactly what we see at node 5.

What is the justification, though, for allowing a covert progression of F-major to D-minor prior to the merger with G-major? Well, here the 'circle of thirds' argument proposed above comes into play. F-major and D-minor are third related, just in the way C-major and A-minor are. In fact, we can transform the F-major triad into a D-minor one through 5-6 motion (by raising the 5th of the F-major triad, i.e. C, to the D of the D-minor triad) just in the way we transformed C-major into A-minor. In this light, we could argue that in the merger of F-major and G-major, F-major covertly progresses to D-minor through 5-6

motion, which then, legally, merges with G-major. If such a “covert progression” proposal seems ‘out there’, consider the fact that it is not that different from Jean-Philippe Rameau’s notion of the *double emploi*, in which an F-major chord and a D-minor (specifically a D-minor seventh) chord are considered to be essentially the same chord with two different roots (i.e. F and D), an idea that was later adopted in part by theorists like Simon Sechter and Arnold Schoenberg (see Meeus (2000) for a brief review of this idea). And I am not even suggesting that F-major and D-minor chords are the same – I am just suggesting that an F-major chord – and any chord interpreted as a subdominant or IV at S-structure – must *progress* to the, closely-related, chord built on scale degree 2, such as D-minor in this case, before it can merge with a dominant-functioning chord such as G-major in the *Tebe poem* phrase. In fact, such a IV-II-V functional harmonic progression is not even all that unusual, and can often be seen *overtly* in many pieces (i.e. where all three chords are heard in the S-structure of a passage), for example in mm. 5-6 of the Minuet from Mozart’s Haffner symphony (shown in Example 1.2-37, which I shall discuss in more detail later).

On the basis of this, one could make a historical case for the presence of such progressions, overt or covert, in tonal music too. That is, since chords built on scale degree 2 were the predominant chord of choice, much before subdominant chords became popular – in fact as far as back as the Renaissance, before the birth of functional harmony (e.g. see Gauldin (1995): 138-139) – a IV-V progression would only make sense when included within the more standard II-V progression, i.e. in a IV-II-V progression as in the Haffner phrase above. Now, the II in such a progression was often metrically-accented via a suspension – but having this full IV-II-V progression in a passage would have been cumbersome, especially if a composer wanted to include a (by definition) metrically-accented cadential 6-4 sonority to the passage too. This might have led to the II chord being gradually phased out of S-structure, resulting in a IV-II-V progression where the II was only implied, and not overt – of the kind we see in the *Tebe poem* example, and as became increasingly common in 19th century tonal harmony.

The above covert progression argument is helpful because it can also help explain the merger of the G-major headed phrase with the prolonged A-minor triad at node 7. A phrase headed by a G-major

chord would normally merge with a C-major one, given their adjacent +cf 00 and +cf 01 scale degree features – which, as discussed before, generates the Tonic Completion constituent at node 6. So, how does a G-headed phrase merge with an A-minor triad at node 7, especially if we take the A-minor triad’s cf value to be +cf 03 as proposed a few pages ago – a value that is non-adjacent to the G-major triad’s cf value? Well, we have already discussed how a C-major chord can progress to an A-minor chord through a thirds-based harmonic motion, and as happens at node 8 in the *Tebe poem* tree. So, what if G-major *does* progress to C-major as expected, which then progresses by thirds to A-minor – except that the initial progression to the C chord is covert, so that the progression at S-structure appears to be G-major to A-minor? This would explain how node 7 arises in the tree – and also how the so-called *deceptive* progression occurs, which is a progression in which a dominant-functioning chord progresses to a chord built on scale degree 6 rather than 1, such as G-major to A-minor in our *Tebe poem* phrase. It would also explain how the deceptive progression is closely associated with a progression to a tonic-functioning chord, since the chord built on scale degree 6 right-prolongs tonic (as happen at node 8), but how it also implies a delay in the arrival of this tonic chord, given that the tonic chord within the deceptive progression was covert.

If the above explanation for the derivation of node 7 sounds feasible, then we will have dealt so far with the cf-feature-based derivation of nodes 1-3, and 5-8 – which means we still have to explain the derivation of nodes 4 and 9 in the *Tebe poem* phrase for my Merge-based account of this phrase to be complete. Now, node 4 also involves an unusual non-circle-of-fifths type progression. Since we have already seen a few progressions like this, it might be worthwhile to enumerate these different unusual progressions, since a complete description of tonal grammar will have to account for such progressions too, in addition to circle of fifths-based ones:

- (a) Thirds-based progressions, such as I-VI, and IV-II.
- (b) Predominant functional use of IV.
- (c) Deceptive functional use of VI (and IV, as in a III-IV progression).

(d) Dominant functional use of VII.

To this list some authors add the “plagal” functional use of IV (e.g. Kostka and Payne (2013): 105), but I mentioned earlier how I consider this progression to be really a borrowed progression from folk and vernacular musics, and I will develop this point in a later subsection too.

Now, progressions (a) and (b) have already figured in our derivation of nodes 5 and 8, and (c) in the derivation of node 7. So, it is to progression (d) we now turn, since this is the unusual progression that features in the derivation of node 4. We discussed earlier how node 4 arises from an $F\#^{07}$ chord merging with a G-major triad (which is itself expanded by the cadential 6-4) to form a prolonged G-major phrase – which would be interpreted at S-structure as a Dominant Prolongation. Therefore, the G-major triad is the head of this phrase, and projects its cf value of +cf 01 to node 4. But notice how in the derivation of this phrase the $F\#^{07}$ is functioning as a VII chord in the key of G major. In other words, the phrase at node 4 would be interpreted as a VII-I progression at S-structure. This raises two concerns:

- (1) Only dominant-functioning chords such as a D-major chord in G major are supposed to progress to tonic-functioning I chords. So, how can a chord built on F# in G major have dominant function, giving rise to the unusual (d) progression above, where a chord built on scale degree 7 is interpreted as having dominant function?
- (2) Chords built on F# have arbitrary cf values of +cf 06, which is far away from a G-major chord’s cf value of +cf 01. So how can these two chords merge, so that a chord built on F# can be interpreted as having dominant function in G major?

Part of the answer to these questions might lie in the fact that the chord built on F# that has a +cf 06 value is really an F#-major chord (cf. the upper, major chord, image in Example 1.2-28 in this regard) – whereas the chord that merges with G-major at node 4 is an F#-*diminished* chord. This might make a difference to (1) and (2) because an F#-major chord would not normally merge with a G-major chord in the way an F#-diminished chord would, which makes sense given F#-major and G-major’s vastly

different cf values of +cf 06 and +cf 01 respectively. And the F#-diminished chord is a dissonant sonority, as particularly is its seventh chord form – which is the actual chord in the *Tebe poem* phrase. Dissonant sonorities have a voice-leading aspect to them, as discussed previously, so it could be that an F#-diminished chord can merge with a G-major chord, in the way an F#-major chord would not, because voice-leading factors make it progress to a G-major chord. This would, in a sense, make the F#-diminished chord's progression to G-major a pre-grammatical, 'morphological' voice-leading phenomenon.

But the difference between F#-major and F#-diminished chords should not be exaggerated. This is because F#-diminished chords can behave just like F#-major chords too – for example, they can progress to chords built on B in the same circle-of-fifths based way that F#-major chords can, because the cf values of chords built on F# and those built on B are adjacent (they are +cf 06 and +cf 05 respectively). Which means that F#-diminished chords can progress to chords built on both G *and* B (see measure 19 in the first movement of Mozart's K. 545 piano sonata for an example of the latter progression.) Moreover, F#-diminished chords do not even have to progress just to chords built on G or B – in their seventh chord form they can progress to a D-flat-major or minor chord, a B-flat-major or minor chord, or an E-major or minor chord too. In this sense, this chord is an independent harmonic entity – a *Stufe* – not tied to G-major because of voice-leading concerns. For this reason, it should be able to participate in grammatical operations, not just pre-grammatical ones, in the way all *Stufen* can – which means that there should be a *Stufe*-based, cf-value specific reason, for why F#-diminished chords can merge with G-major chords.

The answer here could be a similar “covert progression” one as was proposed for some of the other nodes in the *Tebe poem* phrase. The F#-diminished chord is third-related to the D-major chord – the true, dominant-functioning chord in G-major. One can transform the F#-chord into a D-major chord by 5-6 motion too, as was the case with F-major/D-minor and C-major/A-minor previously. In fact, this close connection between the two chords led Walter Piston to suggest that a diminished seventh chord on scale degree 7 is really an incomplete seventh chord on scale degree 5 (i.e. a “dominant 9th chord with missing root” (Piston (1978): 310)). This means that when an F#-diminished chord merges with a G-major chord,

it first merges, covertly, with a D-major chord, through 5-6 motion, which *then* merges with the G-major chord, in the way D-major chords do because of G- and D-major's adjacent cf values (+cf 01 and +cf 02 respectively). The justification for this covert movement is less historical in this case than it was above, and has more to do with voice leading. Since the dissonant nature of the diminished chord requires an *audible* resolution, the intermediate D-major chord must be covert, or else the resolution of the F#-chord's dissonance through voice leading into the G-major chord will not be heard. This might also be the reason why F#-major chords are never seen in direct progressions to G-major chords – not being dissonant entities, they do not need to audibly resolve to a consonant sonority like a G-major chord.

This brings us to the remaining unexplained node in Example 1.2-30, viz. node 9 – the top node of the tree, formed by merging its Tonic Prolongation and Tonic Completion branches. Both of these branches project the +cf 00 values of their C-major heads. As I have said before, which of these two heads, the initial C-major triad or the final one, is the head of the complete tree is a matter not easily resolved, and the literature on this is inconsistent too. Cf values will not help us here, since both heads project identical cf values, which does not help us decide which head will ultimately be the head of the whole tree.

I believe a different kind of scale degree value, a different kind of 'lexical' scale degree feature has to be invoked here, which is semantic in origin. Just as words have both syntactic and semantic features, it is a semantic feature within a *Stufe*'s feature structure – and not one of its syntactic features (like its cf values) – that determines which of the two C-major *Stufen* is the head of the whole tree. The reason for this is that the final C-major triad, the cadential tonic, is often taken to be the head of the whole tree, the entire phrase that is, because it confirms the phrase's tonality and gives it closure – it gives the phrase its authentic cadence, and resolves all the tension inherent in it. But closure is a semantic phenomenon, so the *Stufe* feature that gives a phrase closure – such as the *Urlinie*'s scale degree 1 in the final tonic triad, which arises in the triad after descending from the *Kopfton* – is more a semantic feature than a syntactic one. I will have more to say about this in chapter 2.1, but for now this might explain why *syntactic* discussions about which triad heads a musical phrase have remained inconclusive.

This concludes my discussion of the derivation of a tonal phrase from *Tebe poem* through a Merge-based procedure. I should say, however, that even if the above series of arguments explain, to the reader's satisfaction, how the various chord mergers in the *Tebe poem* phrase occur, my model still has some problems. The main problem is of course my invocation of covert progressions to license some of the *Stufe* mergers in the phrase. In addition to being speculative, there are inconsistencies in this proposal. For example, there is an inconsistency regarding which chord in the above covert progressions is actually covert. In the progression from F-major to G-major at node 5, the intermediate *minor* chord on D was taken to be covert, whereas in the progression from G-major to A-minor at node 7, the intermediate *major* chord on C was taken to be covert. Inconsistencies such as this require explanation. However, I believe my above Merge-based model's ability to describe the derivation of the *Tebe poem* phrase in 'lexical' (i.e. cf value-based) terms – without the attendant problems of a functional harmonic model – gives it some validity that merits further consideration.

Before we wrap up this discussion of musical constituents, harmonic *Stufe* 'lexicons', and Merge, it might be worth contrasting my model with that proposed by the linguists Jonah Katz and David Pesetsky (in Katz and Pesetsky (2011): 57-64) for the same *Tebe poem* phrase. First of all, Katz and Pesetsky, and Martin Rohrmeier before them (whose model Katz and Pesetsky's model is a response to), do not consider voice-leading factors in their models. This prevents them from noticing the right-prolongation of the initial C-major chord by the A-minor chord in measure 6, all of which sustains the *Kopfton* E4 in the *Urlinie* of the phrase. As a result, they take the A-minor chord to be a *left*-prolongation of the final C-major chord, which seems to be an incorrect analysis of the phrase to me. Ignoring the voice-leading aspects of the phrase also leads both authors to overlook the problem of the cadential 6-4 in measure 7, which cannot be a *Stufe* in its own right. Katz and Pesetsky implicitly acknowledge this, since the cadential 6-4 does not play a role in their derivation, but they never say why it does not do so – which would have forced them to countenance voice leading factors in the phrase too. The authors do, however, reject a harmonic function-based approach, such as the one Rohrmeier adopts, for the same reason my model does. But they base the various chord mergers in their model not on *Stufe*-based factors, but on

what they simply call “pitch-class information” – which is all that projects up in the phrase’s tree structure, as opposed to harmonic function information or, as in my model, scale-degree features. The problem with this might be obvious now – it is not clear how pitch-class information by itself can decide how a C-major chord merges with, say, a G-major chord. What order would the two merge in – i.e. how would the merged set be linearized – and what would be the head of the merged set, the C-major chord or the G-major chord? To answer questions like this, the C_{HM} needs something like the scale degree features I have proposed, in order to generate legitimate S-structures. Since Katz and Pesetsky deny outright that music has a lexicon, they do not consider the possibility of such ‘lexical’ scale degree features in generating S-structures, and just discuss the role of an interpretive component in deciding the legitimacy of a derived S-structure, which they call the “Tonal Harmonic Component”. Such an interpretative component surely decides what harmonic function various nodes in the tree have, but for it to make such decisions an S-structure has to be generated in the first place – and it is not clear how the C_{HM} can accomplish this based on pitch class information alone. This seems to imply, in conclusion, that a Schenkerian scale-degree or *Stufe*-based approach to deriving musical phrases, within a broader Merge-based computational system for music, is the correct way to approach the task of modeling musical grammar.

ii. Pesetsky on cadence and Schenkerian “interruption” forms

I claimed in the last chapter that much of this dissertation is aimed at providing a response to Lerdahl and Jackendoff’s landmark work in musical grammar, *A Generative Theory of Tonal Music*, especially since the authors claim in that text that music and language are ultimately not identical – which is a direct rejection of the identity theses for music and language proposed in the last chapter. It is not a coincidence then that much of my own exploration of musical grammar has been influenced by the work of Allan Keiler, as seen in the last section – and given the Schenkerian orientation of this dissertation, since Keiler’s work represents the pro-Schenker approach to generative music theory, whereas Lerdahl and Jackendoff’s approach ultimately rejected Schenker.

Now one of Lerdahl and Jackendoff's most incisive criticisms of the search for music/language identity lies in their claim that musical grammar does not have transformations in it, whereas linguistic grammar clearly does, e.g. in cases of movement – which made Chomsky present his grammatical theory as one of “Transformational Generative Grammar” (TGG) since the origins of the field of generative linguistics. If Lerdahl and Jackendoff's claim is true that would be a devastating blow for any music-language identity thesis. However, in this section I will argue that musical grammar *does* seem to have a transformational component in it too, and the argument will be based again on some observations made by Allan Keiler.

I would like to start though with a discussion of some recent work by the linguist David Pesetsky in this regard, whose Merge-based approach to modeling tonal grammar, developed in collaboration with his graduate student Jonah Katz, we briefly explored in the last section. Pesetsky's work in trying to develop a Minimalist approach to musical structure actually precedes that collaborative project by a few years, and some of this earlier work (e.g. Pesetsky (2007)) was already attempting to locate transformational phenomena in music, specifically in places where *internal* Merge is implicated (internal Merge being, as we have seen before, the grammatical operation behind movement transformations). For example, in the above-cited paper, he made the argument that the setting to rhythm of generated grammatical structures in music happens through the application of internal Merge. More recently, in the collaborative project with Jonah Katz, Pesetsky has claimed that authentic cadences in Western Classical tonal music are the result of internal Merge-based movement transformations too.

The basic argument regarding cadences is as follows. Western Classical tonal phrases are normally considered complete and closed only when they reach an authentic cadence (a phenomenon that is, incidentally, idiom-specific, since the requirement that phrases close with authentic cadences is not seen in idioms like Rock music – a point I shall develop in the next subsection). Such an authentic cadence minimally involves two *Stufen*, a dominant-functioning one (which is almost without exception a root-position triad built on scale degree 5) and a tonic-functioning one that follows this dominant-

functioning one (which is also almost without exception a root-position triad built on scale degree 1). In other words, there is an adjacency constraint on the authentic cadence – viz. a V triad must be adjacent to the final I triad for the authentic cadence to occur. (I use “V” and “I” hereafter not as labels for harmonic functions, but for *Stufen*, as Schenker did himself.)

Also, the importance of this V-I progression in Western tonal music is seen in the fact that this is the progression that constitutes Keiler’s Tonic Completion constituent – the grammatical constituent in a musical phrase that, as is evident in the name, completes the phrase and confirms its tonality. Moreover, the V-I progression, and indeed Keiler’s Tonic Completion constituent, comprise the end part of the Schenkerian *Ursatz*, the part where scale degree 2 in the *Ursatz* descends to scale degree 1, the two scale degrees harmonized, respectively, by the V and I chords of the V-I progression. Often this part of the *Ursatz* is actually manifested in the last two sonorities of a musical phrase too, as indeed happens in the *Tebe poem* phrase above. In such cases, the final authentic cadence in a phrase is the same as the structural V-I progression of the *Ursatz* – which reveals the *Ursatz*’s importance as the grammatical ‘backbone’ of a phrase.

But sometimes the V of the *Ursatz*, the so-called “structural dominant”, is not the penultimate sonority in a phrase, which means that it does not appear adjacent to, and to the left of, the final, cadential tonic. (We will soon see that this is what happens in Classical period forms, in which the V in the half cadence that frequently ends the period’s antecedent phrase is often taken to be the structural dominant. Which means that this V is not only *not* the penultimate sonority in the period, it appears before the onset of the entire consequent phrase too.) In such phrases then, the final cadential tonic has to be preceded by another V chord, in order for the phrase-closing authentic cadence to occur. We have also seen how the structural dominant in the *Ursatz* can itself be prolonged – giving rise to Keiler’s “Dominant Prolongation” constituent. If the head V chord of this DP constituent is right-prolonged within the DP, i.e. if it does not appear on the right edge of the DP, it will be non-adjacent to the final, cadential tonic too. Which means that another V chord must again precede the final, cadential tonic chord for the phrase-closing authentic cadence to occur.

One way to ensure this is to propose a movement transformation that moves the head V chord of the DP constituent to the right edge of the DP constituent – or more specifically to the left of the final tonic triad – so that a V chord will appear adjacent to the final tonic triad, which would therefore allow an authentic cadence to occur (or in other words, it would allow a cadential ‘feature’ to be checked). This would also be a form of head-to-head movement (akin to T- and V-raising in language), because only chords, like words (and not phrases), can be heads, and the movement transformation just proposed takes the *head* V chord of the DP constituent and moves it adjacent to the *head* of the Tonic Completion constituent, i.e. adjacent to the final tonic triad, to realize the final authentic cadence. And this, in essence, is Pesetsky’s movement-based explanation for the phenomenon of authentic cadences in Western tonal music.

One can easily see the importance of such a proposal. Not only does it reveal the possibility of an actual transformational component within musical grammar, it also provides an important piece of evidence against the most damaging of Lerdahl and Jackendoff’s arguments against the identity of music and language. But let us examine this proposal a bit more closely. For one, it seems to suggest that the movement of the head V chord is from the position of the structural dominant in the *Ursatz* to the position immediately adjacent to the final tonic triad. This means that in phrases like the *Tebe poem* one above, where the structural dominant is already adjacent to the final tonic, the movement of this structural dominant to the tonic-adjacent position (to realize the authentic cadence) would not involve moving across any intervening chords at all – since there are no chords in between the dominant and tonic in such phrases. In other words, the movement would be ‘motionless’ in a sense, or “string vacuous” to use the technical term for such movements (Katz and Pesetsky (2011): 47). This idea is illustrated more clearly in Example 1.2-31 (taken from Katz and Pesetsky (2011): 43), where the head dominant chord δ of the DP constituent δP moves, via internal Merge, to the final tonic chord τ in a string-vacuous way – as shown by the dotted arrow – to create the $\delta + \tau$ authentic cadential structure. (Given their, and my, rejection of a harmonic-function based approach to modeling musical grammar, Katz and Pesetsky

Example 1.2-31. Internal Merge in the full cadence from Katz & Pesetsky (2011)

τP
 τ'
 δP
 vP
 $\delta + \tau$

$[\tau P]$... $[\tau']$ $[\delta P]$ v ... $] _]$ $[\tau \delta + \tau]$

intentionally use lower case Greek alphabets to label the constituents of their tree, in order to avoid using labels like D and T, which are more commonly used to label harmonic functions like Dominant and Tonic.)

String-vacuous movements are not problems in and of themselves, since they are common enough in language. In fact, we saw an example just a few pages ago in sentence (1f):

(1f) *What_{obj} does Ulrike believe who_{subj} t_{subj} read t_{obj}?

Here, the subject of the subordinate clause undergoes string-vacuous wh-movement, which is why the moved wh-phrase who_{subj} occupies a position adjacent to the subject NP position where it moved from – where it leaves behind the trace t_{subj}. In other words, the wh-phrase has not crossed over any other words in undergoing its movement transformation. And this string-vacuous movement is not problematic in and of itself – the reason (1f) is ungrammatical is only because it contains a second wh-movement, involving the object wh-phrase “what”, which violates a Bounding constraint. But string-vacuous movements should be invoked with care of course, since where they occur there seems to be no movement at all – which is why I used the term “motionless” to describe them in the last paragraph. So, one should be sure that a movement transformation is actually happening before stating that it is of the string-vacuous type.

And it is not clear that a movement transformation *is* actually happening in the case of authentic cadences. This is reinforced by the fact that it is unclear where the head V chord, which moves to the tonic-adjacent position in a cadence, moves *from*, as Katz and Pesetsky state themselves (Katz and Pesetsky (2011): 49). I suggested earlier that it seems the head V chord is moving from the position of the structural dominant – but Katz and Pesetsky provide evidence against this supposition. For one, they say (following a suggestion from Dmitri Tymoczko) that even a *non*-dominant functioning chord can realize the authentic cadence, as long as it appears adjacent to the final tonic chord – which suggests that this

chord does not move from a structural *dominant* position.⁵⁹ They also say that, unlike Lerdahl and Jackendoff's model, which:

“...adds a third chord to its description of a cadence, by asserting that a cadence is always a cadence *of* something, which they call the “structural beginning” to the cadenced group ... this notion will not figure into our discussion.” (Katz and Pesetsky (2011): 42)

But if the authentic cadence involves a movement of a structural dominant to a final, tonic-adjacent position, as part of the constituent called Tonic Completion, then it *is* clearly a cadence of something – viz. the initial part of the *Ursatz*, i.e. Keiler's “Tonic Prolongation” constituent. So, there is no way Katz and Pesetsky could make the above statement if the head V chord that moves to the final tonic-adjacent position moves from a structural dominant position.⁶⁰

So, if we do not know where the head-to-head cadential movement is taking place from, especially if not from a structural dominant position, and if the movement itself, being string vacuous, seems to be more apparent than real, then the validity of Katz and Pesetsky's proposal that authentic cadences reveal an internal Merge-based movement transformation in tonal music becomes questionable. In fact, we could say that the positioning of a dominant chord next to the final tonic in an authentic cadence is not the result of a movement transformation at all, but rather a different phenomenon – say, an

⁵⁹ This statement is clearly incorrect though. Katz and Pesetsky cite an example suggested by Tymoczko, where a minor iv chord progresses to I, as an instance of a non-dominant functioning chord creating an authentic cadence with the following tonic. But such a “plagal” iv-I progression is *not* a cadential progression, though it is often misunderstood as one in the music-theoretic literature. As the noted form theorist William Caplin says, “such a cadence rarely exists – if it indeed can be said to exist at all. Inasmuch as the progression IV-I cannot confirm a tonality (it lacks any leading-tone resolution), it cannot articulate formal closure ... Most examples of plagal cadences given in textbooks actually represent a postcadential codetta function: that is, the IV-I progression follows an authentic cadence but does not in itself create genuine closure” (Caplin (1998): 43-45). So, the above minor iv-I example does not really further Katz and Pesetsky's case for the authentic cadence.

⁶⁰ In fact, their unwillingness to consider earlier harmonies, prior to the final δ and τ , in their definition of an authentic cadence, raises questions about how accurate Katz and Pesetsky's definition of this term is to begin with. Consider William Caplin's words in this regard again, “We must be careful not to identify a passage as cadential unless we can demonstrate that it logically ensues from previous *initiating* or *medial* functions” (Caplin (1998): 43). This, in addition to the fact that an authentic cadence usually has other factors associated with it (such as a descent in the structural melodic line or *Ursatz*), suggests that Katz and Pesetsky's reduction of the concept of the authentic cadence to two adjacent δ and τ chords is too drastic. And if an authentic cadence should include more than just two individual chords, then the form of movement transformation Katz and Pesetsky ascribe to cadences is probably not of a head-to-head variety either, since it should probably involve phrases – e.g. a dominant-prolongational phrase (Keiler's DP) that includes a predominant harmony of some kind (a II or IV chord, an applied VII of V or V of V, or a cadential 6-4).

insertion operation. Just as the phenomenon of Expletive insertion is invoked in language to fulfill the requirement that a sentence must have a subject, we could propose a Dominant insertion operation to fulfill the requirement that all tonal phrases must have a dominant chord prior to the final tonic, so that the phrase can end with an authentic cadence. This would allow us to explain the phenomenon of the authentic cadence in tonal music without having to invoke suspicious string-vacuous movement transformations to do so.

In fact, I will propose now that this, insertion operation, is indeed the correct explanation for authentic cadences, for the reason just mentioned. This is not the happiest of conclusions though, since authentic cadences are specific to Western tonal music, and idiom-specific phenomena – such as word-order differences across languages – are often best explained in movement-transformational terms, as we have seen before. On the other hand, insertion operations are often invoked to satisfy *universal* grammatical principles, in the way Expletive insertion is invoked to satisfy the universal Extended Projection Principle. So, my invocation of an insertion operation to satisfy Western tonality's idiom-specific cadential requirement can rightfully be questioned.

But the reason I avoid explaining authentic cadences in movement-transformational terms is because I actually do think that movement transformations occur in tonal music – just not in the place where Katz and Pesetsky claim they occur. In fact, I believe that a movement transformation does occur in tonal phrases, but *from* the authentic cadence *to* the position of structural dominant – the opposite of the direction of movement that I initially suggested was implied in Katz and Pesetsky's model. And the reason why I believe such a transformation exists in tonal phrases is because Heinrich Schenker said so himself – albeit only implicitly, as a result of which this important observation has not merited much discussion in the literature. Importantly, I think this observation acts as the strongest piece of evidence against Lerdahl and Jackendoff's critique of music-language identity, so it will be worth our while to look at this phenomenon closely.

Schenker's (implicit) observations about movement transformations occur within his discussion of so-called *interruption forms*. Musical passages are said to have an interrupted form when they begin structurally with the scale degree 3-2 motion of the *Urlinie*, harmonized by a I-V progression reminiscent of the *Ursatz's* harmonic structure, but then follow this “with a second beginning which retraces and completes the opening gesture, perhaps with some elaboration” (Forte and Gilbert (1982): 201). To understand this clearly, consider Example 1.2-32, whose top image provides Liszt's piano reduction of the first phrase from the famous “Ode to Joy” theme in Beethoven's 9th Symphony. (Ignore the two lower images, i.e. the ones marked Fig. 9 and Fig. 11, for the time being.)

The phrase is an example of a Classical 8-bar period. So, the first half of the larger period is the antecedent phrase, which lasts from mm. 1-4, followed by the consequent phrase of mm. 5-8 – which taken together give us the 8-bar period. Now, notice how the antecedent phrase starts in tonic D-major harmony, but then ends at a half cadence over A-major harmony in bar 4 – implying that the antecedent's larger harmonic structure is a I-V progression in the key of D major. This progression harmonizes the initial part of the period's *Urlinie*, which in this case involves the scale degree 3 *Kopfton* F#5, played by the winds on the very first beat of the phrase (and doubled an octave higher), descending to the scale degree 2 E5 over the half cadence – all of which is shown by the capped Arabic numbers. But rather than descend to scale degree 1 after the half cadence to complete the *Urlinie*, the melody at the beginning of the consequent phrase *repeats the opening idea* of mm. 1-3, and thus *resumes* the scale degree 3 F#5 *Kopfton* of the *Urlinie* in measure 5. This pitch then descends again to scale degree 2, in the last measure – but this time scale degree 2 does progress further down to the final tonic pitch D5 in the melody over the final, cadential I chord. This, then, closes the phrase and completes the *Urlinie*.

So, rather than having a unified scale degree 3-2-1 *Urlinie* harmonized within a conventional I-V-I *Ursatz* progression, the Ode to Joy period traces an *interrupted Urlinie* of scale degrees 3-2 || 3-2-1, harmonized by the chord progression I-V || I-V-I, to reveal an *Ursatz* with an interrupted form – the || symbol at the half cadence representing the interruption in the structure of this phrase. Since many if not

Example 1.2-32. Beethoven, Symphony #9, Op. 125/iv: Interruption form in “Ode to Joy” theme

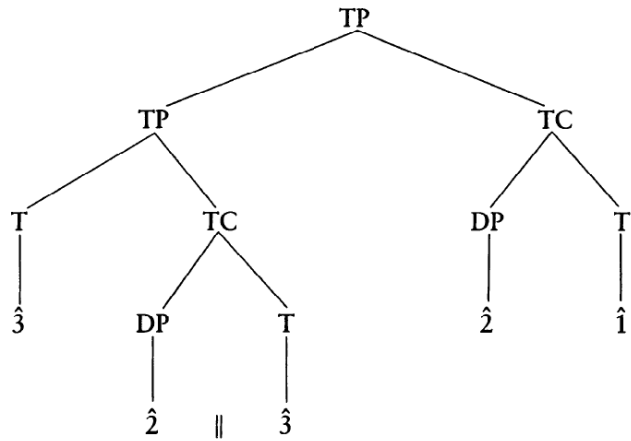


Fig. 9. Interruption technique: tree structure of $\hat{3}-(\hat{2} \parallel \hat{3})-\hat{2}-\hat{1}$.

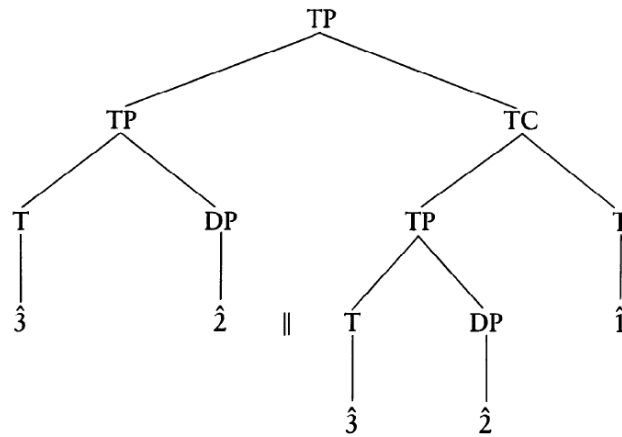


Fig. 11. Interruption technique: alternative version to Figure 10.

most Classical periodic forms contain an antecedent phrase that ends in a half cadence, followed by a consequent phrase that repeats the antecedent's opening harmonic and motivic material, as we see in this famous Beethoven theme, Classical periods therefore provide many of the examples that one sees in the tonal literature of Schenker's interruption form.

The structure of a phrase in interruption form provides a problem for models of tonal grammar. This is because, unlike 'un-interrupted' forms, with their conventional I-V-I structure, interrupted forms have *two* dominant *Stufen* in them and *three* tonic *Stufen*, as we see in their I-V || I-V-I harmonic structure. If we assume that tonal grammar is hierarchical, with the components of the *Ursatz* representing the most hierarchically-superior *Stufen* in a tonal passage, the problem that interrupted forms provide us with concerns *which* of the two V *Stufen* is the structural dominant, and which of the two earlier I *Stufen* is the initial structural tonic of the *Ursatz*, which supports the scale degree 3 *Kopftön* and heads its "Tonic Prolongation" part. (The last I *Stufe* is still assumed to be the final, structural tonic of the *Ursatz* – the tonic that heads the "Tonic Completion" part of the *Ursatz* – because it is the only tonic that supports the final scale degree 1 in the *Urlinie*, meaning that it also serves as the point of cadential arrival, at the end of the phrase.)

Schenker seems to have struggled with this exact problem in his analysis of the interrupted main theme from Brahms' Op. 56a "Variations on a Theme by Haydn", which also happens to be in the form of a Classical period, with a 5-bar antecedent phrase and a 5-bar consequent phrase. (I discuss this theme in more detail, especially its unusual 5+5 bar phrase structure, in chapter 2.2.) In his brilliant account of Schenker's analytical struggle with the Brahms theme (in Keiler (1983-84): 221-228), Allan Keiler describes three possible hierarchical readings Schenker could have entertained of its harmonic structure. Following Keiler, I have listed them here in terms of the scale degrees they harmonize. (A parenthetical scale degree implies that the chord that harmonizes it is hierarchically-inferior to the other chord that harmonizes the same scale degree elsewhere in the theme):

Reading 1: (3 – 2) || 3 – 2 – 1

Reading 2: 3 – (2 || 3) – 2 – 1

Reading 3: $3 - 2 \parallel (3 - 2) - 1$

In the first reading, the tonic harmony after an interruption, e.g. at the beginning of the consequent phrase in Example 1.2-32, is taken to be the initial structural tonic of the phrase, since both chords prior to the interruption are represented parenthetically by the scale degrees they harmonize in the *Urlinie*. This implies that the following, final V in the period would have to be its structural dominant as well. Readings 2 and 3, in contrast, take the first tonic harmony, *before* the interruption, to be the structural tonic that heads TP and harmonizes the scale degree 3 *Kopfton*. But if we take this tonic to be the structurally and hierarchically-superior one, then we will have two dominant harmonies following it – either of which could be the structural dominant. Reading 2 takes the latter, and Reading 3 takes the former of these two harmonies to be that structural dominant.

Now Allan Keiler reads Schenker's own words, in *Der freie Satz*, as implying a strong preference for a reading in which the second V chord is taken as the structural dominant, which we see above in Readings 1 and 2. This is because Schenker felt that the initial descent, before the interruption, was merely an attempt at a structural descent to scale degree 1 in the *Urlinie*, meaning that only the final V and I chords, i.e. the ones involved in the final cadence at the end of the consequent phrase, should be taken as the structural dominant and final structural tonic in the period. This also means that only the scale degrees they harmonize in the *Urlinie*, i.e. the final scale degrees 2 and 1, should be taken as completing the structural descent to tonic in the *Urlinie*. As Keiler says, this is in line with Schenker's "generally left-branching perspective", i.e. one in which the rightmost constituents of a tonal structure are taken to be superior in its hierarchical phrase structure, with more leftward constituents serving to prolong these rightward constituents, as seen in the left-branching architecture of the structure's tree.

Moreover, Schenker generally believed that the first of the two earlier I *Stufen* should be the initial structural tonic too, the one that harmonizes the scale degree 3 *Kopfton* in the *Urlinie*. This means that his reading of choice for interruption forms would be Reading 2. The hierarchical tree structure of such a reading is illustrated in the middle image of Example 1.2-32, which is Figure 9 from Keiler (1983-

84). Here we see how the hierarchically-superior constituents are the initial I *Stufe*, which supports the scale degree 3 *Kopfton*, and the final two V and I *Stufen*, which harmonize the final scale degree descent of 2 – 1.

The middle two *Stufen*, i.e. the V *Stufe* before the interruption and the I *Stufe* after the interruption, can now be inserted into the tree, as an intermediate TC constituent, since this entire constituent continues to support the scale degree 3 *Kopfton* in the *Urlinie*. But this leads to a problem. Although these two *Stufen* must be inserted into the tree as a TC constituent on linear grounds (i.e. to continue supporting the scale degree 3 in the *Urlinie*, prior to its descent at the very end of the whole phrase), they cannot be inserted as a TC constituent on harmonic grounds. This is because they do not form a constituent to begin with. The V *Stufe* in it does not progress to, and therefore does not left-prolong, the following I *Stufe* given that they are on either side of the interruption sign. Put another way, the V *Stufe* occurs in the antecedent phrase's half cadence, whereas the following I *Stufe* begins the next, consequent phrase. So, again, there is no grammatical connection, no prolongational relationship, between these two *Stufen*. However, it is *only* as the two branches of a TC constituent that these two *Stufen* can be inserted in the tree, if its hierarchical structure is to accord with the linear *Urlinie*-based aspects of Reading 2. For this reason, Schenker ends up *abandoning* the preferred Reading 2 analysis of interruption forms, and accepts Reading 3 instead:

“With respect to the unity of the fundamental structure, the first occurrence of the [scale degree] 2 is more significant than the second.” (Schenker (1979): 37)

Allan Keiler sees Schenker's switching to Reading 3 as just more evidence for the former's critique of the latter's harmonic theory that we explored in the last subsection, which can be seen in Schenker's inconsistent use of the term “*Stufe*”. And again Keiler bases his critique on the contrapuntal, as opposed to harmonic, basis for Schenker's ideas – which are ideas that do not seem to deal with the notion of harmonic constituency adequately (seen especially in Schenker's inability to justify the middle TC):

“[Reading 2] requires that DP and T [in the intermediate, inserted TC constituent] be immediate constituents, and in contrapuntal terms, suggests that the initial [scale degree] 3 - 2 || 3 sequence be understood as a neighbor figure prolonging the initial 3. This is, of course, incorrect; the 2 of the first

Ursatz does not connect cadentially to the 3 that returns after the interruption. Schenker, in fact, realizes this and rejects that neighbor note possibility himself. But since Schenker was constrained by the limited possibilities of comparing hierarchically the two (or parts of the two) *Ursatz* structures in these examples in terms either of contrapuntal formulations that made some sense, or in terms of a left-branching perspective that had to remain vague and largely instinctive, he was forced to abandon this formulation of the interruption technique for one that is absolutely contradictory with his usual modes of analysis.” (Keiler (1983-84): 222-223)

Clearly Keiler finds Schenker’s change of heart in this regard indefensible, and as a sign of an inadequacy in Schenkerian theory’s ability to explain harmonic structure and grammatical relationships. However, what is interesting is *how* Schenker gets from Reading 2 to Reading 3. As the Schenkerian theorist David Beach says, in an attempt to defend Schenker against Keiler’s critique:

“I view Schenker’s interpretation as not a single analysis, but as a series of analyses, where events are reinterpreted and renotated at each stage. This does not justify the inconsistencies, but perhaps better explains their existence.” (Beach (1985): 294)

And as Keiler says himself, in his discussion of Schenker’s analysis of another part of the Brahms “Haydn Variations” theme, one can interpret Schenker’s reading of the Brahms theme as a process, rather than just a single analysis – specifically one that requires a *constituent movement* operation of some kind, which allows us to begin with one analysis but end up with another (Keiler (1983-84): 218). Given Keiler’s above thoughts about Schenker’s analysis of this Brahms theme, he finds such a movement transformation to be completely unmotivated. However, I think there is a good motivation for such a movement transformation – in fact, the very motivation that inspired David Pesetsky’s movement proposal we explored earlier, viz. the fact that tonal phrases – even those in interruption form – require authentic cadences to confirm their tonality and to receive adequate closure. In this light, I think it is perfectly reasonable that Schenker starts out analyzing an interruption form in accordance with Reading 2, but then switches to Reading 3. This is because both readings can be justified, albeit on different grounds, and so we need a way to entertain them both – and it is precisely because musical grammar has a transformational component that both readings *can* be entertained, since it is a movement transformation that connects the two readings.

To understand this, let us first see what kind of movement transformation we need to get from Reading 2 to Reading 3. Consider the bottom image in Example 1.2-32, which is Figure 11 from Keiler (1983-84), and which presents Keiler's tree diagram for a Reading 3 analysis of the interruption form's hierarchical structure. Look at the second of the two scale degree 2s here, i.e. the one that is a branching sister to the tonic *Stufe* after the interruption sign (and which represents the second V *Stufe* in the phrase). You will notice that it occupies a hierarchically inferior position to the one it did in the tree diagram for Reading 2, as Reading 3 requires anyway, since it has now been demoted in hierarchical status relative to Reading 2. It now occurs within a TP constituent whose head is the tonic *Stufe* to the left of it – a tonic *Stufe* that is not even part of the *Ursatz* of the phrase. This is instead of its being a branching sister to the final tonic *Stufe* to the right of it that heads the whole phrase, as it was in Reading 2.

Now this second scale degree 2, and the second V *Stufe* that supports it, can acquire this hierarchically inferior position in a number of ways. The way suggested by Keiler's Figure 11 seems to be that the middle tonic *Stufe* in the phrase, which is part of the problematic TC constituent in Reading 2, has to move out of this constituent, and then merge with our second V *Stufe* in the phrase, forming the lower TP constituent in Keiler's figure 11, represented by the parenthetical (3 – 2) constituent in Reading 3. This would demote our second V *Stufe* in hierarchical status from being a left-branching sister of the final tonic *Stufe* in Reading 2, the tonic *Stufe* that supports scale degree 1 in the *Ursatz* and heads the entire phrase, to being a right-branching sister to the middle tonic *Stufe* that has just moved. This would also make the *first* V *Stufe* in the phrase, i.e. the one to the left of the interruption sign, a right-branching sister of the *first* tonic *Stufe* in the phrase – in effect raising its hierarchical status, since this first tonic *Stufe*, unlike the middle one, is part of the *Ursatz* of the phrase.

So, one way to get from Reading 2 to Reading 3 is by a movement transformation involving the middle tonic *Stufe*. But there does not seem to be much motivation for this movement, which is perhaps the reason for Allan Keiler's rejection of this whole endeavor, i.e. Schenker's switch from Reading 2 to Reading 3. If you think about this in another way, notice that the above movement transformation merges the middle tonic *Stufe* with our second V *Stufe*, and in that order – which is the wrong order too, if you

remember our discussion of how such mergers occur in the derivation of the *Tebe poem* phrase earlier. The movement also leaves the first V *Stufe* as the right-branching sister of the first tonic *Stufe*, elevating the first V *Stufe*'s hierarchical status in the process. But this is again the wrong order, the incorrect I – V order, if one buys my cf value-based explanation for how *Stufen* merge. Of course one does not have to buy that explanation to begin with – but it certainly does not make the case for the above middle tonic movement transformation any stronger, given that this movement does not have any independent justification of its own.

However, Schenker did switch analyses from Reading 2 to Reading 3, and I believe that this can be justified, but with a *different* movement transformation. To understand this, let us first find a motivation for starting with Reading 2 as Schenker did – or else we could just start *with* Reading 3, and arrive at an analysis of an interrupted phrase form without requiring any kind of movement transformation whatsoever. So, why even start with Reading 2, as indeed Schenker did himself? The answer might be clear by now – it is to ensure that the phrase has an authentic cadence. If you look at the “Ode to Joy” theme at the top of Example 1.2-32 again, you will notice that its second scale degree 2 is harmonized by the very last V *Stufe* in it – a *Stufe* that is already adjacent to the final tonic, since it joins with that tonic to form the authentic cadence that closes the consequent phrase, and therefore the “Ode to Joy” theme’s period form. And Reading 2 makes this very V *Stufe* the structural dominant in the *Ursatz* of the phrase. So, it makes sense to start with a Reading 2 analysis of the phrase, since this will ensure that the structural analysis of the phrase includes an authentic cadence. (In fact, a good proportion of Schenkerian analyses of tonal pieces do take the final, cadential V *Stufe* to be the structural dominant of the *Ursatz* – and I will even propose this as a stipulation, i.e. a hierarchical analysis of a tonal phrase that does *not* read the final, cadential dominant as the structural dominant of the phrase will be an incorrect analysis of the phrase, *unless* the structural dominant has moved to another position through a movement transformation, a possibility I will explore in the next few pages.)

But Reading 3, as we know, puts the structural dominant at the position of the *antecedent* phrase's *half cadence* – which is the *first* scale degree 2 marked in the “Ode to Joy” theme. So, to switch from Reading 2's *Ursatz* to Reading 3's *Ursatz* requires a movement transformation – not one in which the middle tonic *Stufe* moves, but rather one in which the final V *Stufe* – the structural dominant – moves, *from* the consequent phrase's authentic cadence *to* the antecedent phrase's half cadence. To understand this, let us represent Reading 2 and 3 with scale degrees as follows, with *2a* representing the structural dominant in each reading, and *2b* the other V *Stufe*:

(Reading 2): 3 – (2*b* || 3) – 2*a* – 1

(Reading 3): 3 – 2*a* || (3 – 2*b*) – 1

With this in mind, the movement transformation I am proposing, through which we can begin with a Reading 2 analysis of an interruption from, but end up with a Reading 3 analysis, will have the following derivational steps:

(Step 1): || 2*a* – 1 (merge structural V and final I to derive TC)

(Step 2): 3 – || 2*a* – 1 (merge initial TP with TC to derive *Ursatz*)

(Step 3): 3 – || 3 – 2*a* – 1 (merge middle TP with TC to derive ‘mini’ *Ursatz*)

(Step 4): 3 – 2*a* – || 3 – 1 (move structural V out of ‘mini’ *Ursatz*)

(Step 5): 3 – 2*a* – || 3 – 2*b* – 1 (insert second V to restore authentic cadence)

What is happening here is that we are deriving the *Ursatz* of Reading 2 first, in steps 1 and 2. Then, in step 3, we take the middle tonic *Stufe*, which also supports scale degree 3 in the melody and therefore heads its own TP constituent – and we merge this embedded TP with the TC, to generate a ‘mini’ *Ursatz*, an *Ursatz* within an *Ursatz* as it were. (This accounts for the fact that the consequent phrase, which begins with a TP headed by the middle tonic *Stufe*, *repeats* the opening material of the antecedent phrase – and so has a similar general phrase structure, i.e. *Ursatz*, as the larger phrase it is in.) Also, this is just another

application of External Merge too, totally consistent with how we derived the *Tebe poem* earlier – and the ‘*Ursatz* within an *Ursatz*’ structure just reveals the recursive structure of tonal grammar again.

It is with step 4 that things get interesting. It is here that my proposed movement transformation occurs, and it involves the structural dominant. That is, the structural dominant now moves *out* of the mini *Ursatz* it was previously in and merges, not with the final tonic *Stufe* as it did in Step 1, but with a *projection* of this *Stufe* – *higher up* in the tree. Importantly, this gives the appearance of the structural dominant having merged with the *initial* tonic *Stufe* (or a projection of this *Stufe*), although this is not the case, as just mentioned, and would not be possible because that would lead to an ungrammatical, ‘out of order’ I – V progression, as I argued in my cf value-based derivation of the *Tebe poem* phrase in the last subsection. But as a result of the structural dominant’s moving out of the mini *Ursatz* constituent, the middle tonic *Stufe* now progresses directly to the final tonic *Stufe*, and the phrase has lost its authentic cadence. It is for this reason that a second V *Stufe* is *inserted* in step 5, to realize the cadence – in the manner I proposed in my earlier review of Pesetsky’s Merge-based cadence proposal. (Alternatively, one could say that when the structural dominant moves out of the mini *Ursatz* it leaves a copy of itself behind, if one believes in the copy theory of movement (discussed in footnote 50) – and this copy would allow the authentic cadence to be realized too. Such a proposal would help us get rid of step 5, and avoid invoking an insertion operation, which, as I argued earlier, is not the happiest way to explain authentic cadences anyway.)

So, *this* is how I think movement transformations occur in tonal phrases – especially interruption forms – and *this* is how I justify Schenker’s switching from Reading 2 to Reading 3 in his analysis of the interruption form’s phrase structure. That is, Schenker starts off with Reading 2, given his “general left-branching perspective”, which allows the phrase to have a nice structural, cadential close – but then he switches to Reading 3 because of a need to elevate the hierarchical status of the first V *Stufe*. All of which can be accomplished if we allow a movement-based explanation of interruption forms, and of tonal grammar in general.

But of course the question now is how we can possibly justify moving the structural dominant to another position in the phrase? What can the motivation possibly be for invoking a transformation operation in *music* (contra Lerdahl and Jackendoff's thoughts about this), a movement transformation which we can call "DOM-raising" because of the way it involves *raising* the structural dominant out of the mini *Ursatz* to a position higher up in the tree? But the answer to this is actually quite simple – just as the final tonic *Stufe* in a tonal phrase must be (left) prolonged by a dominant-functioning harmony for the phrase to receive an authentic cadential close, the *initial* tonic *Stufe* of a phrase also needs to be expanded in some way by a *following* V *Stufe* to confirm its tonic status too – hence the need to raise the structural dominant to a position where it appears to prolong the initial tonic rather than the final one. Such dominants, which appear to (right) prolong an earlier tonic *Stufe* (but do not *actually* do so, since that would be ungrammatical), are called "back-relating dominants", and they "close off a musical idea without leading it to a definitive conclusion [i.e. via an authentic cadence]" (Aldwell and Schachter (2011): 189). This is exactly the role of the first V *Stufe* in an interruption form – it finishes off the initial musical idea of the phrase, e.g. the antecedent idea in Beethoven's "Ode to Joy" period, which is accomplished by the *half* cadence that ends the antecedent, without leading the larger structure to a definitive conclusion.

But the role of back-relating dominants is not restricted only to interruption forms, such as those seen in Classical periods. In fact, one could argue that *all* tonal phrases have a back-relating dominant, whose function is to 'secure' the initial tonic's status as a structural tonic *Stufe*. Consider Example 1.2-33 in this regard, which presents a Schenker graph of the first strophe from the Brahms song "Wie bist du, meine Königin", which is the 9th song from his Op. 32 cycle of songs. This strophe, though a complete phrase, does not have an interrupted structure, and it is not in Classical period form. Yet, as the analysis suggests, the phrase seems to be divisible at the middle, as shown by the vertical dotted line in measure 12 of the graph. Prior to the dotted line, the background harmony of the phrase starts with I and ends with V, so the V here is a back-relating dominant harmony. This part of the phrase also supports the scale degree 5 B-flat *Kopftön* of the phrase, as shown by the two half notes in the soprano voice of the phrase, beamed together by the long horizontal line in the graph.

After the dotted line, the *Urlinie* begins its descent back to scale degree 1, starting from the scale degree 4 A-flat in measure 15, which is nothing but the seventh of the back-relating dominant B-flat chord, which has now been turned into a V^7 chord through 8-7 motion down from the *Kopfton*'s B-flat in the *Urlinie*. So, we see that the background harmony that supports this descent back to scale degree 1 in the *Urlinie* is a $V^7 - I$ progression – in other words, Keiler's "Tonic Completion" constituent. The V^7 chord is expanded over measures 12-19, first by the predominant IV chord (which is itself prolonged through a $V^7/IV - IV^6$ progression in mm. 12-14) and then by an applied leading-tone seventh chord and a cadential 6-4 progression in mm. 15-19, all of which support the scale degree 4-3-2 descent in the *Urlinie*. In other words, this whole prolongation of the V^7 over mm. 12-19 is Keiler's "Dominant Prolongation" constituent, which finally progresses to the cadential tonic *Stufe* in measure 20. Since the head of the DP constituent, the V^7 chord that supports scale degree 2 in the *Urlinie* in measure 19 appears adjacent to the final, cadential tonic, it gives the phrase an authentic cadential close too.

So, we see that the above Brahms phrase has a background harmonic structure of "I – V, V – I" or more specifically "I – V, $V^7 - I$ " (and it is quite common for the initial back-relating V to be turned into a cadential V^7 chord too, before it returns to the final I). And again, the initial I – V is only an apparent progression, which I shall explain in more detail in a bit. In this light, it might be better to think of the Schenkerian *Ursatz* not as a "I – V – I" structure, but as a "I – V, V – I" structure, where the first I – V, which contains the back-relating dominant, supports the *Kopfton* in the *Urlinie*, as Keiler's "Tonic Prolongation" constituent, and the final V – I, which contains the cadential dominant, harmonizes the melodic descent back to scale degree 1 in the *Urlinie*, and closes off the phrase as Keiler's "Tonic Completion" constituent. But the *Ursatz* is after all a "I – V – I" structure, with only one V *Stufe* in it, and that one V *Stufe* therefore has to play both the role of back-relating *and* cadential dominant. And the only way this can happen is through a movement transformation, viz. my proposed movement of DOM-raising.

Example 1.2-33. Brahms, *Lieder & Gesänge*, Op. 32 No. 9 “Wie bist du meine Königin”: Analysis, mm. 6-20

mm. 6 11 15 20

Middleground: I V⁷ I⁶ V⁷/V V⁸ V⁷/IV IV⁶ V⁷ : 6 V³ : 4 vii^{♭7}/V V⁴ : 3 I 8 : 7

Background: I ——— V⁸ ——— V⁸ ——— 7 - 6 ——— 5 ——— I

I think Schenker understood this implicitly, which in my opinion is the motivation behind his switching from Reading 2 to Reading 3 in his analysis of the interruption form in Brahms' "Haydn Variations" theme. This is why this observation by Schenker, and Keiler's brilliant, but largely forgotten, exposition of it, acts as the best evidence for the hypothesis that musical grammar is transformational too, just like language.

Before we move on to the next section, I should clarify a couple of aspects of the DOM-raising transformation I described above. First of all, this movement transformation is not of the head-to-head variety, but more an example of phrasal movement (akin to NP or wh-movement in language). This is because the raised constituent could be just the structural dominant, but it could also be an entire dominant-prolongational phrase (i.e. Keiler's DP) *headed by* the structural dominant. Making this movement phrasal, as opposed to the head-to-head movement Katz and Pesetsky describe, also accounts for the fact that authentic cadences – which the DOM constituent originally participates in, prior to being raised – actually involve more than the two head δ and τ chords that Katz and Pesetsky invoke to define authentic cadences. Authentic cadences normally have a melodic component to them too, especially a descent from the *Kopfton* to scale degree 1 in the *Urlinie*, and this is often harmonized by multiple chords, and not just δ and τ . (Consider the second half of the Brahms phrase in Example 1.2-33 in this regard, where an entire *dominant-prolongational phrase*, spanning mm. 12-19, harmonizes the scale degree 4-3-2 descent in the *Urlinie*, and thus prepares the final cadential arrival on the tonic *Stufe* in measure 20.) So, the phrasal, as opposed to head-to-head, nature of the DOM-raising movement seems to fit better with how authentic cadences work in tonal music.

The second, and more important, aspect of the DOM-raising movement is that this movement, as you might recall, makes the structural dominant attach with the leftmost branch of the entire phrase's tree *again*, just with a higher projection of the final tonic *Stufe* this time – and, importantly, *not* with the initial tonic *Stufe* or any of its projections. This might contradict the idea that the structural dominant moves to provide the initial tonic *Stufe* with a back-relating dominant. But there is no contradiction here. The

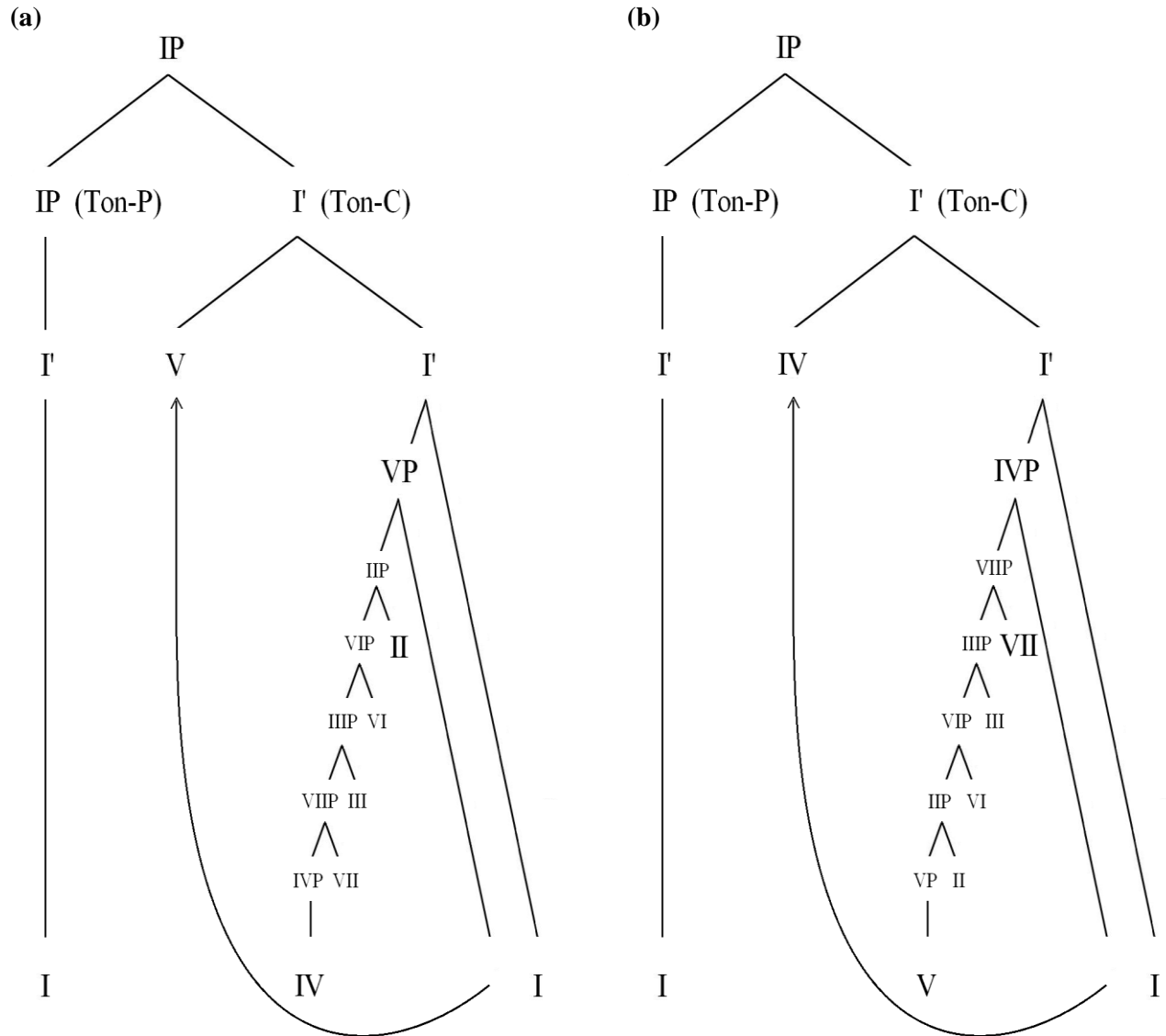
structural dominant does move to a back-relating dominant *position*, but it is not really a back-relating dominant – in fact, there is no such thing as a back-relating dominant in *actuality*, although the moved structural dominant fulfills this function in appearance, given the position it moves to. An actual back-relating dominant can only happen in an ungrammatical I – V progression, where the V right-prolongs the I. Therefore, the moved structural dominant only has the *appearance* of being back-relating. In actuality though, it just left-prolongs the final cadential tonic in a grammatical V – I way, but from higher up in the tree than its original, final-tonic adjacent position, which is why it cannot be the cadential dominant anymore (and takes on a back-relating function instead).

There is an important reason for why the DOM-raising movement must happen in this specific way. There is a regulation on movement transformations, which requires that a moved constituent be able to c-command its original position (Pesetsky (2013): 124). In its original position, the structural dominant is a left-branching sister to the final tonic *Stufe*, whose phrasal projection is Keiler’s “Tonic Completion” phrase. This means that all the higher projections of the tonic *Stufe*, all the way up to the phrasal TC projection, will dominate both the head tonic *Stufe* and its structural dominant branching sister. Now, if the raised structural dominant merges with a higher projection of the final tonic *Stufe*, then, as the branching sister to this projection, it will be able to c-command this projection and all of its daughters and grand-daughters as well – which includes the node the structural dominant occupied prior to movement. So by merging with a higher projection of the final tonic *Stufe*, the raised structural dominant will be able to c-command its original position – thus satisfying the regulation on movement transformations. But if the structural dominant merges with a projection of the *initial* tonic *Stufe*, it will not be a branching sister to a projection of the *final* tonic *Stufe* anymore, and will therefore *not* be able to c-command its original position. And this will violate the regulation on transformations, which is why the raised structural dominant cannot merge with a projection of the initial tonic *Stufe*, even though it appears to be this *Stufe*’s back-relating dominant. This is why the I – V progression in tonal music is only an apparent progression that results from a movement transformation, as I briefly stated in the last subsection.

The above issues are better illustrated in Example 1.2-34a, which provides a more general description of the DOM-raising transformation. (Ignore Example 1.2-34b for the time being – I will deal with it in the next subsection.) The vertical line on the far left of the tree in this example contains the bar-level and phrasal projections of the initial tonic (i.e. “I”) *Stufe*, whose phrasal projection IP, as we know, is Keiler’s “Tonic Prolongation” (labeled as “Ton-P” in the example). On the far *right* of the tree, on the other hand, are the bar-level and phrasal projections of the final tonic *Stufe*. So far I have referred to the phrasal projection of this *Stufe* as Keiler’s “Tonic Completion”, but in this example I have switched to calling the maximal *bar-level* projection of this *Stufe* (i.e. I’) as the Tonic Completion constituent (labeled as “Ton-C” in the example). This is because the final tonic *Stufe*’s phrasal projection is actually the whole *Ursatz* itself, which is formed by merging the final tonic *Stufe*’s I’ projection, Ton-C, with the initial tonic *Stufe*’s IP projection, Ton-P. My switch in labeling here allows us to specify what the head of the whole phrase is, given the earlier ambiguity regarding this issue, and also allows us to model the whole phrase in a left-branching way, which accords with Schenkerian intuitions in this regard.

To the left of the final I *Stufe* in the example is a dominant-prolongational phrase (Keiler’s DP, labeled here as VP), which is headed by the structural dominant, and which merges with the final I *Stufe* to form that *Stufe*’s lowest I’ projection. You will notice that the rightmost position in this VP (not to be confused with a linguistic verb phrase!), is the position occupied by the structural V *Stufe* head of this phrase, which allows this *Stufe* to be adjacent to the final tonic, and thus provide the phrase with an authentic cadence. But the structural V *Stufe* *itself* does not merge with this final tonic – what merges with the final tonic, i.e. in Reading 2, is the structural V *Stufe*’s phrasal projection VP, as the branching structure of Example 1.2-34a shows. There are other constituents within the VP, which serve to left-prolong the structural V *Stufe*, not only in accordance with Schenkerian left-branching intuitions, but also in accordance with the cf-value based system of phrase derivation we discussed in the last subsection. It is only when the structural V *Stufe* has merged with one or more of these other constituents, to generate its phrasal projection VP, does this *phrasal projection* then merge with the final tonic *Stufe*. And then it is

Example 1.2-34. Grammatical movement in music: (a) DOM raising vs (b) SUBD raising



this phrasal projection that moves to the position higher up in the tree in the DOM-raising movement too. This is why the structural *V Stufe* merges *first* with the predominant *II Stufe*, with whom *V* shares adjacent cf values, to generate the VP phrase, before any further mergers or movement transformations take place – and which is why DOM-raising is a form of phrasal, and not head-to-head, movement. In Example 1.2-34a, this merger involves the *V Stufe* merging with a phrasal projection of the *II Stufe*, viz. IIP – which makes the *V Stufe* the head of the resulting set, i.e. the VP phrase.

The arrow in the example suggests, however, that what moves in the DOM-raising movement is the structural V *Stufe* itself, since the arrow originates from that *Stufe*'s original position at the bottom of the tree, adjacent to the final tonic *Stufe*, rather than from the VP position – which is what we would expect if it is the entire VP phrase that is moving, and if DOM-raising is indeed a form of phrasal, and not head-to-head, movement. There is a reason for why I have placed the arrow in the way it appears in the example – it is indeed the case that the VP is what moves in DOM-raising in general, but the only part of the VP that *actually* moves in Example 1.2-34a is the structural V *Stufe* head of this VP. The IIP branching sister of the structural V, and all its daughters and grand-daughters (which make up all of the remaining parts of the VP) do not move in this particular case because IIP is an adjunct to the structural V *Stufe*, and therefore does not have to move with it. (This is akin to how adjunct PPs within an NP do not have to move when the rest of the NP moves, e.g. in verbal passives (cf. Roeper and van Hout (2006): 8).)

The reason for why IIP is an adjunct – and not a complement or a specifier, which would normally require it to move with the head – is as follows. II functions as a predominant sonority in tonal phrases, which is why it (specifically its phrasal projection IIP) precedes the dominant-functioning structural V *Stufe* in the above example. But consider a crucial difference between the way II behaves as a predominant, when compared with the predominant *Stufe* known as the “applied” or “secondary” leading-tone chord, i.e. VII⁷/V, which does not appear in Example 1.2-34a, but which appeared in both the *Tebe poem* phrase in Example 1.2-30, and in the Brahms phrase in Example 1.2-33 – and also to the immediate left of the cadential V *Stufe* in each case (the latter expanded through voice leading by a cadential 6-4 in each case too). In tonal phrases, II can progress directly to V (as in (19a) below), or have the VII⁷/V intervene between it and V (as in (19b)), but it cannot *follow* VII⁷/V in a progression to V (as in (19c)):

(19a) II – V

(19b) II – VII⁷/V – V

(19c) *VII⁷/V – II – V

This set of facts also occurs when we use a secondary dominant, i.e. a V/V or a V⁷/V, instead of the VII⁷/V, and has a voice-leading reason behind it. For example, in C major, the V/V, V⁷/V, and VII⁷/V chords all have an F# in them, which normally resolves up to the root of the V chord, viz. G, for voice-leading reasons, unless it moves down to the F-natural of a G⁷ chord, where the chord progression still places the applied chord and the V chord right next to each other. In other words, the F# of the applied chord would not normally progress down to the F-natural of a II chord, and so II's intervening between the applied and V chords leads to ungrammaticality, as happens in (19c). This means that predominant applied chords *have* to be adjacent to the chords they resolve to, whereas a predominant like II does not. Moreover, a predominant chord that is closely-related to II, such as the third-related IV, can occur alongside it in a progression to V, such as in (19d), but the closely-related applied chords V/V and VII⁷/V do not occur next to each other, as in (19e), suggesting that they are in complementary distribution with respect to each other:⁶¹

(19d) IV – II – V

(19e) *V/V – VII⁷/V – V

All of these facts suggest that predominant chords like II and IV on the one hand, and VII⁷/V and V/V on the other, behave in rather different ways – specifically, II and IV (and their phrasal projections, more precisely) behave like adjuncts to V, whereas the V/V and VII⁷/V behave like complements. For this reason, in a movement transformation that involves the phrasal projection of a V *Stufe* – as is the case with DOM-raising – the phrasal projection of a II *Stufe*, as an adjunct to the head V *Stufe* of this larger phrasal projection, does not have to move. Which explains why in Example 1.2-34a only the structural V *Stufe* moves to the higher position in the tree, leaving the IIP and all its daughters and grand-daughters behind, even though the movement transformation here is still of the phrasal kind.

⁶¹ There are passages where a VII⁷ chord seems to be followed by a V⁷ chord, such as in mm. 123-131 of the first movement of Beethoven's Op. 57 "Appassionata" piano sonata. But such passages would be heard as prolonging really just the V⁷ chord, through a neighboring motion involving scale degrees 5 and 6. This is particularly true of third inversion VII⁷ (i.e. VII^{4/2}) chords, which are invariably voice-leading elaborations of root position V⁷ harmony (cf. Aldwell and Schachter (2011): 427).

What the above facts also suggest is that, if the structural V *Stufe* in Example 1.2-34a were to be preceded by the phrasal projection of a VII⁷/V chord, this phrase *would also* move with the V *Stufe* to the new position higher up in the tree, unlike the way IIP behaves in the tree in the example. This is because complements, unlike adjuncts, normally *do* move along with their heads in movement transformations. A testable hypothesis arises out of this. Since adjunct II phrases do not have to move with the head V *Stufe* of a VP phrase in a DOM-raising transformation, but VII⁷/V phrases do, as per the above theory, we can make a prediction that the first V *Stufe* in tonal structures that appear to have two V *Stufen* in them, such as the interruption forms we have looked at, *has to be preceded by the same predominant as the one that precedes the second V Stufe, only if the predominant is an applied chord, but not if it is II or IV*. This is because when the V *Stufe* raises from its cadential position to its back-relating one, a predominant II *Stufe* will not necessarily raise with it – which means that another predominant can occur before the V *Stufe* in its raised position. However, since a VII⁷/V *Stufe* will raise *with* the V *Stufe*, according to our hypothesis, this VII⁷/V *Stufe* will continue to be the predominant to the V *Stufe* even in its raised position.

A thorough investigation of, say, Classical periodic forms in the tonal literature can verify this hypothesis. I shall not do so now, in the interests of time, but hope to do so in the near future. Although evidence for this hypothesis is likely to be found in specifically *interrupted* periodic forms – not other, non-interrupted forms. For example, this phenomenon does not seem to obtain in a different kind of periodic form, which constitutes the first of William Caplin’s “hybrid” themes (Caplin (1998): 59-63). This form begins with an antecedent phrase, just as a standard period does, which often ends with a half cadence as well. But instead of a consequent phrase that repeats the opening harmonic-melodic material of the antecedent, the antecedent in this hybrid theme is followed by a *continuation* phrase, which presents new harmonic-melodic material, before ending, more often than not, with a typical authentic cadence. (Such a continuation phrase was sometimes called a *contrasting* consequent phrase in earlier theories of musical form, and the larger, hybrid, structure a “contrasting period”.)

So, instead of a I – V || I – V – I harmonic structure, the above hybrid theme can have a harmonic structure such as I – V, IV – II – V – I. (Notice the absence of interruption here, since the theme does not

repeat the antecedent's opening tonic material at the beginning of the consequent – as a result of which a simple comma can suffice to depict the division between the two phrases of the theme.) Exactly such a hybrid theme constitutes the main theme of the Minuet from Mozart's 35th symphony, K. 385, also known as the “Haffner”. Example 1.2-35 depicts this phrase in mm. 1-8, where we can clearly see an antecedent phrase ending on a half cadence in mm. 1-4, followed by a version of the above contrasting progression, viz. IV – II⁶ – V⁷ – I, in the continuation phrase of mm. 5-8. The entire theme has two V *Stufen* in it, at the half cadence and in the final perfect authentic cadence. We can explain this phenomenon as the result of a DOM-raising movement now, with the structural V *Stufe* starting out in the perfect authentic cadential position in bar 7, and then moving to bar 4 to realize the half cadence. What is interesting here though, in terms of the preceding argument about predominants, is that the predominant ii⁶ chord in bar 6 does *not* appear before the half cadential V in bar 4 as well. Which means that if the half cadential V is just the DOM-raised authentic cadential V, then the latter's predominant II *Stufe* did not raise with it to the half-cadential position. This adds evidence to the above idea that predominants like II are adjuncts, and not complements, to V *Stufen* – which is reinforced by the fact that two predominants, IV and II, occur in this passage, which would not be possible if these predominants were in complementary distribution. (However, as I said in the last paragraph, even complement-like predominants will probably not raise with the raised V *Stufe* in such non-interrupted formal structures – the reason for which is puzzling.)

This adjunct-like behavior of the common predominants II and IV (and their common inversions and seventh-chord forms, like II⁶, II^{6/5} and IV⁶) makes it possible not only for phrases like the Haffner Symphony Minuet's main theme to arise, but also for the variety of phrases seen in tonal music to arise – otherwise, all tonal phrases would have the same predominants and the same predominant-to-dominant functional progressions in them. So, it makes sense that adjunct-like predominants would behave in this non-raising manner.

Example 1.2-35. Mozart, Symphony #35 "Haffner", K. 385/iii: I-V / IV-II-V-I progression, mm. 1-8

The musical score is divided into two systems. The first system (measures 1-5) features the Oboe, Clarinet in D, Bassoon, Horn in D, and Timpani in D-A. The Oboe, Clarinet, and Bassoon play a melodic line starting on G4, moving to A4, B4, and C5. The Horn in D plays a rhythmic accompaniment of eighth notes. The Timpani in D-A plays a steady eighth-note pattern. Dynamics include *f* and *p*. The second system (measures 6-8) features Violin 1, Violin 2, Viola, and Cello & Bass. The Violins and Viola play a melodic line starting on G4, moving to A4, B4, and C5. The Cello & Bass play a rhythmic accompaniment of eighth notes. Dynamics include *f* and *p*. The harmonic progression is I-V / IV-II-V-I.

Instrument labels: Oboe, Clarinet in D, Bassoon, Horn in D, Timpani in D-A, Violin 1, Violin 2, Viola, Cello & Bass.

Harmonic progression: I — V — IV — ii⁶ — V⁷ — I

But now the question arises as to how the Haffner Minuet theme is derived to begin with. We have seen how periodic structures, with their DOM-raising movements, can be generated, and we have seen in the last subsection's discussion how chord progressions like IV-II-V can arise in tonal music, such as in the *Tebe poem* phrase. But how can a phrase be generated that has *both* of these characteristics, as is the case with the Haffner Minuet phrase?

To answer this, let us return to Example 1.2-34a and examine the tree diagram it depicts more closely. We have already seen how the larger *Ursatz* of this phrase is derived, i.e. by merging its Ton-P and Ton-C constituents, and how the DOM-raising movement arises within this *Ursatz* structure. We have also discussed the dominant-prolongational VP phrase here, which participates in the DOM-raising movement, and which is derived by merging the structural V *Stufe* head of this phrase with the IIP phrasal projection of the predominant II *Stufe*. Now, my very use of the term "head" to describe the V *Stufe*, and the suggestion that the predominant IIP is an "adjunct", implies an X-bar structure for this phrase, and possibly tonal music in general – which is an interesting implication in itself, given the linguist Ray Jackendoff's prior rejection of it (e.g. in Jackendoff (2009): 200-202).

But if this is the case, then my tree representation of the VP in the example is clearly inadequate. This is because in our earlier X-bar theoretic description of how a head's phrasal projection is derived, we observed that a head merges first with its complement to derive a bar-level projection, X' , and only after this X' projection has merged with another constituent, viz. the specifier, is the phrasal projection, XP , of the head generated. In fact, we see an illustration of this phenomenon in the way the VP in Example 1.2-34a merges with the final tonic *Stufe* of the phrase. The VP clearly acts like a complement to the final tonic *Stufe* in this case. Unlike predominants like II and IV, it occurs only by itself in tonal phrases, unaccompanied by any other dominant-functioning *Stufe*, which it is presumably in complementary distribution with. (Just compare the two predominants II and IV at the end of the Haffner Symphony Minuet phrase in Example 1.2-35, with the solitary V^7 that follows them, to witness an example of this phenomenon.) Also, the VP *has* to be adjacent to the tonic *Stufe* for cadential reasons. That is, it has to be a branching sister to the final tonic *Stufe*, and no other constituent can intervene between them, or else

they cannot form an authentic cadence because the adjacency constraint on cadences will be violated. In consequence, the VP merges directly with the final tonic *Stufe* – to yield a bar-level projection of the tonic *Stufe*, viz. I'. It is with this I-bar level projection that the VP merges again in the DOM-raising movement to generate the Ton-C constituent – presumably as the specifier of this constituent, as moved phrases often are (recall how *wh*-phrases move to the specifier of CP position in *wh*-movement, and how subject noun phrases move to the specifier of TP position to check for nominative case in NP-movement).⁶²

In light of the above, the VP in Example 1.2-34a should be derived by the V *Stufe*'s merger first with a complement (e.g. a VII⁷/V, in light of previous arguments in this regard) to derive the V-bar level projection V', and only then should a merger occur between this V', and the IIP phrasal projection of the II *Stufe*, to derive the VP phrase. So, I should have fixed the inadequate tree representation of the VP in the example by introducing complement constituents, like a VII⁷/V or a V⁷/V, into it. I did not do so firstly for reasons of space and visual clarity, i.e. to avoid cluttering up the example with too many branches, but more importantly because a Minimalist approach to musical grammar does not commit us to any specific (X-bar) representation of tonal phrases scheme anyway. This is important because if we forced a true X-bar representational schema on to Example 1.2.34a's phrase structure then we would have to read the IIP constituent in the VP as its specifier, and not as an adjunct – which would problematize the whole treatment of dominants as either adjuncts or complements discussed above. The rejection of a specific X-bar based schema also allows the merger of a variety of chords that do not necessarily have X-bar type relationships among them, as long as they have the kind of scale degree features that allows them to be merged successfully (even if musical phrases *do* have X-bar structure).

⁶² If this is the case, then Ton-C should be a phrasal projection, like CP and TP in language. The I-bar label assigned to the Ton-C node militates against this reading though, especially since this reading is motivated by the belief that the phrasal projection of the final tonic *Stufe* is the entire phrase itself, and not just the Ton-C constituent – in accordance with Schenkerian left-branching intuitions about tonal structure. However, we *could* argue, instead, that Ton-C is the phrasal projection of the final tonic *Stufe* – which merges with the *initial* tonic *Stufe*, as its phrasal complement, to generate the complete phrase. This would make the initial tonic *Stufe* the head of the whole phrase though, and the whole phrase, i.e. Ton-P, a phrasal projection of this initial tonic – a violation of the generally left-branching intuitions about tonal structure we have been entertaining so far. But since this would also allow the entire phrase to have an almost exact X-bar structure, within which movements like DOM-raising occur in language-like ways, this is a conclusion that merits more serious consideration.

We can see this in the internal structure of the VP in Example 12.34a, where a number of constituents appear, all related by circle of fifths features – the point being that this Merge-based, transformational model of grammar I have been proposing is not restricted to phrases where the only chords being merged are manifestations of I and V *Stufen*, as was the case with the regular and interrupted *Ursatz* models we explored a few pages ago. So after the II *Stufe* left-prolongs the structural V *Stufe* via its IIP phrasal projection, this II *Stufe* can itself be left-prolonged via the cf-value adjacent VI *Stufe*, via its phrasal projection VIP as the example shows, and in the typical ‘anticlockwise in the circle of fifths’ manner we explored in the last subsection’s *Tebe poem* phrase derivation. But now the VI *Stufe* can itself be left-prolonged too, by the cf-value adjacent III *Stufe*’s phrasal projection IIIP; the III *Stufe* then, in turn, can be left-prolonged by the cf-value adjacent VII *Stufe*’s phrasal projection VIIP, and finally this last *Stufe* can be left-prolonged by the IV *Stufe*’s phrasal projection IVP – giving us all the possible circle-of-fifths progressions available to tonal grammar.⁶³ (Which I have shown in smaller letters in the example, so as not to clutter it with too many large branches.)

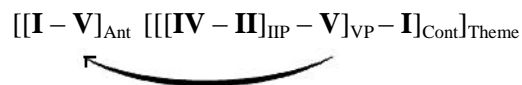
The above, cf-value based, internal branching structure of the VP in Example 12.34a occurs within a larger phrase in which a DOM-raising movement occurs as well. Consequently, we can now see how we can work our new DOM-raising transformational approach to tonal grammar, based on interruption forms, into the earlier cf-value based approach to deriving tonal phrases, explored during our derivation of the *Tebe poem* phrase in the last subsection – and this can show us how a tonal structure like the theme from Mozart’s Haffner Minuet might be generated.

But this model is inadequate when it comes to non circle-of-fifths relationships in tonal music, such as the thirds-based relationship between the predominant *Stufen* IV and II, which I used as part of

⁶³ An interesting characteristic of all these circle-of-fifths progressions is that they include major, minor and diminished triads. For example, in C major, the IIP that merges with the G-major V *Stufe* is the phrasal projection of a D-minor triad, whereas the IVP that merges with the B-diminished VII *Stufe* is the phrasal projection of an F-major triad. In contrast, if the predominant that merged with the G-major V *Stufe* were the phrasal projection of an *applied* dominant, viz. a V/V – which bears a circle-of-fifths relationship with the V *Stufe* just like IIP – this applied chord *has* to be a major triad, D-major in this case, and the V *Stufe* it progresses to *has* to be a major or minor triad. (This is actually true of all cadential V-I progressions.) This difference in the two kinds of circle-of-fifths relationships presents another piece of evidence for treating IIP as an *adjunct* to V (and IVP as an adjunct to VII), but V/V as a *complement* to V.

my “covert progression” justification for how IV can merge with V in the derivation of the *Tebe poem* phrase. I have displayed these two *Stufen* with larger Roman numerals in Example 1.2-34a, so that their relation in the example can be understood more easily. As you can therefore see, to merge IV with II in the model requires merging IV with VII first via IV’s phrasal projection IVP – resulting in a phrase headed by VII, viz. VIIP. This VIIP will then have to be merged with III and so on, until we finally end up merging II into the tree – all because the branching structure shown here is based on circle-of-fifths relationships. This is why the specific phrase model shown in Example 1.2-34a needs to be replaced with a more general Merge-based one, in which any kind of chord merger can occur, including IV with II. We can implement such a model in the example by allowing thirds-based mergers as well, through 5-6 motion as discussed in the *Tebe poem* derivation – and this will allow for direct IV-II mergers in a tonal phrase too.

As a result of all of this, we can derive a phrase with the harmonic structure IV-II-V-I, but within a larger phrase structure in which movement transformations occur as well – and in this manner we can generate the hybrid theme of the Mozart Haffner Minuet in Example 1.2-35. Therefore, the derivation of that theme would involve merging IV to II to create a IIP phrase, which would then merge with the structural V to generate a VP phrase. This VP would then merge with the final tonic *Stufe* to generate the Ton-C constituent, and therefore the continuation phrase of the Haffner’s hybrid theme, and the Ton-C would then merge with the Ton-P constituent to generate the entire theme – but not before the VP raises to the back-relating dominant position to create the antecedent phrase within the theme too. All of this can be visualized thus:



The above discussion also allows us to finally understand now the back-relating nature of the raised VP in the DOM-raising movement in more formal terms. In this movement, the VP moves to its final position

higher up in the tree, where it merges with a projection of the final I *Stufe* again. As a result, this moved *Stufe* can now c-command its original position, which allows the movement to satisfy the regulation on movement transformations discussed earlier. But because of the VP's merger in this particular position, it will not be the branching sister to *any* projection of the initial I *Stufe*, which means that it does not right-prolong this *Stufe* in any way. This implies that back-relating dominants do not really exist. However, the position the VP has moved to makes it the very next constituent immediately to the right of the initial I *Stufe* and its projections. So, it has moved to the *position* of a back-relating dominant (in terms of being right adjacent to the initial Ton-P), without actually right-*prolonging* that constituent. So the VP, in its moved position, can appear to fulfill the function of a back-relating dominant without actually being one – *this* is what I was implying, in my discussion of back-relating dominants and the structural V *Stufe* a few pages ago.

But notice an interesting structural connection the moved VP *does* have with the initial Ton-P constituent. As a daughter of the Ton-C constituent, which is a branching sister to the initial Ton-P constituent, the moved VP can be c-commanded by the initial Ton-P. In fact, the initial Ton-P fulfills three conditions simultaneously in a way no other constituent in the tree does:

- (a) It is a, or the projection of a, tonic *Stufe*.
- (b) It *precedes* the VP in its *moved* position.
- (c) It c-commands the VP in its moved position.⁶⁴

There are other constituents that fulfill one or two of these conditions, but none that satisfy all three of them. (For example, the final, cadential tonic *Stufe*'s first bar-level projection I' is a branching sister to the VP in its moved position, and so can fulfill (a) and (c) above – but it does not precede the VP.) Therefore, since only the initial Ton-P constituent satisfies all three of the above conditions, we can work this into a more formal definition of the back-relating dominant. That is, the phrasal projection VP of a

⁶⁴ “Precede” has a more technical definition in generative linguistics, according to which a constituent that precedes another constituent c-commands that second constituent too. (This is called the *Linear Correspondence Axiom* or LCA (Kayne (1994): 33).) So, if we adopt this definition of “precede” in (b), this automatically implies (c) as well.

structural V *Stufe* appears to (but does not actually) right-prolong a *Stufe*, as the back-relating dominant of that *Stufe*, only when that *Stufe* “is a tonic *Stufe*, or the projection of a tonic *Stufe*, that both precedes and c-commands the VP”. With this statement, the DOM-raising movement has now been properly defined, and the first part of my discussion of the movement-based nature of musical grammar has come to an end.

iii. DOM-raising and parametric differences between Classical and Rock music

There is, however, another, maybe even more important, way in which movement phenomena appear in music – and this will be the basis, in this subsection, for the second part of my discussion of the transformational nature of musical grammar. This has to do not with the way they might help us better explain certain structures *within* an idiom (such as interrupted forms in Western Classical tonal music), but with how they might help us better explain certain structures *across* idioms.

To begin, consider the fact that the Minimalist model of musical grammar we have developed so far has focused exclusively on just Western Classical tonality. This owes to the model’s foundation in Schenkerian thought, and Schenker’s own focus solely on Western Classical tonality in his theorizing. Now, just focusing on one idiom might not be a problem in and of itself as long as that idiom reveals the kinds of recursive and transformational phenomena that only a specifically generative model can explain. We see examples of this in linguistics all the time – for example in English, where a generative model is specifically needed to explain a phenomenon such as NP-movement, and the Case Filter that constrains it. Moreover, the presence of recursive and transformational phenomena even within a specific language or musical idiom provides us with a theoretically infinite number of structures to analyze and theorize about – and this also accords more generally with Cartesian and Humboldtian ideas about the infinite nature of the human mind and language of the kind Minimalism subscribes to. So, restricting generative theory to just one language or musical idiom is sufficient for a Minimalist research program.

But what if one believes that the language or musical idiom being described by generative theory does *not* have a recursive or transformational grammar? In such a situation, all the claims made by the theory about that musical idiom or language – in the way Schenkerian theory does for Western Classical

tonal music – will be taken as being purely speculative and empirically unfounded. This, in fact, is how the music theorist David Temperley evaluates Schenkerian theory in his recent critique of the system (in Temperley (2011): 157-163). A short digression to explore Temperley’s critique of Schenkerian theory might be worthwhile at this point, not only to see why extending the scope of Schenkerian theory beyond the Western tonal canon might help justify it, but also to correct some general (mis-) conceptions about Schenkerian theory in Temperley’s arguments, which can actually be found in the writings of various other authors too. Temperley evaluates Schenker’s ideas from two different perspectives, viz. as a theory of how listeners perceive grammatical structure in Western tonal music, and as a theory of how musicians create such grammatical structures when they compose music. In the last section of the previous chapter, I criticized the attempt to understand Schenkerian theory as a theory of musical perception, given that this system should be understood, instead, as a *cognitive* theory – of the *knowledge* of musical structure without which musical perception would not be possible to begin with. That Temperley even attempts a critique of Schenkerian theory on perceptual grounds might have to do with his own views of musical structure, which take a perceptual, and even physicalist, approach to understanding music – which I criticized in the last chapter as being inherently problematic in and of itself. But more importantly, such an approach misunderstands the essentially *anti*-perceptual basis for Schenkerian theory, making any attempt to refute, or even defend, Schenkerian theory on perceptual grounds a misguided enterprise.

Unlike his ‘straw man’ critique of Schenkerian theory as a theory of perception, Temperley does get the goals of Schenkerian theory right, however, when evaluating it as a theory of composition – i.e. as a theory of how musical phrases are generated, as opposed to perceived. After all, this is the perspective from which I have been defending Schenkerian theory as, or as the basis for, a (Minimalist) generative grammar of music.⁶⁵ But Temperley still finds Schenkerian theory to be flawed, even as a theory of

⁶⁵ As long as ones accepts the caveat that “composition” refers here not to the conscious, artistic acts of individual composers, but rather the unconscious, intuitive acts that any competent, native ‘speaker’ of a musical idiom can engage in (as long as the appropriate performance conditions obtain). (In this regard, see Brown (2005): 222-233.) This is an important caveat, in the light of Fred Lerdahl’s distinction between *compositional* grammars and *listening* grammars, which we explored in the last chapter. As Lerdahl rightly says, composers often compose music in idiosyncratic ways – in ways that involve the idiosyncratic, *compositional* grammar of an individual composer,

composition, on grounds of its using unnecessarily complicated theoretical constructs, which also happen to be empirically unwarranted. His main argument seems to be that Western tonal phrase structures can be modeled adequately with simple, non-hierarchically organized chord progressions, i.e. in a non-hierarchical finite-state model of tonal phrases, which is an argument made earlier by Dmitri Tymoczko (2003) as well. So, Schenker's attempt to reveal large-scale hierarchical, and often recursive, relationships between constituents is, in Temperley's opinion, an unnecessary theoretical extravagance. Moreover, since this theoretical apparatus does not seem to explain the structure of tonal phrases any better than the simple sequential model Temperley proposes, Schenker's system is empirically unwarranted too:

“The *Ursatz* reflects the principle that a tonal piece or tonally closed section typically begins with a I chord and ends with V–I, though the initial and final events may be separated by an arbitrarily long span of music. However, this dependency might also be captured in simpler ways, without requiring all the apparatus of a CFG [context-free grammar, i.e. the kind of grammar Chomsky proposed for natural language]. At local levels (for example, a tonally closed phrase or period), it could be modeled quite well with a finite-state model of harmonic progressions ... At larger levels, too, it is not clear what the *Ursatz* adds, in predictive power, to a more conventional model representing a piece as a series of key areas with a progression of chords in each key, beginning and ending in the same key, and ending with a perfect cadence in the home key. In short, it is unclear what is to be gained by modeling Schenkerian theory as a CFG. To justify this approach, one would need to find “long-distance dependencies” that are not predicted by more conventional principles of tonal theory. I am not convinced that such long-distance dependencies exist.” (Temperley (2011): 152-153)

Now, the claim that hierarchical, long-distance dependencies do not exist in music has to be considered as partly polemical. After all, one often sees a similar rejection of hierarchical, long-distance dependencies in language too, which is usually motivated by an ideological commitment to some form of anti-Rationalist, anti-nativist belief system (see for example Elman et al. (1996)) – despite the large amount of

rather than the more general listening grammar possessed by native listeners in an idiom. Therefore, a theory of composition that focuses on compositional grammars cannot have the sort of generality that a truly scientific music theory aspires to. (Indeed, it is for this reason that more anti-scientifically-inclined music theorists interpret Schenker's ideas as describing the idiosyncracies of individual composers – that, to their mind, precludes a more scientific interpretation (see Dubiel (1990) for an example of this).

One could argue, however, that Schenkerian theory does not describe the idiosyncratic composing habits of individual composers, since it describes the general structural properties of a vast swath of Western music. In this sense, it is closer to what Lerdahl calls a “natural” (versus an “artificial”) compositional grammar (Lerdahl (1992): 100-101) – it is a grammar of the natural musical ‘language’ of Western common-practice tonality. Lerdahl has no problem with natural compositional grammars as objects of serious inquiry, although he thinks that such an inquiry should be based on a study of listening grammars first – which led to Lerdahl and Jackendoff's perceptual/reductive approach to tonal structure, and their subsequent break from a more Schenkerian, generative perspective. That approach has its own problems, which I examined in the last chapter, which to my mind justifies persisting with a (natural) compositional approach to musical structure based on Schenkerian theory.

evidence in favor of hierarchical structure in language, some of which we have looked at in this chapter, and despite the fact that accepting such structure allows us to develop a theory of language that has considerable explanatory power. In the last chapter, I discussed the four Ps of music scholarship – perception, physicalism, pedagogy, and poetics – that act as ideological barriers to pursuing a generative approach to the study of musical structure. I believe that part of the reason for denying hierarchical structure in music has to do with one of the above ideological motivations. It would be unfair for me to speculate about what Temperley’s motivations are, but it seems that part of it at least has to do with his commitment to a perceptual explanation of musical structure – seen in his above evaluation of Schenkerian theory as a theory of perception, and in his earlier work in the listening grammar tradition (e.g. Temperley (2001)). Part of it might also have to do with an interest in mathematical explanations of musical structure (e.g. Temperley (2004a, 2007)), which he shares with Dmitri Tymoczko (Tymoczko (2011)), and which lies at the heart of the physicalist paradigm in music theory.

However, ideology aside, the fact remains that there are many theorists of an anti-Schenkerian bent, like Temperley, who deny that Western tonal phrases are structured hierarchically, or display long-distance dependencies. This is a problem if one believes that Schenkerian theory provides unique insights into Western tonal structure, or if one believes that the Minimalist interpretation of Schenkerian theory provides unique insights into tonality in general – as this dissertation does. This is because the explanatory power of Schenkerian theory lies in its ability to reveal the complex hierarchical organization of grammatical phrases in the Western tonal idiom – but if the very existence of hierarchy or long-distance dependencies is denied to Western tonality, then this reduces much of the explanatory force of Schenkerian theory or musical Minimalism in general. This is why focusing on just the one idiom of Western tonality is problematic for generative approaches to musical structure, even if it is sufficient for generative theorists to make the case they want to make for musical generative grammar. In other words, focusing on just one idiom can be rendered ineffective if other music-theoretic approaches deny the

existence of the very features (like hierarchy) that are supposed to make one's explanation of that idiom superior to rival ones.

For this reason, *cross-idiomatic* evidence is needed to support a theory of musical structure, especially if the language or musical idiom that has been the sole focus of prior study provides evidence for that theory only controversially. For this reason, if Schenkerian theory *is* a theory of composition, a theory of how phrases are generated in musical idioms, it is of paramount importance that evidence from other idioms be brought to bear on it – even if this means extending it beyond Schenker's own, initial, focus on the Western Classical tonal idiom. To compare this with linguistics, this is precisely why the generative study of language has become an increasingly cross-linguistic research program. Cross-linguistic research has allowed generative linguistics to test its claims against a wealth of data from across the world.⁶⁶ In the process, this has allowed it to defend its claim that C_{HL} is a universal system, innately present in the minds of all humans by virtue of our shared biology, because the data from cross-linguistic research suggests that sentences across all languages seem to be generated by this system (despite its generating different S-structures across languages).

So, David Temperley's criticism of Schenkerian theory as a theory of composition can be easily falsified by demonstrating hierarchy, recursion, long-distance dependencies etc. in other musical idioms, which only a generative approach based in Schenkerian theory might be able to explain. Alternatively, one could show how Temperley's anti-generative model is unable to account for important cross-

⁶⁶ One example of this can be found in how generative theory deals with the unusual, verb-initial, word order of Irish sentences in comparison with other theories of sentence structure. Some, particularly anti-generative, approaches have suggested that the verb-initial order of Irish sentences implies that Irish sentences do not have a hierarchical, generative structure – similar to how Temperley rejects hierarchical structure in Western tonal phrases. More specifically, 'anti-generativists' have claimed that Irish sentences have a non-hierarchical 'flat' structure, in which the verb, subject, and object are all branching sisters to each other, in a *ternary*-branching tree, rather than being hierarchically organized into a binary-branching tree of the kind proposed by earlier X-bar forms of generative theory. But if this is the case, then each branch of such a tree should be able to c-command its branching sisters, meaning that the subject NP of an Irish sentence should be able to c-command and *be c-commanded by* the object NP of the sentence. And given that both the subject and object NPs are in the same clause, i.e. the main clause of the sentence, they should be able to bind each other as well. But as Andrew Carnie argues, this implies that Irish versions of the sentences "Saw Sheila herself" and "Saw herself Sheila" should both be grammatically correct – but this is not true, because the subject NP must c-command the object NP in Irish sentences, but the opposite leads to ungrammaticality (Carnie (2002): 200-201). This suggests that Irish sentences do *not* have a flat structure, thus falsifying anti-generative claims in this regard. Similar cross-idiomatic data can therefore help defend music-theoretic assertions too – it can provide vital evidence to decide between rival theories of musical structure.

idiomatic similarities between musical idioms, which only a universal generative grammar of music, again based on Schenkerian theory, might be able to explain. And it is here that movement transformations – the focus of this section – play the greatest role in supporting a generative, and specifically Minimalist, approach to musical or linguistic structure. As we have seen in earlier sections, the Minimalist emphasis on transformations, i.e. the workings of *internal Merge*, can explain how two different S-structures can be generated from a similar D-structure. In this manner, the different S-structures seen in paradigmatic English versus Irish sentences, with their different word orders, can be explained as the result of movement transformations like NP-movement in English, which transforms the same D-structure (of the kind proposed by the VP-internal subject hypothesis) into different S-structures in the two languages.

This implies that a comparable Minimalist Program for music, based in Schenkerian ideas, should ideally be a cross-idiomatic research program – it should propose hypotheses about how the musical mind, i.e. C_{HM} , works across idioms, and should test such hypotheses through an analysis of musical phrase structures across idioms too. This would allow us to compare Temperley's non-hierarchical model of musical phrase structure with a hierarchical Schenkerian one based on music outside of the Western tonal canon, and would therefore allow us to test the validity of his claim that his non-hierarchical model deals more adequately with the data than a Schenkerian one does. Moreover, such a cross-idiomatic research program might provide us with more evidence for the claims made by the generative model I have been proposing, such as the claim that musical grammar is transformational too.

Such a cross-idiomatic exploration of musical structure will be the focus of the rest of this section. Specifically, I will examine the harmonic and phrase structure of musical phrases in the rather different idiom of Rock music, on the basis of which I will argue that only a generative approach to music, based in Schenkerian ideas, can explain both these Rock structures *and* the structures of Western Classical tonal phrases – thus falsifying Temperley's contention that a non-hierarchical model does a better job of explaining 'the data' than a hierarchical, Schenkerian one does. In addition, I will show that the ability of

a generative model to explain both Western Classical and Rock phrase structure simultaneously *depends on the existence of movement transformations in music too*, just as similar transformations in language allow generative models to explain cross-linguistic phrase structure more adequately than non-generative ones can, for example in the way NP movement can explain the differences between English and Irish sentences.

So, what is it about Rock music that makes it so hard for a finite-state model, like the one Temperley proposes, to simultaneously explain phrase structure in this idiom and in Western Classical tonality? The answer to this has much in common with the way sentences in different languages have their words ordered in different ways, such as the ordering of subject NPs and VPs in English versus Irish sentences. We have seen how Western Classical tonal phrases are generated by merging chords in an ‘anticlockwise’ direction in the circle of fifths, which leads, for example, to the canonical V – I phrase (Keiler’s *Tonic Completion*) in this idiom. In Rock music, however, phrases are often generated by merging chords in a *clockwise* direction in the circle of fifths, leading to phrases in which chords appear in the *exact opposite order* in which they appear in Western Classical tonal phrases.⁶⁷ For example, when a G and a C chord merge in Rock music, the G chord rather than the C one often heads the resulting phrase. Keeping Schenkerian left-branching intuitions in mind, this leads to the non-head C appearing as a left-branching sister to the head G chord – which means that the resulting phrase is linearized as C – G, rather than G – C as happens in Western tonality. And this yields a IV – I progression in G-major, a progression that is as common and popular in Rock music as V – I is in Western tonality, and which far outweighs V – I as the canonical progression in Rock music too.

This means that chords normally appear in grammatical Rock phrases in an *ascending fifths* (or circle of *fourths*) order, rather than the typical *descending* fifths order of chords in Western tonal phrases. (Of course, ascending fifths chord progressions can be found in Western tonal phrases too, but these are

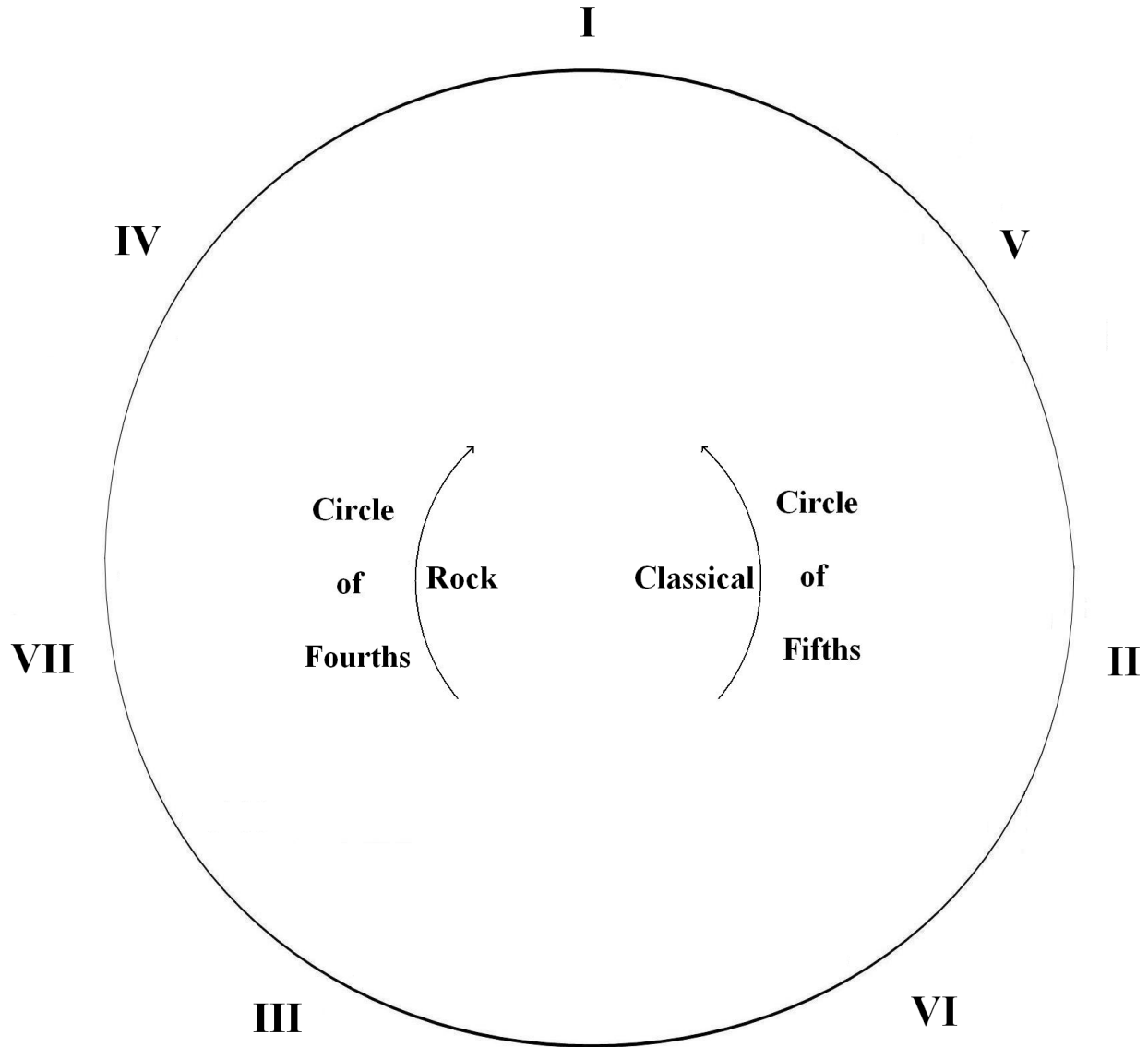
⁶⁷ It is worth pointing out that I am taking “Rock music” here as referring to the music of Blues-based guitar bands, such as the Rolling Stones, Led Zeppelin etc. As such, Rock music is a very diverse idiom, with a wide variety of influences – including Classical music, in the case of some (more keyboard-driven) performers. So, one can find Classical tonal harmonic practices in Rock music too – although the more Blues-based strain within Rock music can be distinguished from this, which is the focus of my investigation here.

usually part of an ascending fifth *voice-leading sequence*, where chords do not have functional, grammatical status.) So, we could say that Rock chords are often related in terms of “circle of fourths” features – or rather, they are related in a circle of fifths fashion like Western tonal chords, but in the opposite direction, i.e. going clockwise (i.e. ascending) in the circle rather than anti-clockwise as is the case in Western tonality, and as was discussed earlier in this chapter. This idea is illustrated in Example 1.2-36.

As the example illustrates, the opposite ordering of chords in Rock music leads to chord progressions that are very atypical in Western Classical tonality, but that are canonical in Rock music. Whereas traveling in an anti-clockwise direction on the circle leads to descending fifths progressions like II-V-I, which are typical in Classical tonality, traveling in a clockwise direction gives us ascending fifths of circle of fourths progressions like VII-IV-I or III-VII-IV-I, which are omnipresent in Rock but rare in Classical tonality outside of sequences. This difference in chord ordering between Rock and Classical tonality is similar to the different, and often opposite, ordering of words and phrases in different languages, as we have seen before. Importantly, generative linguists often explain such differences in parametric terms, as we saw in section 1.2.3. So, it could be that the above chord-order differences between Rock and Classical tonality are evidence for parametric differences between musical idioms too – which would be important evidence for a Principles and Parameters-type generative theory of music.

I will explore this significant theoretical point in a moment, along with an examination, as promised earlier, of what role transformations play in all of this. But let us look at more data for this chord-order difference between Rock and Classical tonality, just to hammer in the reality of this difference, and its foundation in circle of fifths directionality issues. Example 1.2-37 lists four kinds of chord-order differences between Rock and Western Classical tonality, all of which relate to the fundamental difference in how progressions are built from the circle of fifths in these idioms. The first difference lies in what the example calls “cyclic progressions”. These are nothing but straightforward circle of fifths progressions, of the kind that were being discussed in the last few paragraphs. This is clear

Example 1.2-36. Parameters in music: “Circle of Fourths” vs. “Circle of Fifths” settings



Example 1.2-37. Parameters in music: A list of examples from Rock music vs. Classical music

Circle of Fourths parametric ‘setting’
(Rock music)

Cyclic progressions: [V-II-VI-III-VII-IV]-I

Jimi Hendrix (“Hey Joe”, VI-III-VII-IV-I)
Skid Row (“Eileen”, VI-III-I)
Rolling Stones (“Jumping Jack Flash”, III-VII-IV-I)
Guns N’ Roses (“Right Next Door To Hell”, III-VII-IV)
Lynyrd Skynyrd (“Simple Man”, III-VII-I)
Audioslave (“Like A Stone”, III-VII-I)
Van Halen (“Unchained”, III-VII-I)

Opening progressions: I-VII-IV

Lynyrd Skynyrd (“Sweet Home Alabama”)
Steppenwolf (“Magic Carpet Ride”)
Led Zeppelin (“Good Times, Bad Times”)
Rolling Stones (“Midnight Rambler”)
The Beatles (“Get Back”)
AC/DC (“Shoot to Thrill”)
Def Leppard (“Animal”)
Green Day (“Welcome to Paradise”)

‘Subdominant’ expansions: II-IV

Thin Lizzy (“The Boys Are Back in Town”)
4 Non Blondes (“What’s Up”)
Def Leppard (“Hysteria”)
Guns N’ Roses (“Bad Obsession”, II-VII-IV)

Model phrase progressions: I-V-IV-I

Rolling Stones (“Dead Flowers”)
Aerosmith (“Angel”)
The Who (“Baba O’Riley”)
The Black Crowes (“Descending”)

Circle of Fifths parametric ‘setting’
(Western Classical tonal music)

Cyclic progressions: [IV-VII-III-VI-II-V]-I

Any descending 5ths progression. Numerous examples in the tonal literature.

Opening progressions: I-II-V

Many, if not most, antecedent phrases ending with a half cadence in the Classical period literature.

‘Subdominant’ expansions: IV-II

Any expansion of subdominant harmony through 5-6 voice leading motion. Numerous examples in the tonal literature.

Model phrase progressions: I-IV-V-I

Numerous examples in the tonal literature, especially in phrases ending with an authentic cadence in the 19th century tonal literature.

from the right side of the example, which deals with the Classical tonality side of the equation, where “cyclic progressions” are exemplified by the standard descending fifths progressions omnipresent in this idiom. The reason I am calling them “cyclic” is because such progressions often cycle through the complete circle of fifths, usually starting and ending with the tonic triad (i.e. I-IV-VII-III-VI-II-V-I), which yields a descending fifth sequence (which can be seen, for example, in mm. 18-21 of the first movement of Mozart’s K. 545 piano sonata). The three last chords of the progression, i.e. II-V-I, are seen

more frequently in isolation from the entire sequence in tonal phrases, where they either prolong tonic harmony at the beginning of a phrase, or where they constitute an authentic cadence at the end of a phrase. (William Caplin calls the former an example of a “prolongational progression”, and the latter an example of a “cadential progression”, see Caplin (1998): 25-29.)

In contrast, cyclic progressions in Rock music, as illustrated by the left side of Example 1.2-37, tend to involve chords that progress by ascending fifths – thus reversing the order in which chords appear in Classical tonal circle of fifths progressions, i.e. to I-V-II-VI-III-VII-IV-I. Now, this entire progression is not normally seen in a Rock phrase, but parts of it certainly are. One reason for this might have to do with the fact that Rock harmonies are often derived from the Blues scale, unlike Classical tonal harmonies, which are derived from the major and minor scales. The Blues scale is made up of scale degrees 1, 3, 4, 5, and 7, so it is more common to find only that part of a cyclic chord progression in Rock made up of chords whose roots correspond to these scale degrees. As a result, the progression made up of the last four chords of an ascending fifths progression that returns to tonic harmony (i.e. III – VII – IV – I) is more common in Rock music (e.g. in the chorus of the Rolling Stones’ “Jumping Jack Flash”) because it avoids using the VI chord, which would appear right before the III chord that begins the progression – but which is also a chord built on the non-Blues scale degree 6. Frequently, the progression III – VII – IV – I is found with one or more chords omitted too, such as in the cyclic progressions in the Guns N’ Roses and Van Halen examples shown in 1.2-37. But sometimes Rock songs make use of an extended Blues scale, in which scale degree 6 is included, which does allow the chord progression VI – III – VII – IV – I, one of the best examples of this being Jimi Hendrix’s “Hey Joe”, which just cycles through this progression again and again for the entirety of the song.⁶⁸ In any case, the main point here is that cyclic

⁶⁸ One conclusion that might be derived from this data is that the ascending fifths progressions seen in Rock music, like VI – III – VII – IV – I, are not sequential progressions, but functional ones. This is why the complete sequence I – V – II – VI – III – VII – IV – I is so rarely seen in Rock music, as opposed to Classical music, where it is relatively common. A sequential progression arises from voice leading, rather than from functional harmony, and can therefore include chords in what appear to be ungrammatical orders. In Classical tonal harmony, an ascending fifths progression is, strictly speaking, ungrammatical – no functional harmonic progression in Classical tonality involves chords whose roots ascend by fifth. But it is precisely because the ascending fifths *sequence* is *not* a functional harmonic progression that it can occur relatively frequently in Classical tonal phrases. In contrast, the ascending

progressions, whether stemming from a simple or extended Blues scale, or a diatonic major or minor scale, are ubiquitous in both Rock and Classical tonal music – but they involve chords that are ordered in exactly opposite ways in the two idioms, due to the direction in which these chords progress in the circle of fifths in each idiom.

The second kind of chord-order difference between Rock and Western Classical tonality involves what Example 1.2-37 calls an “opening” progression. This is the kind of progression that normally begins a phrase, by prolonging the initial tonic of the phrase (i.e. Keiler’s “Tonic Prolongation” constituent). In Western Classical tonality, this is the kind of chord progression that might harmonize an antecedent phrase (such as I – II – V), ending in the half cadence that ends the phrase – but without the final tonic triad at the end to give the phrase an authentic cadential ending. (In other words, this is *not* the chord progression that would yield Keiler’s “Tonic Completion” constituent, which is something one would normally see in *consequent* phrases.) Such opening progressions often have a circle of fifths origin too – e.g. the II – V part of the I – II – V progression just mentioned involves descending (i.e. anti-clockwise) motion in the circle of fifths. It is just that such progressions do not complete the motion around the circle of fifths back to the tonic, as cyclic progressions do. So, opening progressions are essentially just the initiating part of a cyclic progression.

But it is worth distinguishing them from cyclic progressions because of the way they help distinguish Rock from Classical tonal harmony. In both idioms, one finds opening progressions that begin with tonic harmony, followed by a circle of fifths progression. In Classical tonality, this gives rise to common progressions like the I – II – V progression just mentioned. However, it is precisely here that Rock harmony again shows opposite directionality – not by reversing the I – II – V progression, as was the case with the above cyclic progressions, but by following the initial tonic harmony by chords that are directly opposite on the circle of fifths. So, instead of I – II – V, we would see a progression like I – VII – IV. Here, the VII in the latter progression is the same distance from I in the circle of fifths as is II in the

fifths progression *is* grammatical in Rock music, and arises from functional Rock harmony, which is why there are restrictions on what chords can appear in it, and in what order – making the *complete* ascending fifths progression rare in this idiom, for the reasons discussed above.

former progression, i.e. two steps away – but from the opposite direction (review Example 1.2-36 to confirm this). Similarly, IV in the latter progression is a fifth away from I in the circle of fifths, just like the V in the former progression, but again from the opposite direction – i.e. the root of IV is a fifth lower than the root of I, whereas the root of V is a fifth higher than the root of I.

The reason this distinction between opening progressions in Rock and Classical tonal harmony is important is not only because it shows the opposite ordering of chords in the circle of fifths in these two idioms, but also because these progressions are *extremely* common in these idioms – therefore demonstrating how common this difference between the two idioms is too. I – II – V is very common in Classical tonality, as the typical harmonization of an antecedent phrase. But I – VII – IV is equally if not more common in Rock music, as the opening progression of countless Rock songs – the most famous examples being the intro to Lynyrd Skynyrd’s “Sweet Home Alabama” and Steppenwolf’s “Magic Carpet Ride”. (This progression is often used at the beginning of choruses in the middle of a Rock song too, if not right at the beginning of the entire song – the AC/DC and Def Leppard opening progressions cited in 1.2-37 being cases in point.) So, again, Rock and Classical tonal harmony seem to be distinguishable because of another important chord-order difference in their phrase structures.

The third chord-order difference between Rock and Classical tonal harmony has to do with the predominant harmonies that exist within larger, circle-of-fifths based chord progressions. In Classical tonal harmony, multiple predominant harmonies can precede the V chord that then leads to I by descending fifth root motion. IV and II are the most common triadic examples, and they often appear in that order too, i.e. IV – II – V – I, as we saw in the Mozart Haffner Minuet passage in Example 1.2-35. Now, the II of this progression is related to the following V by descending motion in the circle of fifths, as is the V to the final I, but the initial IV is related to the following II by root motion in thirds. In section 1.2.4.i, I proposed a way of thinking about these third-based chord relations – but that is not particularly relevant here. What is important here, again, is the order in which these chords appear, with the IV *preceding* the II. We can refer to this as the order in which *subdominant harmony is expanded* in Classical

tonal phrases, which often takes place via a 5-6 motion between the bass and an upper voice, as we have seen before.

And again, an important order difference exists between the above Classical tonal practice and Rock music, because in the latter it is quite common to see IV *following* rather than preceding II, as the examples on the left side of Example 1.2-37 illustrate. So, “subdominant expansion” in Rock music also reveals a chord-order difference compared to Classical tonality. Importantly though, the IV – II subdominant expansion normally precedes dominant harmony in Classical tonal phrases (as we saw in the Mozart Haffner example), but the opposite II – IV subdominant expansion does not normally lead to V in Rock music. In fact, the function of IV in Rock phrases seems to be analogous to that of V in tonal phrases, as we saw in the opening progressions I – II – V (in Classical) versus I – VII – IV (in Rock). This means that the IV – II subdominant expansion in Rock is *not a predominant phenomenon*, as it is in Classical tonality. If anything, the II in a II – IV subdominant expansion in Rock might be considered a “presubdominant” leading to IV, analogous to how a *predominant* II leads to V in Classical tonality. For this reason, the presubdominant II in Rock can be further expanded (e.g. by VII, see the Guns N’ Roses “Bad Obsession” example in 1.2-37), just as the predominant II in Classical tonality can be further expanded by (or can itself expand) IV, in the phenomenon of subdominant expansion we just discussed. Compare:

Rock		Classical
II – IV – I	versus	II – V – I
II – VII – IV – I	versus	IV – II – V – I

The final chord-order difference between Rock and Classical music lies in what Example 1.2-37 calls “model phrase progressions”. In Classical music, an example of such a progression is I – IV – V – I, as the bottom right side of the example illustrates. This progression represents, of course, the typical Tonic-Predominant-Dominant-Tonic functional chord progression that forms the basis for so many tonal phrases, especially in the 19th century, although I – II – V – I was more common in earlier eras of

Classical composition. For this reason, such progressions are the basis for model tonal phrases in Classical music – which gives these progressions their name.

In Rock music though, another chord-order difference exists, in the well-known fact that IV often *follows* V in this idiom – as opposed to preceding it, as happens in Classical model phrase progressions. This yields the model I – V – IV – I progression ubiquitous to Rock music phrases, and which can be traced all the way back to the Blues too. Notice that in this Rock progression IV appears right before the final tonic harmony, just as *dominant* harmonies normally would in Classical tonal phrases – which reaffirms the analogy I made a few paragraphs ago between IV in Rock and V in Classical music. So, even though the model Rock phrase has both IV and V harmonies, in these phrases IV appears in the place that V normally would in a Classical tonal phrase. This point not only illustrates the significance of chord-order differences between Rock and Classical phrases one last time, it is a point that will be relevant to our impending discussion on transformations.

In the face of all of the above chord-order differences between Rock and Classical music, one thing is clear – no finite-state model, like the one proposed by David Temperley, could ever model this data. A finite-state machine could model the grammar of an idiom, by computing the greater probability of, say, V following, rather than preceding, IV, or the greater probability of V, rather than III, following IV in that idiom – both of which do usually happen in Classical tonal phrases. To this extent, Temperley might seem to have a point in asserting that a finite-state model is good enough to describe how Classical tonal phrases are generated. But including Rock data in the picture changes everything. That is, as soon as we try to use the same finite-state model to describe phrase structure in *both* Classical and Rock phrases, we run into insurmountable problems. For in this picture, IV is *as* likely to precede V as it is to follow it (the former happening in Classical, the latter in Rock, phrases). Moreover, the model would have to compute the fact that II in this scenario, is as likely to be preceded by IV and followed by V (as happens in the canonical IV-II-V-I Classical phrase, e.g. in the Mozart Haffner Minuet theme), as it is to be preceded by V and followed by IV (as happens in certain Rock phrases harmonized by the I-V-II-IV progression, e.g.

in Buckcherry's "Slit My Wrist"). So, a finite-state model would have to model both a specific set of chord orders and the exact opposite set of orders *simultaneously*, in order to be able to model how phrases are generated in both Classical and Rock idioms – which is impossible.⁶⁹

So, a cross-idiomatic comparison of Classical and Rock harmony makes it clear that a finite-state approach to modeling musical grammar does *not* do a better job of explaining 'the data'. This would appear to falsify David Temperley's anti-generative assertions about Schenkerian theory – but of course, this would be so only if we can demonstrate that a generative approach, based on Schenkerian theory, *can* actually model this cross-idiomatic data, in a way that a finite-state model cannot. And it is here that a theory of transformations will come to our rescue too, since there have been attempts to understand Rock harmony from a Schenkerian perspective, the results of which have sometimes been unsatisfactory – my contention being that it is the lack of transformations in these previous Schenkerian attempts to describe Rock harmony that leads to their at times unsatisfactory results. Once transformations are invoked in Schenkerian analyses of Rock music – which I claimed in the last section as being something Schenker implicitly did himself in his analyses of Classical tonal passages – then this approach can successfully model both Classical and Rock music in a way that no finite-state approach can.

The most famous Schenkerian theory of Rock music is arguably Walter Everett's work in this area, particularly in his Schenkerian analyses of the music of the Beatles. Everett takes a more traditionally-Schenkerian approach to Rock music, in which he views this idiom as being just an

⁶⁹ One way around this is to assert that a separate finite-state model should be proposed for Rock harmony, separate from the one that, for example, Temperley proposes for just Classical tonal harmony. In other words, one should not attempt to model Rock and Classical tonal harmony simultaneously, not only because a finite-state model can never achieve this (for the reason discussed above), but also because Rock and Classical music, in this view, are orthogonal systems, and should therefore be modeled with separate grammars. This is in fact the solution Temperley subscribes to himself (personal communication). But this is also a solution that reveals an extreme anti-universalist position, one that is inconsistent with a generative approach to music. Such an anti-universalist position also involves believing that every musical idiom is associated with a separate mental faculty, and must therefore be learned from one's environment or culture (since it does not make sense to assume that each and every musical idiom has been separately hardwired into our minds – for this would imply, putting it very crudely, a separate gene for Classical music, Rock music, and every other possible idiom, which is of course nonsense). This further implies an anti-nativist, Empiricist attitude towards music, the mind, and human nature. Not only is such a position clearly incompatible ideologically with the current Minimalist project, it is rife with problems and inconsistencies, as I argued at length in the last chapter – and is therefore not a viable solution, in my opinion, to the problem of getting a finite-state model to describe Rock grammar.

extension of Classical tonality (as opposed to a separate idiom circumscribed by its own, separate, grammatical parameters). An example of this can be seen in his analysis of the Beatles' "With a Little Help From My Friends" from the *Sgt. Pepper* album (Everett (2001): 102-104). The chorus of this song is made up of the standard Rock ascending fifths (or circle-of-fourths) cyclic progression VII-IV-I. As we just saw in the last few pages, this progression is ubiquitous in Rock music, but it also reverses the standard circle-of-fifths order that governs chord progressions in Classical tonal phrases. This can pose problems when applying Schenkerian methods to Rock phrases, since those methods are designed to reveal the structure of *tonal* phrases, which are normally harmonized by *descending* fifths progressions. For example, the descending fifth progression V-I plays an important role in Schenkerian descriptions of Classical tonal phrase structure, since it makes up the second half of the Schenkerian *Ursatz*, and also Allan Keiler's "Tonic Completion" constituent. So, in a standard Schenkerian analysis it is this progression that usually harmonizes the descent from scale degree 2 to scale degree 1 in the *Urlinie*. But in an idiom like Rock music, where V-I type descending fifths progressions are rare, a similar harmonization of a descent from scale degree 2 to 1 in the *Urlinie* will not be available. In a Rock passage harmonized with a VII-IV-I progression, for example, the final I will of course harmonize scale degree 1 in the *Urlinie* of the passage, but IV is not able to harmonize scale degree 2 – even though its role in Rock phrases is analogous to that of V in Classical tonal phrases.

In light of this problem, Everett suggests in his analysis of "With a Little Help" that it is the VII of the VII-IV-I progression in the song that harmonizes scale degree 2. Now VII *can* harmonize scale degree 2 – but this would make VII-I (without the IV) as the fundamental harmonic progression of this passage, which supports the descent from scale degree 2 to 1 in the passage's *Urlinie*. Problem is, this would be akin to saying that II-I, without the V (in a II-V-I progression) is the fundamental progression of a Classical tonal passage, since II can support scale degree 2 as well. But of course, this would completely misunderstand how tonal harmony works – so, for the same reason treating VII as the support for scale degree 2 in a Rock phrase harmonized by VII-IV-I seems to misunderstand how Rock harmony works – VII merely prolongs IV, with IV-I being the fundamental progression that harmonizes a Rock passage.

My proposal here, contra Everett, is to suggest that Rock harmony is *not* an extension of Classical tonal harmony – meaning that the typical descending *Urlinie* found in tonal phrases might not be found in Rock phrases, and Rock phrases will also, therefore, not reveal a I – V – I *Ursatz* structure. But this is not to say that Rock and Classical tonality are orthogonal systems, as the anti-generative theorist might say – my claim instead is that phrases in both idioms are generated according to general (perhaps universal) principles of musical grammar, it is just that these principles are parameterized in different ways in different idioms, just as happens in language. So, I would argue that Rock and Classical phrases are both governed by a circle of fifths principle – according to which chords are merged to form phrases in both idioms according to certain “circle of fifths” features, of the kind I described in earlier sections. However, this principle is parameterized differently in Rock and Classical music, leading to the different directions in the circle of fifths that chord progressions follow in these two idioms – ‘anti-clockwise’ for Classical and ‘clockwise’ for Rock music – and which yields the chord-order differences between the two idioms discussed in Example 1.2-37. This different *linearization* of chords in the two idioms also implies that analytical concepts that depend on directionality, such as the Schenkerian *Urlinie*, will not manifest themselves in the same way in both idioms – which is why one cannot find a typical descending tonal *Urlinie* in Rock phrases, or why one runs into trouble if one tries to find one, as Everett does in his analysis of “With a Little Help”. However, one might find a different kind of *Urlinie* in Rock phrases, perhaps a scale degree 7 – 6 – 5 descending line, which can be harmonized by the VII-IV-I progression, and which would not involve emphasizing the VII over the IV as Everett does.

This might lead to us proposing a different *Ursatz* form for Rock music too – perhaps a I – IV – I structure, which treats IV in Rock phrases as analogous to V in Classical tonal phrases (as I have argued for above), making the I – IV – I Rock *Ursatz* analogous to the I – V – I *Ursatz* Schenker proposed for tonal phrases. This would fit in with the general model of grammar I have been proposing in which the *Ursatz* is not taken to be a primitive in an axiomatic system, but rather a complex entity that arises from a more basic Merge-based generative procedure. So, when this generative procedure generates complex entities, it can generate different *Ursätze* too, in different idioms, depending on how these idioms are

parameterized. So, rather than focusing on what the *Ursatz* or *Urfinie* of a Rock passage is, my proposal is to focus on the common generative procedure that yields both this passage and a passage in a Classical tonal piece. Such an approach is inherent in how Minimalism views linguistic structure too – and is precisely the kind of approach anti-generativists like David Temperley reject for music. However, it is also the kind of approach that has *not* been the focus of extant Schenkerian Rock theory. As I have mentioned earlier, it has not been the focus of extant Schenkerian Rock theory because the notion of a transformation has not been brought to bear on this issue. So, now I turn to discussing how this might be done.

Let us return to Example 1.2-34 for this discussion. We have already examined the grammatical tree (a) on the left side of this example, which illustrates how a dominant-phrase raising movement transformation might occur in Classical music, to generate the kinds of interrupted forms one sees in Classical periodic structures. However, this tree only represents the structure of such Classical tonal phrases – it cannot account for the phrases of Rock music, with its opposite-ordered chord progressions. To address this, the tree (b) on the right side of Example 1.2-34 describes how a typical Rock phrase might be generated. The larger structure of the tree is identical to the one on the left, being comprised of an initial Tonic Prolongation branch, and a final Tonic Completion branch, which merge to generate the entire phrase, whose head is the final tonic triad of the phrase.

The difference between the two trees in the example lies primarily in the branching structure of the final Tonic Completion branch. In Example 1.2-34a, the final tonic triad head of the entire tree was merged with a Dominant Prolongation phrase (VP) to generate a Tonic-bar level of representation. This Tonic-bar level representation was then merged again to generate the Tonic Completion constituent. In Example 1.2-34b, however, the Tonic-bar level representation arises by merging the final tonic triad head, not with the Dominant Prolongation phrase, but with a *Subdominant Prolongation* phrase (IVP), given the analogous role of such IV harmonies in Rock compared to structural V harmonies in Classical tonality. This IVP constituent now reveals the reverse chord-order direction of Rock phrases, since it is generated according to circle-of-fourths harmonic factors. This is why, as the tree illustrates, it is generated by

merging the IV head of the constituent with a VIIP phrase, with which it has a circle of fourth relationship – of the kind we see in VII-IV-I cyclic Rock progressions. If you remember my earlier discussion about these kinds of mergers in section 1.2.4.i, they involve *negative* cf values – as opposed to the positive cf values involved in Classical tonal phrase generation, of the kind I explored in my generation of the *Tebe poem* phrase by Dmitri Bortniansky in Example 1.2-30.

Now, the VIIP phrase that merges with IV to generate the IVP phrase, is itself generated by merging the VII head of this phrase with a IIIP constituent, again due to their circle-of-fourths relationship. And the IIIP constituent is generated by merging the III head of this phrase with a circle-of-fourths related VI phrase – and so on, until we get a complete cyclic progression, which looks like:

I_(Tonic Prolongation) – (V-II-VI-III-VII-IV-I)_(Tonic Completion)

The above structure is what the tree in Example 1.2-34b represents. So, in this way the harmonic progression underlying a complete Rock phrase can be generated – in a manner identical to how Classical tonal phrases are generated by this model, but after accounting for the important chord order difference between these models. As you will notice, this results in V chords being treated rather differently in Rock phrases compared to Classical tonal phrases. In the tree, V enters the structure only by merging as a Dominant phrase with a II triad – according to circle-of-fourth factors – to generate a II phrase in which the II triad is head. This reveals the relative structural *unimportance* of V chords in Rock harmony compared to Classical tonal harmony.

Now, a V chord can, of course, be part of a Rock harmonic progression without it first merging with II – in fact, that is how we get the standard Rock model phrase progression I-V-IV-I. Here, V merges directly with IV, in the way a predominant IV (in a Classical phrase) can merge directly with V in the tonal progression I-IV-V-I. So, this is a shortcoming of the tree in Example 1.2-34b. However, this shortcoming arises, as you might remember, from the *representational*, as opposed to *derivational*, view of music structure depicted in the example. A derivational approach, which is more consistent with the Minimalist view on these things anyway, would allow us to merge anything with anything, including V

directly with IV in Rock progressions – just as it allows a predominant IV to be merged directly with V in a Classical tonal progression, which is something I discussed earlier, in my examination of the Classical phrase tree shown in Example 1.2-34a.

What this suggests is that within my current generative model a straightforward I-V-IV-I progression can be generated, in a way that is identical to how a similar I-IV-V-I tonal progression is generated, but in a way that also accounts for the important, parametric, chord-order differences between these idioms. This, to my mind, represents an improvement over the more traditionally-Schenkerian approach favored by theorists like Walter Everett. However, Example 1.2-34 also presents *different* models for Classical and Rock phrase structures, presumably because it generates the Classical phrase from a typical I-V-I *Ursatz*, but switches to a I-IV-I *Ursatz* for the Rock phrase, to account for this idiom’s different ordering of chords. A truly Minimalist generative model should be able to generate both the Classical and Rock trees shown in Example 1.2-34 from a simpler, shared foundation – where only Merge is the relevant factor.

This is where the role of transformations becomes so important, because it is through a movement transformation (i.e. through an application of internal Merge) that *both* trees in Example 1.2-34 can be generated, from a common foundation. But before we explore this important conclusion, recall that a transformational element already exists in the trees in Example 1.2-34 – as the title of the example already states. In Example 1.2-34a, this is the DOM-raising transformation implicated in Classical periodic forms, which occurs in such forms because of the way the structure of such forms are conceived in the example – i.e. because of the way a dominant phrase can *move* from its original position in the tree to its final position higher up in the tree. This implies that a similar transformation should exist for the tree in Example 1.2-34b, in Rock music phrases, as the arrow in that example illustrates. That is, there should be a “Subdominant-raising” (or SUBD-raising) movement transformation in this tree involving the IVP, which occupies the same position in this tree as the VP in Example 1.2-34a that raises in the DOM-raising transformation. In this SUBD-raising transformation, the IVP should raise to a position higher up in the tree – i.e. to a “back-relating subdominant” position, as depicted by the arrow.

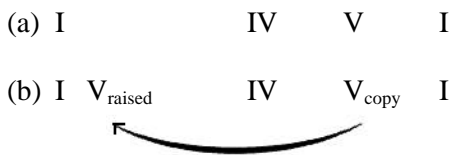
Interestingly, this prediction of a movement transformation in Rock music, based on the existence of a similar transformation in Classical music, *is* borne out by the data. In certain Rock songs, phrases that show typical model phrase harmonizations like I-V-IV-I, often have a IV chord inserted after the initial tonic, whose function is to prolong this tonic in a back-relating manner. This yields the following structure I-IV, V-IV-I – an example of which can be found in the Troggs’ song “Wild Thing”, a song probably better known through Jimi Hendrix’s memorable performance of it at the Monterey Pop Festival in 1967, and in which the I-IV, V-IV-I progression is the basis for every chorus in the song. A possible explanation for how this back-relating subdominant arises in the model phrase progression is that it *raises* from its original position adjacent to the final tonic of the progression, to its final, back-relating position, to generate the S-structure of phrases in songs like “Wild Thing”.

But there seems to be an important difference between the DOM-raising and SUBD-raising movement transformations. The DOM-raising transformation is a necessary one – without it the requirements that a tonal phrase have an authentic cadential (i.e. V – I) ending *and* that the initial tonic receive back-relating dominant support cannot both be simultaneously satisfied. The SUBD-raising transformation does not seem to have such necessity, since Rock phrases do not seem to require an analogous IV – I ending, meaning that the IV phrase involved in the SUBD-raising transformation does not *have* to raise from its ‘cadential’ position – in fact, it does not have to originate in that cadential position to begin with, in the way the V phrase that participates in DOM-raising does.⁷⁰ So, in a sense, the DOM-raising transformation is a more fundamental one compared to the SUBD-raising one.

In this light, one could think of Rock music phrase structure as arising from this, more fundamental, DOM-raising transformation *as well*. To understand this, consider the fact that when a movement transformation takes place, the moved constituent is said to leave something behind in its

⁷⁰ Moreover, the phenomenon that SUBD-raising is supposed to give rise to, i.e. ‘back-relating’ subdominants, might have an entirely different explanation. So, one could think of these harmonies, not as prolonging the I that precedes them, but rather the V that *follows* them, as in: I-(IV-V)-IV-I. So, if there is a SUBD-raising movement, it could be raising the IV phrase to a different position altogether (i.e. to a predominant one) rather than to the position DOM-raising raises the structural V phrase to – or there could be no movement transformation here to begin with, making the notion of a “SUBD-raising” conceptually unnecessary, in a way that DOM-raising is not.

original position. Earlier versions of generative theory took this something to be a “trace”, and later versions (especially Chomsky (1995b)) take this something to be a copy of the moved constituent itself. If we assume this latter “copy theory of movement” position, then a dominant phrase that raises to a back-relating dominant position in a DOM-raising transformation will leave a copy of itself behind in the original position of the dominant phrase. In the standard tonal model phrase progression I-IV-V-I, this gives rise to the following derivation:



Now, you might remember from our earlier discussion on language that the copy of a moved linguistic constituent may or may not be overtly pronounced after S-structure has been generated, and this difference is parameterized across languages. For example the copy of a *wh*-phrase that has been left behind after *wh*-movement is either left unpronounced (as in English, where only the moved phrase is pronounced), or both the moved phrase and its copy are pronounced (as in vernacular German), or only the copy (and not the moved phrase) is pronounced (as happens in “*wh*-in-situ” languages that show *covert wh-movement*, such as Mandarin). This leads to three possibilities for whether the copy of the dominant phrase in (b) above, i.e. V_{copy} , is ‘pronounced’ (i.e. whether it is sounded in a performance of the phrase) or not:

- | | | | | | | |
|------|---|---|----|---|---|---|
| (b1) | I | V_{raised} | IV | V_{copy} | I | (only V_{raised} pronounced, a.k.a. “overt movement”) |
| (b2) | I | V_{raised} | IV | V_{copy} | I | (V_{raised} and V_{copy} pronounced, a.k.a. “copy movement”) |
| (b3) | I | V_{raised} | IV | V_{copy} | I | (only V_{copy} pronounced, a.k.a. “covert movement”) |

We have discussed the second possibility here, i.e. (b2), already – as being the kind of transformation that generates the S-structure of Classical periodic forms (both parallel and contrasting), since this is what allows for the generation of both the back-relating and authentic cadential dominant phrases in them. One might argue that (b3) is what happens in non-periodic Classical phrase structures too – although this requires evidence.

But I am primarily interested here in the first possibility, i.e. (b1), since this – quite strikingly – generates the surface form of the characteristic *Rock* progression I – V – IV – I. This implies that the DOM-raising transformation can generate the periodic, and possibly non-periodic, S-structures of Classical tonal phrases *and* the S-structure of characteristic *Rock* phrases. Specifically, an *overt* DOM-raising movement generates the I-V-IV-I S-structure of *Rock* phrases, whereas *copy* or *covert* DOM-raising movement generates the I-IV-V-I or I-V || IV-V-I S-structures of Classical tonal phrases. But this just means that the DOM-raising movement is parameterized, just as wh-movement is in language – and it is the parameterization of these movement transformations that allows them to generate different phrase structures in different languages or musical idioms.

After DOM-raising generates the I-V-IV-I *Rock* S-structure, SUBD-raising can be applied to it to generate the I-(IV)-V-IV-I S-structure of *Rock* phrases of the kind found in the Troggs’ “Wild Things”. But the important point here is that this happens *after* DOM-raising, the latter being the more fundamental movement transformation. Moreover, DOM-raising, as is the case with all movement transformations, is just an instance of internal Merge, as we have seen before. So, just applying Merge to the lexical items of music, which I have been taking to be chords, seems to be sufficient for generating the different phrase structures of *Rock* and Classical tonality, as long as it is appropriately parameterized. That is, first external Merge merges chords based on their cf values, and then internal Merge transforms the resulting structures into the diverse S-structures we see in idioms like *Rock* and Classical tonal music. This is a thoroughly Minimalist model of musical phrase generation – one that does not assume any *Ursatz* forms, as axiomatic primitives, for either *Rock* or Classical music, and one that is identical to the model that generates the differently-ordered S-structures of English and Irish (using appropriately parameterized V-raising and NP-movement transformations) in human natural language.

In this way, a Minimalist generative model, based on Schenkerian ideas, but one that also invokes transformations, is able to describe the structure of both Classical and *Rock* phrases, without reducing one to the other (as more traditional Schenkerian theorists often do), but also without treating them as totally

separate systems as the anti-generative theorist would. This demonstrates that a hierarchical, generative model *is* better than a finite-state one at describing musical phrase structure – thus falsifying David Temperley’s anti-Schenkerian claims to the contrary in the process.

All in all, this confirms, in my opinion, the validity of a transformational approach to modeling musical grammar. Many models of musical grammar have been proposed over the years, but none have noticed this transformational component within it (apart from Katz and Pesetsky’s model of course, in their discussion on cadences, which we examined in the previous subsection). Moreover, none have examined how a specifically Schenkerian approach to musical grammar reveals not only its transformational nature, but also its Minimalist foundation. This is why I disagree with Martin Rohrmeier’s assessment that “The core rules [of musical grammar, as discussed in his model] are sufficiently simple to be expressed by any of the current main models of linguistic syntax, without requiring the use of the special features of any of these models” (Rohrmeier (2011): 51). I believe a Merge-based, Minimalist approach to musical grammar *does* reveal aspects of musical structure that other formalisms cannot, such as its derivational and transformational nature – hence my choice of this approach in the current dissertation.

iv. Bare Phrase Structure and Schenkerian reduction

The past three subsections have explored certain ways in which the human computational system for music (i.e. C_{HM}) can be understood from a Minimalist perspective, and how this reveals certain points of identity between C_{HM} and the human computational system for language or C_{HL} . These explorations have been quite extensive; so in the interests of space, the remaining subsections, starting with this one, will continue this exploration, but much more briefly. This subsection will explore the Minimalist notion of “Bare Phrase Structure”, and its relation to a generative model of musical grammar.

In the 1990s, Bare Phrase Structure was proposed by the Minimalist Program as an alternative to the old X-bar theoretic description of linguistic phrase structure. X-bar theory had conceived of linguistic phrase structure in terms of the “Specifier/Head/Complement” representational schema, implying that the

labels of “specifier”, “head” etc. had to be assigned to any constituent that was included in an X-bar tree. As we examined in 1.2.3, this can lead to problems regarding what categorical label to assign to a given constituent. This, in turn, can lead to what the linguist Cedric Boeckx calls “spurious” or “vacuous” projections (Boeckx (2006): 174) – since every constituent has to be assigned a categorical label in an X-bar tree, every constituent ends up being forced a label, e.g. that of a “head”, even when there are no accompanying specifier, complement or adjunct constituents in the tree, leading to the tree’s having vacuous specifier, complement, and adjunct branches. (We saw an example of this in Example 1.2-20a, when considering the X-bar structure of determiner/noun phrases.) To deal with this problem, Minimalism proposed a derivational approach to phrase structure instead, in which X-bar theory’s top-down representational schema is not assumed – i.e. linguistic phrase structure is taken to be “bare” – and phrases are generated bottom-up, by merging constituents two at a time, X-bar theory’s categories of head, specifier etc. then emerging as by-products of such derivations. This eliminates spurious categories, since a category will emerge within a bare approach to phrase structure only when an actual constituent merges with another constituent, and thus merits a categorical label. So, “Sean”, in the sentence “Mary kissed Sean”, will emerge as the complement to the head verb “kissed” only because of its merger with “kissed” to form the verb phrase “kissed Sean” – the categorical label *complement* arising only as a result of the bottom-up derivation of the verb phrase “kissed Sean”.

What this means is that the bare approach to phrase structure, in addition to being bottom-up and derivational (as opposed to top-down and representational like X-bar phrase structure), is also *relational*, as opposed to *categorical*. “Sean” is *related* to “kiss” as one of the arguments of this transitive verb – and can therefore merge with it. This is in contrast with Sean’s participating in the phrase structure of “kissed Sean” because it fills some *a priori* categorical label, such as “complement”, as would be the case in an X-bar theoretic description of the phrase’s structure.

Turning to music, we have seen that a bottom-up, derivational, and relational approach to phrase structure – i.e. a bare approach to phrase structure – has advantages even when describing musical phrase structure. As we saw in section 1.2.4.i, this approach helped us deal with some problems in Allan Keiler’s

quasi-X-bar-theoretic, representational view of tonal phrase structure. Moreover, a derivational approach helped us avoid the mess of branches in the representational trees of Example 1.2-34, when we wanted to accomplish the simple merger of IV with V. So, a bare phrase structure approach is advantageous when describing the structure of musical phrases too, even if musical phrases can be considered to have an X-bar theoretic representational structure, which, in section 1.2.4.ii, I suggested might be the case for certain kinds of predominant-dominant relationships in tonal phrases.

What is interesting though, is that the novel bare approach to phrase structure proposed by Minimalism is actually the view of phrase structure inherent in Schenkerian theory – a point that has been noted by some scholars (e.g. Jackendoff (2009): 201, Katz and Pesetsky (2011): 20). Schenkerian theory describes tonal phrases as groups of hierarchically-paired sonorities, using the binary labels familiar to most music theorists – “prolonged” vs. “prolonging”, “structural” vs. “non-structural”, “harmonic vs. non-harmonic”, “consonant vs. dissonant” etc. Normally the sonority labeled with the first of the above pairs of terms is the one that is given greater hierarchical status in the structure of a musical phrase too, meaning that it would be the one to get a stem or a white note head in a Schenkerian analysis of that phrase. So, in a passage where a leading-tone diminished seventh chord is followed by a root position tonic triad, assuming no further voice-leading or harmonic complications, the standard Schenkerian analysis would be to interpret the seventh chord (with its two tritones) as a dissonant, non-structural sonority that prolongs the consonant tonic triad. The issue is trickier when one root position tonic triad follows another in a musical passage, e.g. when a predominant IV triad is followed by a V triad, because both triads would appear to be equally consonant (in acoustic terms) – so, the hierarchical structure of such a passage is more difficult to describe. (Although, this is the situation where my proposal regarding cf values discussed in earlier subsections might be of use.) But in any case, a standard Schenkerian analysis would normally analyze the two triads as forming a hierarchically-ordered pair, with the V as the hierarchically-superior member of the pair.

This means that the Schenkerian approach to tonal phrase structure is one in which sonorities are thought of in *relational* terms, as prolonging or prolonged etc., which allows them to be grouped into

binary pairs on this basis too (often by means of slurs or beams).⁷¹ Applying this procedure to more and more sonorities therefore allows one to generate a larger musical phrase in a bottom-up fashion. The resulting phrase might be described in representational terms (as Allan Keiler does, for example) – but the basic generative procedure here – the basic generative procedure described by Schenkerian theory – is strictly derivational.

So the move from X-bar theory's representational approach to Minimalism's derivational approach in linguistic theory not only allows one to think of both musical and linguistic phrase structure in similar, if not identical, ways – as the previous sections have argued – it also reveals a deeply Schenkerian, music-theoretic, approach to issues of grammar and hierarchy. Bare Phrase Structure allows one to conceive of a verb + noun pair like “kissed Sean” as essentially a prolongation of “kissed” rather than a pair of constituents that can be inserted into a “head + complement” representational schema – which forcefully unifies generative musical and linguistic theory, and acts as evidence for their identity.

The Bare Phrase Structure proposal has not been without its critics though. One of these criticisms has to do with what *projects* in the derivation of a phrase. To understand this, recall that in the top-down X-bar schema, the pair {kissed, Sean} would be understood as a head (i.e. “kissed”) and its complement (i.e. “Sean”), which form the V-bar-level projection “kissed Sean”. But categorical labels like *head* and *complement* do not exist as primitives in Minimalism and the bare approach to phrase structure. So, in this approach, “kissed” and “Sean” are just taken to be two lexical items that are combined, bottom-up, into a two-member set by Merge. Now when two lexical items are merged only the

⁷¹ The Schenkerian emphasis on describing how a line of pitches (which Schenker termed “*Zug*”) arises, through voice leading, in polyphonic passages – which are what the slurs or beams in a Schenker graph usually depict – might obscure the binary nature of the pitch relationships described by the graph. This is because a large number, i.e. not two, pitches can be slurred together in a Schenker graph, to show how that particular voice leads from one pitch to the next, within the polyphony of the passage. But the very fact that a slur or a beam necessarily has *two* end points, an initial one and a final one, suggests that the use of this notation assumes a binary approach to pitch relationships – the first and last pitches being connected by a slur or a beam, and (the chords they belong to) being the grammatical entities whose hierarchical relationship the graph depicts, especially since one of these pitches is normally stemmed or given a hierarchically-superior note head than the other. All other intermediate pitches merely arise through voice leading, which means that their grammatical function in a phrase arises at a more superficial level of structure. One could describe these in binary terms too, but most Schenkerian graphs do not, since this would just obscure the hierarchical relationships that occur at deeper levels of structure, whose depiction is often the main goal of a Schenker graph.

lexical information of one of the items (viz. the hierarchically-superior one) – but all of it – projects up the derivational tree, as the merged pair merges with other constituents to generate a phrase. So, in “kissed Sean”, only the lexical information of “kissed”, which happens to be the X-bar-theoretic head of “kissed Sean”, projects up the derivational tree. Therefore, the Bare Phrase model talks about phrase generation in terms of the bottom-up projections of a lexical item, rather than in terms of top-down categorical labels like “head” – even though it is the head that projects its lexical information even within bare phrase structure. However, Jonah Katz and David Pesetsky say that:

“For language, the Bare Phrase Structure view is probably too strong. Certain properties of the heads of phrases, though relevant to other components of the grammar, do not project to the phrasal level in the syntax. For example, phonologically relevant information such as “ends with a fricative consonant” or “stressed on the penultimate syllable” appears to be irrelevant to syntactic phrases. Syntax is also oblivious to morphological information such as “forms its past tense by adding -t” or “belongs to the third conjugation”. No process of syntactic agreement, for instance, cares about conjugation-class in languages like Spanish that have such phenomena – even though conjugation-class is of great relevance to the phonological and word-formation components of the linguistic grammar.” (Katz and Pesetsky (2011): 20)

So, Katz and Pesetsky think that Bare Phrase Structure is suspect because its stipulation that all of the head’s lexical information projects up a derivational tree does not seem to be borne out by the facts – since certain kinds of lexical information do not seem to project to the phrasal level. What is interesting about this critique though, is *what* kinds of information do not project to the phrasal level. Katz and Pesetsky mention phonological information as an example – but this is precisely the kind of information that is specific to language, and not shared with music. To the extent that Bare Phrase Structure reveals a ‘music-theoretic’ way of thinking about linguistic phrase structure, or to the extent that Bare Phrase Structure reveals a point of identity between C_{HL} and C_{HM} , it seems a *virtue* of the model that the only kinds of information that are implicated in its description of how phrases are generated are ones *not* specific to language. That it might not be able to account for how and why certain kinds of language-specific information (such as phonological information regarding fricatives) enter (or don’t enter) into a phrase derivation, might actually be a good thing about the model.

Moreover, Bare Phrase Structure has also had positive empirical payoff too. As Cedric Boeckx says, X-bar theory used to make a distinction between two kinds of intransitive verb, viz. unaccusative

verbs like “leave” and unergative verbs like “laugh”, which was borne out by its different representational schemata for phrases in which these verbs appear. But Bare Phrase Structure’s rejection of X-bar theory’s schemata meant that this distinction between the two types of verb could not be maintained anymore – which might seem to be a drawback for the theory. However, arguments made by some linguists suggest that the tree structure of unergatives *should* look like those of unaccusatives – meaning that Bare Phrase Structure was right to collapse the distinction made by X-bar theory between these two types of verbs (Boeckx (2006): 175). So, Bare Phrase Structure seems to have some empirical justification. And this justification comes from its ability to collapse artificial distinctions between entities that might actually be the same, or least of the same type, as was the case with the above verbs – meaning that Bare Phrase Structure’s push toward greater generality, i.e. towards a greater streamlining of grammatical explanation, is not only consistent with a Minimalist approach, it is the phrase structural model to prefer if one wants to find generalizations between, or collapse artificial distinctions between, language and music too.

v. Empty categories and implied tones

Another unique feature of generative theory is its discussion of “empty categories” in language. Empty categories are constituents, usually noun phrases, that are said to appear in the S-structure of certain sentences, but which have no phonological content – meaning that they are not included in the pronunciation of the sentence. (In other words, they are “null” constituents.) An example of such an empty category can be seen in situations that appear to violate the Extended Projection Principle, which requires that sentences in all languages have a subject. In some languages, like Italian, sentences often appear not to have subjects, e.g. in the Italian sentence “Parla” (i.e. “(He) speaks”), even though the sentence is grammatical. (This is in contrast to English, where the subject has to be overtly present for a sentence to be grammatical, hence the “he” in the English version of “Parla”.) But the absence of a subject in “Parla” also violates the Extended Projection Principle. So, linguists argue that such sentences do have a subject, but just a null one that is not overtly pronounced, which allows sentences like “Parla” to appear as if they do not have subjects. This null subject is the pronominal noun phrase mentioned in an

earlier footnote called (little) “pro”, and its appearance in sentences is parameterized across languages (according to the “pro-drop” or “null subject” parameter), since it can appear, for example in Italian sentences, but is dropped in favor of overt subjects in English sentences.

Another example of such an empty category can be found in phrasal movement transformations, such as in *wh*-movement. From our discussion of movement transformations in the preceding sections, we know that a moved constituent leaves something behind in its original position, which in earlier versions of generative theory was thought of as a “trace”, and in more recent proposals is said to be just a copy of the moved constituent. This copy is not pronounced in languages that display overt movement, such as English, but is pronounced in languages that display covert movement, such as *wh*-in-situ languages like Mandarin. So, a phenomenon like *wh*-movement leads to an empty category in the original position of the moved *wh*-phrase in a language like English, whereas it leads to an empty category in the moved position of a *wh*-phrase in a language like Mandarin. This means that whether the copy of a moved constituent is pronounced or not is parameterized across languages. So, again we see a parametric aspect to how empty categories appear in sentences across languages.

Now the theory of empty categories depends, for one, on the distinction between ‘deep’ grammar, where both overt and empty constituents participate in phrase generation, and ‘surface’ levels of grammatical representation like Phonetic Form, where only overt constituents have a presence. Moreover, in cases where empty categories arise through movement transformations, a theory of empty categories depends on a transformational approach to grammar. Since notions like Phonetic Form and movement transformations are unique to generative theory, especially in its Minimalist revision, this makes the theory of empty categories a unique feature of generative linguistic theory too.

What is striking though, is that a similar theory of empty categories forms an important part of generative *music* theory, and in many ways is unique to the Schenkerian approach to musical structure. If you look back at Example 1.2-5, which presents an essentially Schenkerian reduction of a passage from Beethoven’s Kreutzer violin sonata, which we discussed at length earlier in this chapter, you will see an example of a musical ‘empty category’ there too. This is the *implied* G5, shown in parentheses in the

soprano in the penultimate measure of the example. This G5 is not heard when this passage is played because it is not present as a note in the score for the performer to play.⁷² In this sense, this G5 has no ‘phonological content’ – it is an abstract tone, placed by the theorist in an abstract structural description of the passage, but which does not map to the sound structure of the passage, the latter being what is represented by the notes in the score that the performer plays.

Just like a linguistic empty category, music theorists place such implied tones in structural descriptions of musical passages for grammatical reasons too. In the above Beethoven passage, the soprano G5 of the penultimate measure provides scale degree 2 of the *Urlinie*, shown by the horizontal beam in the example. In this sense, the G5, supported by V⁷ harmony, can be considered the background sonority from which the S-structure of the penultimate measure is derived. Why this G5 has to appear in the *Urlinie* has to do with Schenkerian thoughts about how the *Urlinie* itself is structured, as a conjunct descending line (represented here in its scale degree 3 – 2 – 1 form) – which betrays, importantly, a rather parsimonious attitude toward voices and voice leading that I will take up more in the next subsection. But the point worth noting here is that the implied nature of scale degree 2 in the penultimate measure of the Kreutzer passage, and the way it seems to be substituted by an E5 in the soprano (which is actually the alto pitch of this measure), is a very common phenomenon in tonal music. (The substitution of a scale degree 2 soprano with an apparent scale degree 7 soprano, i.e. G5 with E5 as happens in the Beethoven passage, is known, unsurprisingly, as “substitution” in the Schenkerian literature, e.g. see Cadwallader and Gagne (1988): 23-24, Forte and Gilbert (1982): 120.)

So, a theory of empty categories seems to exist in Schenkerian descriptions of tonal passages too, where certain tones are placed in these passages for grammatical reasons, but which do not get articulated in the performance of these passages. And just like in the case of generative linguistic theory, such a theory of empty categories is unique to Schenkerian theory, primarily due to its insistence on

⁷² The soprano G5 on the downbeat of the following measure does not count, since that is a dissonant pitch, suspended over the final tonic triad of the measure. Also, there *is* a G in the piano’s right-hand part, a G3, which forms the tenor voice of this passage. This G cannot be considered a displaced soprano pitch either, since it prepares the tenor B-flat seventh of the V⁷ chord of this measure, through a 5-7 upper neighboring motion over the bass.

hierarchically-organized levels of grammatical structure in musical passages – which separates Schenkerian theory from the various surface-oriented, anti-generative perspectives found throughout the discipline of music theory. Specifically, the Schenkerian approach allows one to talk about a tonal passage as being generated from certain deep harmonic-contrapuntal archetypes, most importantly the *Ursatz*, all of whose tones might not be heard in the surface of the passage after the passage is generated, but whose existence is *implied* by the deep harmonic-contrapuntal grammar of the passage. This means that the theory of empty categories in its Schenkerian musical form is really a theory of *implied tones* (which is possibly a better name for the concept, given that empty categories are not really ‘empty’).

In the most thorough description of implied tones in the music-theoretic literature, i.e. Rothstein (1991), the Schenkerian music theorist William Rothstein shows a similar attitude to implied tones as generative linguists show towards empty linguistic categories. For example, he draws the important distinction between “notes”, i.e. entities with specific pitch, duration, and maybe even a specific dynamic or other articulation that are actually performed by a musician, and “tones”, i.e. the abstract representation of such notes in a structural description of a tonal passage. This means that only tones can be implied, since notes are always actual – they are what one sees in actual *notated* scores of musical passages. And just like empty linguistic categories, musical implied tones do not have a specific ‘phonological’ articulation either. They do not have a specific duration (even though they may be represented with one for convenience of explication, as with the *half-note* implied G5 in the above Beethoven example, and as is the case with all Schenker graphs). Also, they normally have a specific pitch class, as evident from the harmony they belong to, but they do not have to have a specific pitch, and therefore register.⁷³

⁷³ Rothstein claims that implied tones often have definite pitch (Rothstein (1991): 294), and this makes sense given that implied tones often belong to specific voices in a tonal passage, with a specific associated register. This is why the implied tone in the Beethoven passage of Example 1.2-5 was specifically a G5, and not a G4 or a G6. But I would argue that this has less to do with the harmonic-contrapuntal grammar of a passage, and more with the general principle that voices have conjunct melodic structure – a principle of musical parsimony that I will discuss more in the next subsection. As far as grammar goes, an implied tone can be in any register. Rothstein might even agree with this, given his thoughts on the relationship between implied tones and imaginary continuos (see below). However, I believe Rothstein is wrong in another claim that he makes, viz. that implied tones sometimes have specific timbres (op. cit.). He cites a passage from an aria in Bach’s *Magnificat* BWV 243 as an example, where implied tones seem to occur at the end of the two flute parts of the piece, and therefore appear to have a specific timbre, viz. that of a flute. I would argue that the issue is more complicated. For example, if one of the flute parts

The latter point is important since it shows how implied tones will normally be chord tones in a chord that occurs within a grammatical harmonic progression. Rothstein makes this point with reference to the notion of the “imaginary continuo” (Rothstein (1991): 295-297). Every tonal piece can be thought of as having such an imaginary continuo, which is an abstract realization of the upper voices of the piece, given its bass line, of the kind a continuo player might improvise. Such a continuo represents, therefore, a middleground level of structure not quite as abstract as the I-V-I *Ursatz* of the piece, but abstract enough to illustrate the essential chord progression underlying its surface. The imaginary continuo illustrates the essential contrapuntal structure of a piece too, in the way its upper voices are realized from the bass line according to principles of good voice leading. This means that the imaginary continuo will contain all the tones, in all registers, *implied* by this harmonic and contrapuntal structure, without which the piece would be ungrammatical.

Several historical theorists, such as Johann Philipp Kirnberger and Carl Czerny, recognized the immense utility of the imaginary continuo, as a way of modeling the basic harmonic-contrapuntal structure of a piece. But it was Schenker who highlighted the role of the imaginary continuo in illustrating the hierarchical, grammatical structure of a piece, given the way the voices of this continuo can be elaborated, through melodic and rhythmic diminution and displacement, to generate its surface. And this is what allows one to place the surface of a piece in its grammatical context too, ascribing implied tones to it when necessitated by tonal grammar. As Rothstein says:

were actually scored for flute and another instrument, say a violin, with the flute and violin alternating the notes of the part in an irregular manner, it would be unclear as to which instrument should play the final implied tone, although the pitch of this implied tone *would* be clear for grammatical and voice-leading reasons. So, the timbre of an implied tone does not arise necessarily, out of, say, grammatical requirements, in the way its pitch does. The larger issue here is that there are various non-grammatical aspects of music that often seem to be implied in a musical passage, and which form the basis for larger theories of musical implication too (such as those of Leonard Meyer, which William Rothstein also discusses in his above essay). But not all these aspects of music are relevant to a *grammatical* study of implied tones. A similar point can be made about language. Linguistic sentences have all kinds of implications, e.g. if someone asks a Frenchman a question, one might expect the Frenchman to answer with a sentence pronounced with a French accent, i.e. a sentence that has the phonological properties of French – just as one expects the implied tone in the above Bach aria to be ‘pronounced’ with a ‘flute accent’. But this is by no means necessarily the case – the Frenchmen might have been brought up in Australia, thus speaking with an Australian accent – meaning that these extra-grammatical implications have little bearing on the grammatical structure of the sentences under consideration. Of course this could have a bearing on the meaning, or some other contextual aspect, of the sentence – but the fact that it has no bearing on grammar is precisely the point.

“What is more significant ... is [Schenker’s] recognition of the principle we have enunciated – namely, the implied presence of all chord tones in all registers. These implied chord tones are efficiently represented by the imaginary continuo ... the implied presence of all chord tones helps the listener to account for [any newly-introduced pitch in a passage, which] must either be an instantiation of some implied tone, or [must come] from – and is syntactically dependent upon – some implied tone.” (Rothstein (1991): 300)

From the above discussion, it might be apparent now that implied tones can be of two types, viz. those that are ascribed to passages for contrapuntal reasons, and those that are ascribed for harmonic reasons. Contrapuntally-implied tones are by far more common, since the other pitches of the harmony to which these tones belong are usually present in the musical surface – which reinforce the presence of an implied tone there. Sometimes even the very pitch *class* of such an implied tone is present in the surface, albeit in a different voice. This was the case in Example 1.2-5, where the pitch class of the implied soprano G5 is already present in the G3 of the tenor there. Probably the most common form of such contrapuntal implication occurs in what William Rothstein calls “contrapuntal elision”. This is the phenomenon where a dissonant tone, such as the seventh of a V⁷ chord, appears directly in a passage, without a preceding or following consonance to prepare or resolve it. An example can be found in the progression ii – V⁷ – I in C-major, voiced bass upwards in the following manner:

C	B	C
<i>D</i>	<i>F</i>	E
F	G	G
D	G	C

The two italicized pitches in the alto voice of the above progression show how the seventh of the V⁷ chord, viz. F, is leapt into from the preceding D, the F thus arising as an incomplete upper neighbor between the D and E of the alto. This means that the seventh of the V⁷ chord is not prepared by step, as should happen according to the rules of strict counterpoint. To account for this infelicity, Schenkerian theorists often ascribe an implied tone to the above progression to prepare the seventh, normally a tone an octave above the bass of the V, and which therefore prepares the seventh via 8-7 motion:

C	B	C
<i>D</i>	<i>G - F</i>	E
F	G	G
D	G	C

But in the actual surface of this progression the seventh is *elided* with the preceding G that prepares it, giving the appearance of the seventh arising directly, by leap, in the progression – which illustrates the role of contrapuntal elision in preparing dissonances through implied tones, without which the progression would violate the rules of counterpoint.

In contrast with such contrapuntally-implied tones, implied harmonies (or *elided* harmonies, given the previous paragraph's discussion) are found only infrequently. This is because in such cases a theorist ascribes an entire harmony to a passage, all of which is left unarticulated in the musical surface – which is a more difficult ascription to make compared to contrapuntal implication, where only a tone in one voice is normally ascribed to a passage, and which is reinforced by the bass line and the other chord tones that are present in the surface. Moreover, implied harmonies do not even exist in the imaginary continuo, since the latter is derived from the *actual* bass line of a piece, and therefore illustrates the actual chord progression heard in the surface of a piece – even though this surface might not articulate certain (contrapuntally-implied) tones in the progression, as we have just seen. This makes it doubly hard to ascribe implied harmonies to tonal passages.

But one could argue that implied harmonies are more important for a theory of empty categories in music than contrapuntally-implied tones because they parallel more closely the entire *words* or *phrases* that are implied in linguistic sentences. In contrast, a contrapuntally-implied tone is just part of a chord, and therefore is more analogous to a *part* of a word that is implied in a sentence – which is not the focus of the generative theory of empty categories. For example, the phrase “do not” is often contracted to “don't” in English, by subtracting the second *o*. So, when someone says, “don't do that”, we could ascribe an implied letter *o* to the first word in the sentence – and one could argue that this is akin to what happens in contrapuntal elision, when one ascribes an implied tone to a melodic line. But the contraction of “do not” to “don't” is precisely *not* the kind of phenomenon that the theory of empty categories in linguistics engages with. So, if chords are like words – which is a belief I have advocated several times in this chapter – then *harmonic* implication in music is more like the linguistic phenomena of wh-traces or pro.

Even if empty categories make sense theoretically, as I have tried to demonstrate in this subsection, justifying them on empirical grounds is important too. However, this has been a challenge for generative linguistic theory – which, in light of the argument just made, makes justifying harmonic implication on empirical grounds equally important, and equally challenging too. (Contrapuntal implication seems less controversial, which might be related to the fact that it is governed by the imaginary continuo – a concept that several (non-Schenkerian) music theorists have subscribed to as well.) But linguists have suggested some evidence for empty categories in language. One piece of evidence comes from the relation between wh-movement and the phenomenon called “wanna-contraction” (Carnie (2002): 288). “Wanna-contraction” simply refers to how “want to” is often contracted to “wanna” in vernacular English, just as “do not” is often contracted to “don’t”. This by itself is uninteresting for a theory of empty categories. But there are restrictions on wanna-contraction that *are* interesting. When wh-movement targets a clause’s subject, wanna-contraction leads to ungrammaticality:

- (20a) You want me to read the book. (prior to wh-movement)
 (20b) Whom do you want to read the book? (after wh-movement)
 (20c) *Whom do you wanna read the book? (after wh-movement and wanna-contraction)

But when wh-movement targets the *object* of this clause, wanna-contraction can also occur without issue:

- (21a) You want to read the book. (prior to wh-movement)
 (21b) What do you want to read? (after wh-movement)
 (21c) What do you wanna read? (after wh-movement and wanna-contraction)

What’s happening here is that the subject “me” in (20a) is *ex hypothesi* leaving a trace or copy of itself behind, in its original position between “want” and “to”, after wh-movement. This prevents “want” and “to” from contracting over it in (20c), which leads to the ungrammaticality of that structure. In contrast,

there is no such subject constituent between “want” and “to” in (21a), since wh-movement is targeting the object here, and so “want” and “to” can contract into “wanna” in (21c) without any problem.⁷⁴

This seems to demonstrate, empirically, the reality of empty categories – in this case wh-traces or copies. In the particular case discussed here in sentences (20a-21c), however, the empirical reality of empty categories is demonstrated in the context of a wh-movement transformation. Since such transformations have not been discussed in the case of music, a corresponding empirical justification for empty categories in music is more challenging, particularly for implied harmonies. However, an empirical argument, albeit an admittedly weak one, might be made for implied harmonies in music, in the light of a striking passage from the first movement of Beethoven’s Op. 59 No. 1 string quartet – i.e. the first of the “Razumovsky” quartets. In mm. 85-86 of this movement, Beethoven writes an A major-minor seventh chord, followed by a G major-minor seventh chord. He repeats this in mm. 87-88 and 89-90, and prolongs the G major-minor seventh harmony for a couple of measures more, before writing a C-major triad on the downbeat of measure 93, which is the tonic triad of this passage. So, the essential harmonic progression of this passage appears to be:

G	F	E
E	D	C
C#	B	G
A	G	C

⁷⁴ The matter is actually more complicated than this. You might have noticed the difference between (20a) and (21a), in which the subject NP “me” in (20a) is left out in (21a). This might seem like a sleight of hand, since it is the absence of “me” in (21a) that might seem to be the basis for the acceptable wanna-contraction in (21c) – not the fact that wh-movement targets the object of (21a) as described above. That is, if (21a) were to have “me” in between “want” and “to”, like (20a), wanna-contraction might seem impossible in it too, *even if* wh-movement is targeting the object, and not the subject, of that sentence. The reality of the matter is that there *is* something between “want” and “to” already in (21a) – an empty category called “PRO” (not to be confused with little “pro” in null subject situations). The structure of (21a) is really “You_i want PRO_i to read the book”. Here PRO refers to the “you” subject of the sentence (hence the co-indexing), and so (21a) might be read as “you want you to read the book”. PRO appears in (21a) to provide the predicate “read” with its *agent* argument, otherwise the sentence would violate the Theta Criterion, and would be ungrammatical. But as an empty category, PRO is not pronounced, which is why the sentence, when articulated, *sounds* like (21a). Now, since PRO already exists between “want” and “to” in (21a), we cannot put another subject NP like “me” there, as happens in (20a) – and this is why “me” has been left out of 21a. What is interesting though, is that despite the presence of PRO in (21a), wanna-contraction is still possible, in a way it was not with “me” in (20a). I will not explore the reason for this here.

There are several things that are wrong with this progression. First, there are parallel fifths between the bass and alto of the first two chords. Secondly, the dissonant seventh member of the first chord (i.e. the soprano G) does not resolve to a consonance, as dictated by the rules of strict counterpoint, because the F that follows it itself forms a dissonant seventh with its bass G pitch. Finally, and related to the previous point, the dissonant F in the second chord is not prepared by a consonance, as the rules of strict counterpoint dictate, since it is preceded by the dissonant seventh G of the first chord. (Note that these are all contrapuntal problems, a point that will become relevant soon.)

These problems suggest that there might be an implied harmony in the deeper structure of the passage, which fixes these problems – meaning that this apparently ungrammatical surface progression is actually fine when one examines how it is generated. Now there are several implied harmonies that might be ascribed to this progression to fix its faults. One could ascribe a G-major triad in between the first and second chords. This would make the first chord a *common-tone dominant seventh* chord (Aldwell and Schachter (2011): 606-607), the dissonant soprano G of the first chord being held over, as the chord’s eponymous common tone, to the consonant soprano G of the next chord. This consonant G would then prepare the dissonant F of the G major-minor seventh chord, through 8-7 motion with the bass, thus fixing that problem with the progression too:

<i>G</i>	<i>G</i>	<i>F</i>	E
E	D	D	C
C#	B	B	G
A	G	G	C

This does not get rid of the parallel fifths between the bass and alto in the first two chords though, so one could ascribe a different implied harmony to the progression that takes care of this problem as well, e.g. an F-major triad instead of the G-major triad:

<i>G</i>	<i>F</i>	<i>F</i>	E
<i>E</i>	<i>A</i>	D	C
C#	C	B	G
A	<i>F</i>	G	C

The point is that even this implied harmony, which fixes all the *contrapuntal* problems in the surface progression, still does not seem right *harmonically*. My conclusion is that this is because an implied harmony already exists in this progression, which William Rothstein (following Schenker’s own description of this passage in his *Kontrapunkt* text) takes to be a D-minor triad:

G	F	F	E
E	D	D	C
C#	A	B	G
A	D	G	C

This D-minor harmony occupies the position between the A major-minor and G major-minor seventh chords, albeit implicitly, thus preventing any other harmony from occupying this position – however suitable that other harmony might be in grammatical terms.

In sum, empty categories seem to exist in both music and linguistic structure. More work needs to be done to justify this empirically – but speaking theoretically, this does serve as a striking point of identity between music and language, and the generative theories of music and language that reveal such empty categories.

vi. Principles of economy and voice-leading parsimony

One of the key features of the Minimalist Program is its emphasis on how general principles of economy help form cognitive systems like language. Noam Chomsky has referred to such principles as “third factor principles”, whose examination takes the study of language beyond a mere search for descriptive and explanatory adequacy (Chomsky (2005): 6-11). We have seen in section 1.2.3 how this concern for economy has led to the streamlining of description and explanation within generative theory, which is how the Minimalist Program gets its very *minimalism*. For example, we have seen how a concern for *representational economy* led to the development of the Minimalist idea that linguistic computation should be the province of a single operation known as Merge, constrained solely by Full Interpretation, and which gives rise to only the levels of structural representation (i.e. PF and LF) conceptually necessary for C_{HL} to meet the external conditions imposed on it.

Similarly, we have seen how its efficient workings have led linguists to explore the *computational economy* with which C_{HL} generates sentences. For example, in the context of wh-movement, we have seen that wh-phrases move to the closest landing site available, without which a Bounding violation will occur. (This is why wh-movement must occur cyclically, i.e. a wh-movement that requires a wh-phrase to move out of a subordinate clause to the specifier position of the main clause, must first move to the specifier position of the subordinate clause itself, so that each clause is derived in successive cycles.) This computational economy is something that generative linguists have observed in a variety of other movement transformations, which prompted the Minimalist Program to formulate it as an economy condition on movements, which has come to be known as the “Minimal Link Condition”.⁷⁵ The condition states that x can move to y if and only if y c-commands x , and there is no z such that y c-commands z and z c-commands x , z being of the same type as y – i.e. they should both be specifiers, heads and so on. This means essentially that x cannot move to y if there is an intermediate landing site z of the same type as y for it to move to, since a movement to y in this situation would not be computationally economical. (The c-command aspect of the condition just ensures that movement happens to a position, either z or y (if z does not exist), that is a branching sister to the branch that contains x lower down the tree. This seems to be a constraint on movement transformations – if you remember the discussion in section 1.2.4.ii, this is precisely the reason for why my proposed DOM-raising movement moves a dominant phrase to a position higher up in the tree that is a branching sister to the final tonic triad’s bar-level projection, i.e. the branch that contains the dominant phrase itself, lower down in the tree.)

So, the question now is whether there is something akin to the Minimal Link Condition for music – i.e. is music economically structured like language, both representationally and computationally, as Minimalist concerns dictate?

⁷⁵ Noam Chomsky first described this condition in an earlier proposal called “shortest move” (Chomsky (1995b): 181).

Despite its deep antecedents in Schenkerian theory, of the kind I have tried to illustrate throughout this dissertation, the generative approach to music is still a relatively novel research program within music theory. Therefore, at this point, even descriptive adequacy, and especially explanatory adequacy, remain elusive goals for this approach. In this light, an excessive concern within generative music theory for third factor principles, such as those dealing with representational or computational economy, would be premature. However, if generative music theory wants to learn from the history of generative linguistics, and if it does not want to repeat its mistakes, a general concern for issues of economy within musical description and explanation should not be beyond the pale either.

For this reason, I believe that a concern for representing levels of structure in music in the most economical way possible is a relevant task for contemporary generative music theory, and in this light I have already suggested that music can be thought of as having the two minimal LF and PF levels of representation like language too – which is something I will look at in more detail in Part II of the dissertation. (As before, my argument there will examine the idea that such levels of representation are implicit in Schenkerian descriptions of phenomena like musical meaning and rhythm.) On the other hand, the world of music theory has already been discussing certain economies of the computational sort in how musical phrases are generated, so this is worth examining in greater detail now.

The observation that certain music theorists have made about economical computations in music does not arise, of course, from an explicit Minimalist concern for computational economy in music theory, given the novelty of this research program – so no music theorist has tried, for example, to formulate something like a Minimal Link Condition for music. In fact many of the statements regarding economy in musical structure have come from theorists working within anti-generative frameworks. A good example of this is the neo-Riemannian framework in music theory, which I discussed in the last chapter and in section 1.2.2 of this chapter. You might remember from those discussions that neo-Riemannian theory has been particularly interested in understanding certain chord progressions in terms of the semitonal connections between their constituent chords. That is, one can construct certain chord progressions by changing *just*

one pitch of a chord, by a semitone, to get another chord, which you change likewise by just one pitch – until you get a complete chord progression that cycles around back to the first chord. Such cyclical chord progressions (an instance of which we saw in the maximally-smooth hexatonic cycle in Example 1.2-10), therefore display what neo-Riemannian theorists call *voice-leading parsimony*, since they involve chord-to-chord voice-leading motions that move by just one note at a time. So, a large part of neo-Riemannian theory has been dedicated to describing how certain tonal passages can be generated by combining chords through parsimonious voice leading, rather than through certain prolongational procedures (as a Schenkerian theorist might). (You might recall that this has also led neo-Riemannian theorists to explore certain parsimonious voice-leading *operations*, which can be used to combine chords into neo-Riemannian chord cycles – such as the L, P, and R voice-leading operations.)

But neo-Riemannian theorists have not been the only ones to notice parsimonious voice leading as a governing factor in how tonal structures are generated. I discussed in the last chapter how neo-Riemannian theory takes a physicalist approach to understanding music, but there are more psychologically-oriented approaches that have also focused on voice-leading parsimony. Most of these approaches tend to see voice leading as a *perceptual* phenomenon, which guides how we hear musical structures – and therefore they understand voice-leading parsimony as something that is governed by perceptual laws. For example, David Huron has discussed how not just voice-leading parsimony, but voice-leading phenomena in general, can be understood in terms of principles of perception (Huron (2001, 2006)). Others, like Eugene Narmour, have explored voice-leading parsimony in the context of individual melodies – i.e. through the phenomenon of conjunct (i.e. stepwise) melodic motion – which Narmour describes as his principle of melodic “Proximity”, and which is related to more general Gestalt-based principles of perception (Narmour (1990)). Finally, Richard Parncutt has attempted to describe harmony and voice leading in general in terms of acoustic principles (Parncutt (1989)).

Now, none of the above approaches are generative ones – in fact one, viz. that of Narmour, is explicitly anti-Schenkerian. However, what is interesting is that Schenker himself believed that parsimonious voice leading, and conjunct melodic motion more broadly, is an important factor in how

tonal phrases are generated, which is something he discussed under his concept of “melodic fluency” (or “*fliessender Gesang*”, Schenker (1987) Book 1: 94-100). The Schenkerian *Ursatz* is itself a contrapuntally parsimonious structure – its constituent *Urlinie* is a conjunct descending melodic line. Also, Schenker was more generally interested in describing polyphonic tonal passages as complexes of individual conjunct lines (or *Züge*, singular *Zug*), which is an interest that comes from Schenker’s emphasis on the principles of species counterpoint in determining tonal structure.

So the broader focus on voice-leading parsimony in the music theory literature shows us that music theorists generally understand tonal phrase structure as something that arises from combining pitches in an economical way – irrespective of whether one explains such combinations from a generative perspective or not. Which means that music theorists value economy in descriptions of tonal structure.⁷⁶

Now, if voice-leading parsimony is understood as a perceptual phenomenon (which several theorists do, as we just saw), then this might be understood as a ‘phonological’ matter within music, since it has to do with how music sounds, and is heard. In this sense, voice-leading parsimony, or any other voice-leading phenomenon in music, might be considered an aspect of musical phonology, independent from musical grammar or C_{HM} – which is a conclusion that accords with the sentiments of those theorists who prefer a more perceptual approach to music, or who reject the grammatical focus of generative theory.

⁷⁶ The musicologist Janet Levy (in Levy (1987)) has criticized this emphasis on economy as a covert value that influences music-theoretic description more generally, without its being acknowledged as such – which she sees as a negative trait in music analysis. There is no harm in desiring that such implicit influences be made explicit, but Levy also seems to misunderstand the role of economy in music-theoretic or analytical descriptions, as is evident from statements such as, “the provenance of economy qua artistic value seems both varied and suggestive. A few likely sources may be ... the ingrained, though perhaps mistaken, belief that nature itself is economical and that to mirror nature is a priori “good”” (Levy (1987): 10). Levy does not provide any evidence for why the belief that nature is economical is mistaken (especially in the light of all the evidence discussed in the Minimalist literature to the contrary). Which implies that a naturalistic (i.e. scientific) description of music *should* attempt to understand music in terms of natural properties, principles, and phenomena – not as a “value” of some sort, covert or overt, but as something that conforms to proper scientific method. Of course, this also means that music theorists and analysts should be careful to distinguish between those projects and statements that are naturalistic from those that are ‘aesthetic’ or ‘poetic’. But this is not always easy to do, as we saw in section 1.2.2, in the matter of defining what a musical ‘sentence’ is. So a music-theoretic or analytic description that appears to be an aesthetic one might actually be a naturalistic one, and vice versa – one should not assume that a musical statement is necessarily aesthetic or that it possesses certain covert aesthetic values, which is unfortunately something many musicologists do quite frequently, and which Levy seems to do here as well.

But there are important connections between phonology and grammar. We have seen this especially in the Minimalist notion of PF, which is a level of grammatical structure that is meant to satisfy the external conditions imposed on grammar by phonology. And in light of Schenker's own observation about the importance of a 'phonological' phenomenon like voice leading within musical generative grammar, we could conceive of voice-leading parsimony as a phonological economy condition on the generation of musical PF structures by Merge, so that musical surfaces generated by Merge must have conjunct voice-leading structure – or else they will not map to the sensorimotor systems, and the computation will crash. This would not only reveal an aspect of *computational economy* in music, it also allows us to think about voice-leading parsimony in generative grammatical terms (and not just in perceptual terms). And this makes sense practically too, since there is a commonly-held belief that musical passages are easier to sing if they have conjunct melodic structure – which means that when musical surfaces are generated in contrapuntally parsimonious ways they correctly map to the sensorimotor system too, which then allows them to be *interpreted* (i.e. performed).

Thinking about the issue slightly differently, postulating a PF level of representation for musical grammar, which accounts for voice-leading phenomena, not only shows both representational and computational economy in musical structure, it also clarifies how facts about perception are related to, but are not *intrinsic* to musical structure. This can give us a concrete way to talk about how musical grammar is connected to the way the ear parses sound, or how it is connected to performance (or certain performance-practice traditions), without reducing one to the other. This reveals the greater descriptive and explanatory power that a Minimalist approach (and its associated notion of PF) can bring to the study of musical structure, relative to one that looks at music in purely perceptual, or even physical, terms.

Clearly more work needs to be done in order to make a case for economy principles in music. But the above discussion hopefully illustrates the relevance of this work to a study of music, and how some aspects of music already reveal economies, of the kind implicated in descriptions of third factor principles in generative grammatical theory.

vii. Musical grammar and ambiguity

After the considerably technical discussions in the past subsections about the economical, transformational, nature of Minimalist musical grammar, I would like to end this section by moving to a rather different kind of connection between C_{HM} and C_{HL} , and of music to language more generally – one that will involve some fairly old ideas in generative theory, rather than the novel ones proposed by Minimalism. This connection has to do with the issue of *ambiguity* in musical and linguistic structure.

Ambiguity is a major issue in the theoretical discourse on musical structure. It is well known that different analyses of a musical passage by different scholars often reveal conflicting interpretations of that same passage. In fact, many music theorists, including those of a Schenkerian persuasion, see this as a virtue of the analytical enterprise, since they see such conflicts as encouraging imaginative, new ways of hearing pieces of music. Marianne Kielian-Gilbert reflects this position when she censures the Schenkerians Allen Forte and Steven Gilbert for insisting on a specific analysis of a Mozart passage:

“Not only are multiple readings sometimes – often – possible, they may also be a significant way to render the specificity of a particular reading or the dynamic of a progression over time. Might the sense of an ‘oscillation’, a back-and-forth of different hearings, characterize the relationships of such conflicting and/or multiple harmonic readings over time? Should we be wary of the fact that our theoretical tools often compel us to make ‘impossible’ unitary decisions, or should we welcome the fact that they force them, impossible as they are?” (Kielian-Gilbert (2003): 55)

Kielian-Gilbert does not seem to be suggesting that unambiguous readings of tonal structure are impossible or undesirable in general, since she recognizes that such unambiguous readings:

“...can also function to encourage an attention to specifics, so as to highlight the tension between normative and compositional presentation. In this sense, presentational ambiguities of various sorts are often carefully handled within the confines of a particular theory, such that the musical context can ‘clarify’ which of a range of situations a composer might intend.” (Kielian-Gilbert (2003): 65)

So, Kielian-Gilbert just seems to be expressing the desire that a theoretically-driven reading of a tonal passage (especially a Schenkerian one) be open to the possibility of the passage having multiple analyses, in light of the interpretive advantages this brings to the table – even if the theoretical system used by the analyst is capable of producing an unambiguous reading of the passage. Note that the point here is not one

that merely values multiple readings of a passage or ambiguity in musical structure – few would deny such an obvious position. Rather, the point is that ambiguous readings should be entertained even, and perhaps especially, within those analytical systems that are often used to defend specific readings of a passage, as Allen Forte and Steven Gilbert do in their Schenkerian reading of the aforementioned Mozart piece. In other words, theories that are capable of disambiguation should also be able to discuss ambiguity in a meaningful way – a *theory* of ambiguity should be inherent within something like Schenkerian theory, which is often used with a disambiguating purpose.

It is worth noting that Kielian-Gilbert is a music theorist who has often explored issues of tonal structure from within a Schenkerian perspective. So, rather than being anti-Schenkerian, Kielian-Gilbert's argument really has to do with how one countenances ambiguity *within* Schenkerian theory, with Kielian-Gilbert representing, to put it crudely, the 'pro ambiguity' position, and Forte and Gilbert the 'anti ambiguity' one. Such an opposition of positions can also be seen in the way the music theorist Kofi Agawu, much of whose work also falls within the Schenkerian tradition, responds to the noted Schenkerian theorist Carl Schachter's well-known essay on musical ambiguity "Either/Or" (Schachter (1990)). In his response, Agawu develops the premise that "the concept of ambiguity is meaningless within the confines of an *explicit* music theory" (Agawu (1994): 88). By this Agawu means that a theoretically-driven approach to analysis, such as a Schenkerian one, is designed to eliminate all but the one analysis deemed to be 'correct' or 'true'. This is especially relevant in actual musical situations, where a context for eliminating undesirable readings is always available. As an example, Agawu cites Schachter's analysis of a Chopin mazurka, where the latter eliminates all but one of three probable analyses of the piece in light of what the prevalent context – in this case a motivic one – seems to dictate. To this extent, Schachter seems to be taking the 'anti-ambiguity' view that Forte and Gilbert seem to endorse as well. And in response to this, Agawu seems to take the 'pro-ambiguity' view that Kielian-Gilbert takes, when he criticises Schachter's position:

"There is something programmatic about Schachter's "Either/Or" title, for rather than pursue genuine alternatives, he provides, I fear, mostly weak alternatives that are promptly discarded. Has Schachter set up straw alternatives? ... It seems that, although his rhetoric includes references to 'a true double

meaning' and 'genuine ambiguity', none of Schachter's analyses demonstrates a final-state awareness of ambiguity." (Agawu (1994): 104)

Again, Agawu's argument here, contra Schachter, is not one that rejects the Schenkerian impulse towards disambiguation, especially given Agawu's assertion that this is necessary in any theory-driven analytical enterprise. Agawu just seems to be expressing the desire, akin to Kielian-Gilbert, that a greater accommodation be made for genuinely ambiguous readings within a Schenkerian approach to analysis:

"It is entirely possible that, abstractly-composed examples aside, no musical situation in a tonal work is unequivocally clear. It is in the nature of interdimensional interaction that while some dimensions close, others remain open. Never do all dimensions proceed in parallel fashion. If by 'clarity' we mean a particular state of interdimensional tension, then we must be ready to specify the equations that distinguish these relatively clear structures from relatively unclear ones. The clarity/ambiguity dialectic therefore urgently demands explication in order to counter the true but trivial proposition that either all tonal music is clear or all tonal music is ambiguous." (Agawu (1994): 106)

So the issue of ambiguity has an important place in generative, and especially Schenkerian discourses on musical structure, which is one of the reasons I thought it was worth spending some time on this issue in this section. However, my main reason for bringing this issue up is because I actually think the whole discussion of ambiguity within Schenkerian discourse, as represented by the pro- and anti-ambiguity positions discussed above, is rather misguided. To put it another way, I do not believe that a "clarity/ambiguity" dialectic, to use Agawu's words from the above citation, exists when it comes to analyzing complicated musical passages. Part of the problem here has to do with a misunderstanding over the connection between music and *language* that music theorists often reveal when discussing ambiguity – which is why this issue is particularly relevant to our present concerns in this dissertation. A problematic statement that Carl Schachter makes, in his aforementioned essay discussed by Kofi Agawu, exemplifies this misunderstanding:

"[Ambiguity and multiple meanings] certainly do exist. But their function, in my opinion, is more narrowly circumscribed than some analysts, perhaps misled by *false analogies to language* [my emphasis], seem to believe. It is just as much a part of the composer's art as it is of the sculptor's or painter's to be able to create clear and distinct shapes; the more clearly and vividly the listener perceives these shapes, the more fully and deeply will he live the life of the composition as he hears it." (Schachter (1990): 169)

So, the idea here seems to be that if a clarity/ambiguity dialectic exists in music, it might arise out of linguistic notions of ambiguity, which influence how music theorists sometimes think about this issue. But the question is whether ambiguity and multiple meanings are *really* less circumscribed in language, as Schachter seems to believe – because without this, part of the reason for entertaining the clarity/ambiguity dialectic in music will disappear. More specifically, how does ambiguity arise in language to begin with, so that we can even entertain the idea that ambiguity is less circumscribed in language than in music?

To find an answer to this, consider the two following sentences (which are based on a joke in Douglas Hofstadter’s *Gödel, Escher, Bach* (Hofstadter (1979)):

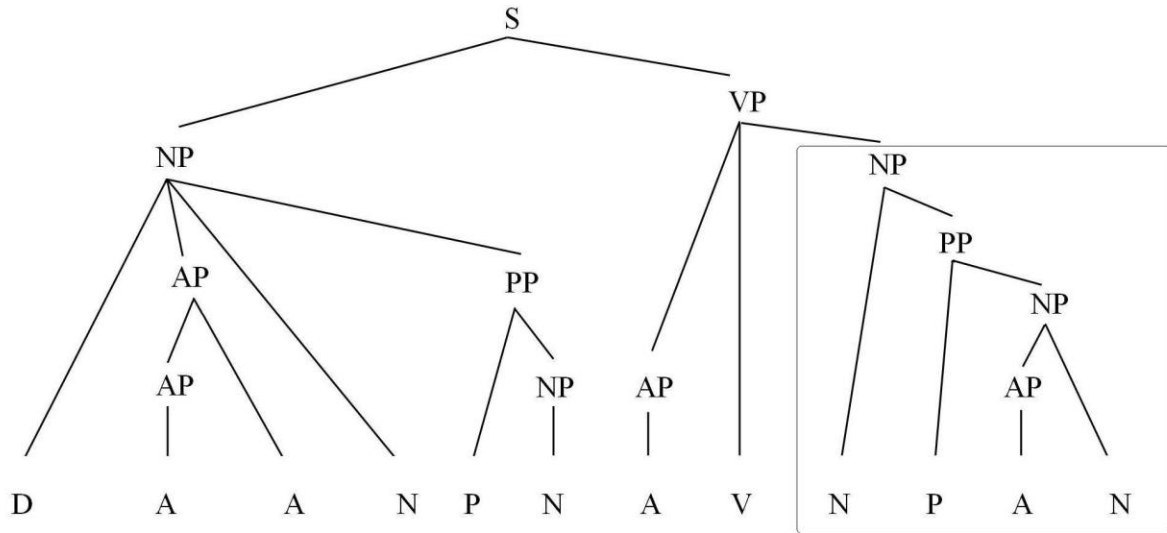
(22a) The Schoenberg tonic factory in Vienna later made cereal with great flavors.

(22b) The Schoenberg tonic factory in Vienna later made cereal with great panache.

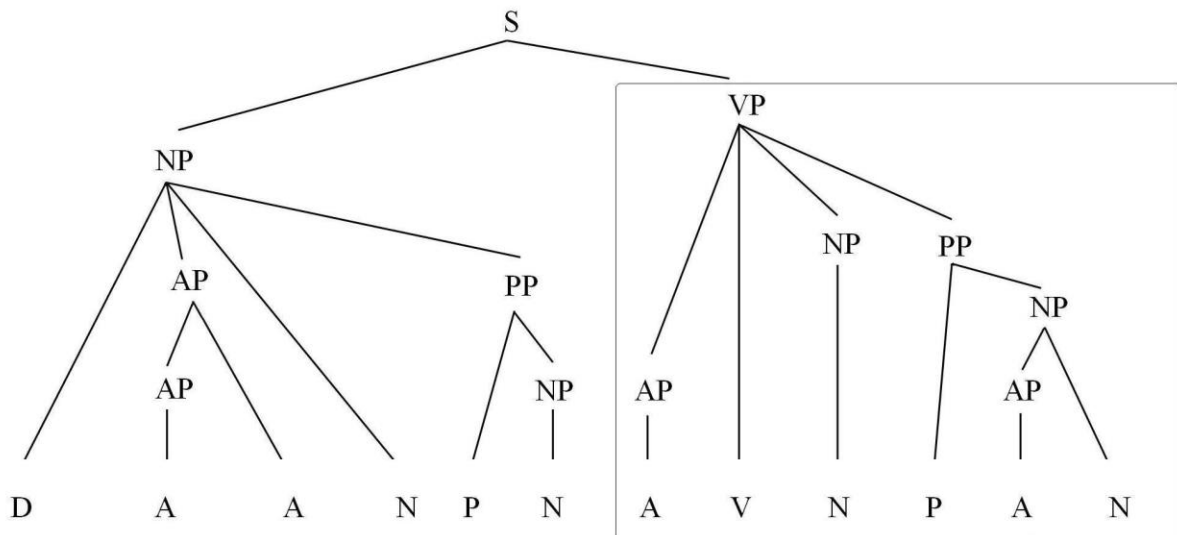
The two sentences are nearly identical, i.e. in their S-structures, differing only in their last words. How does this affect how we interpret them? As the simple Standard Theory phrase structure tree in Example 1.2-38a shows, the likely analysis of the first sentence takes the preposition phrase “with great flavors” to modify the noun “cereal”, to form the noun phrase “cereal with great flavors” (shown in the box). But Example 1.2-38b suggests that in the second sentence the preposition phrase is more likely modifying the verb “made” in the generation of the *verb* phrase “later made cereal with great panache”. So, despite the near identity of their S-structures, we see that the two sentences have rather different tree structures. And this is not all. All that Examples 1.2-38a-b suggest are the *likely* analyses of these two sentences. But what is this likelihood based on? Nothing grammatical surely, since the two sentences are formed from identical strings of parts of speech (the last word, though different, is still a noun in both sentences).

In fact, in light of how closely related the two sentences are, we could analyze them in identical ways too. So, we could analyze the preposition phrase “with great flavors” as modifying “made” rather than “cereal”, meaning that the cereal was perhaps made with flavorful ingredients, rather than it being flavorful itself (after all, dishes can lose flavor through bad cooking, even if they are made with

Example 1.2-38. Ambiguity in language seen through tree diagrams



The Schoenberg tonic factory in Vienna later made cereal with great flavors



The Schoenberg tonic factory in Vienna later made cereal with great panache

flavorful ingredients). We could analyze 1.2-38b in a similar fashion, so that “with great panache” is seen as modifying the noun “cereal” rather than the verb “made”, meaning that the cereal itself, rather than the way it was made, had great panache (maybe it was ‘full of character’ as people often say of fine wines).

So what we have here is a case of ambiguity in language. That is, we have two sentences that are very similar in their appearance but both of which have at least two analyses. For both sentences, one analysis seems more likely than the other – and this ultimately leads to both sentences being assigned a specific tree analysis despite their similar appearance. But the likelihood of one analysis over the other in the case of both sentences is *not based on anything within the grammar of the sentences*, especially since they are made up of identical parts-of-speech strings. All that grammar does is to show us that there is more than one possible analysis of the two sentences. So, which analysis we ultimately accept does not have much to do with grammar, but rather with extra-grammatical factors such as the pragmatics of discourse – ‘real world’ circumstances that determine, for example, the greater likelihood of an edible substance being described as flavorful rather than in terms of the process through which it was made.

In other words, the task of grammar, i.e. C_{HL} , is to generate S-structures, in a way that allows them to map to the external, conceptual-intentional (i.e. ‘semantic’) and sensorimotor (i.e. ‘phonological’), systems via the LF and PF levels of structural representation, and it is the task of grammatical theory to describe and explain this generative procedure. We have discussed before how this amounts to a description and explanation of what can be called linguistic (or musical) “competence”. Which means that it is *not* the job of a grammatical theory to describe and explain “performance” phenomena, such as how generated S-structures get interpreted, after C_{HL} has generated them. That is up to the external systems, and is governed by extra-grammatical, often pragmatic, factors. C_{HL} is concerned solely with the generation of structures that *can* be interpreted, irrespective of whether they *are* interpreted. Similarly, a theory of C_{HL} , i.e. a grammatical theory, is concerned with the conditions that govern the generation of interpretable structures, again regardless of whether these structures *are* interpreted or not, by the external systems of semantics and phonology.

And that's all. Which means that whether a generated structure is deemed ambiguous or not is something that happens beyond grammar, at the level of interpretation, i.e. after grammar has done its job. This is why a structure that is deemed ambiguous is *disambiguated* by extra-grammatical factors too, i.e. the pragmatic factors that help us choose one analysis of the structure over others. So, ambiguity does not even factor into a purely grammatical description of structure – the consideration of ambiguity in language falls beyond the pale of grammatical theory, and in the world of semantics and pragmatics. This is why a clarity/ambiguity dialectic really does not have much of a place within grammatical theory.

What this implies for music is that it would be equally misguided to ask a *grammatical* theory of music to include an account of musical ambiguity, as Kielian-Gilbert and Agawu desire. All such a grammatical theory can do is describe and explain how musical S-structures arise. If these structures are ambiguous, an *extra-grammatical* music theory needs to be invoked to describe and explain how and why these structures are ambiguous, and how they might be disambiguated. So, to the extent that Schenkerian theory is a theory of musical grammar, it is misguided to situate a clarity/ambiguity dialectic within it, or to expect it to include a theory of musical ambiguity too.

But it is equally misguided to think that when we do not see ambiguity, this is because of grammatical factors, because even the resolution of ambiguity is something that lies beyond the limits of grammatical theory. So, when Carl Schachter illustrates how the structure of the above Chopin mazurka can be disambiguated by rejecting all but one analysis of the piece, he is actually engaging in an act of *interpretation*, which really lies beyond the pale of a grammatical description of the mazurka's structure. This is why the factors invoked by theorists to disambiguate structures deemed ambiguous should really be considered extra-grammatical factors too. Schachter's invocation of motivic factors to disambiguate the Chopin mazurka is a case in point here, since this takes his discussion of the mazurka from the realm of grammatical description to that of (semantic) interpretation, even if he does not explicitly recognize this. (Related to this, it has been noted how the (inappropriate) inclusion of motivic factors by Schenkerian theorists, in purportedly grammatical descriptions of musical passages, can lead to

significant problems, e.g. see Cohn (1992b).) So, Schachter's statement that ambiguity is more narrowly circumscribed in music than language is problematic, because it involves a false comparison between *grammatical* description in music (as seen through a Schenkerian lens), and *extra-grammatical* interpretation in language. Ambiguity has no connection to grammatical description in either music or language, as a result of which it is no more circumscribed in music than it is in language – unless it is for extra-grammatical reasons, which lie beyond Chomskyan descriptions of language, and which should lie beyond Schenkerian descriptions of music too.

I think that part of the reason for why ambiguity has (inappropriately) played such a large role in discussions of musical structure – including Schenkerian ones – has to do with the emphasis on *analysis* in these discussions. The conflict between the above statements by Kielian-Gilbert, Agawu, Schachter, and Forte and Gilbert all arise in debates over how to analyse certain musical pieces or passages therein, such as the Mozart and Chopin pieces these authors discuss. But analysis is something that happens after the fact – after the fact of *generation* that is. One can only analyse a musical passage, and subsequently deem it ambiguous, when it has been generated to begin with – which is why ambiguity arises only *after* grammatical generation has played its part. In this sense, analysis is equivalent to an act of interpretation – something that takes us from a discussion of musical competence to that of musical performance. So, by focusing on analysis over generation, one inadvertently creates an intellectual space for ambiguity to flourish in – but only because we have privileged something that should really be the focus of extra-grammatical interpretation after the fact of grammatical generation has been accounted for. If our initial focus is, instead, on describing the phenomenon of generation, the fact that some generated structures are ambiguous will, in contrast, not emerge as a central concern in the way an analytical approach makes it out to be.

Moreover, we saw in section 1.2.4.i how Schenker himself took the study of tonal structure to be one that privileges generation over reduction – i.e. one that focuses on how backgrounds give rise to surfaces rather than how surfaces can be reduced to backgrounds. Therefore, to focus on analysis (which is equivalent to reduction in a Schenkerian framework) over generation not only introduces ambiguity,

incorrectly, into grammatical descriptions of musical structure, it also misrepresents the Schenkerian way of describing tonal structure. In contrast, a Schenkerian study could first think of the different S-structures that can be generated from a certain middleground structure (an example of the latter being something like an imaginary continuo, of the kind discussed in 1.2.4.v), and only then proceed to a specific analysis (or reduction or interpretation) of a musical surface, i.e. by relating it to one of these generated S-structures. This would be a properly grammatical approach to the study of that surface – after which one can engage with the extra-grammatical issue of why a specific analysis/reduction/interpretation is preferable to others. In a sense this is what Carl Schachter does when he discusses the three possible S-structures that a piece might have, in his aforementioned discussion of the Chopin mazurka. Schachter then proceeds to eliminate all but one S-structure as his interpretation of the structure of the mazurka, on motivic grounds (which I take to be an extra-grammatical reason), thus displaying a truly grammatical orientation in his study – and one that is consonant with a genuinely Schenkerian background-to-foreground approach too. This is why Schachter’s ‘anti-ambiguity’ approach gives us, in my opinion, the correct context for ambiguity within a grammatical, and especially Schenkerian, approach to the study of musical structure – so long as one disregards the false statements he makes about *linguistic* ambiguity in this regard.

It is interesting to note that many proposals in the music-theoretic literature that make ambiguity out to be a bigger issue than it really is – at least for truly grammatical approaches to this issue – are often ones that privilege analysis over generation, or which take a surface-, rather than background-, oriented approach to the study of musical structure. A case in point is the music semiologist David Lidov’s critique of generative models of musical structure (in Lidov (1997)), where he argues that the generative theorist’s impulse to understand musical surfaces in terms of discrete tree structures is misguided, since surfaces are often continuous, and cannot therefore be parsed into specific unambiguous tree structures. I think this evaluation of generative music theory is unwarranted because parsing a musical surface is not the focus of such theories to begin with. Instead, the focus of such theories is on generating S-structures, and a

generated S-structure can then be interpreted as ambiguous for a variety of extra-grammatical reasons, including, say, acoustical ones, which can make a surface seem continuous and hard to parse. But that such ambiguity occurs has nothing to do with the grammatical structure of a surface in itself, which Lidov's 'surface-first' approach fails to notice.

Just as a surface-oriented approach can privilege ambiguity in problematic ways, such an approach can also disambiguate ambiguous surfaces through appeal to extra-grammatical factors too, even when that is not the goal of the relevant theory. An example of this can be found in Lerdahl and Jackendoff's analysis of the initial measures of Mozart's 40th symphony, K. 550 (Lerdahl and Jackendoff (1983): 22-25). They propose two S-structures for the passage, and say that they "will refrain from choosing between these competing alternatives", which is the appropriate stance to take from within a generative perspective. But Lerdahl and Jackendoff's approach really amounts to a surface-oriented, parsing theory of musical structure, similar in some respects to semiological theories of music as discussed in the last chapter, and of the kind subscribed to by theorists like David Lidov – and this ultimately makes it impossible to refrain from choosing one of their S-structures over the other. Some combination of preference rules will always privilege a specific reading of the passage, as did Carl Schachter's invocation of extra-grammatical, motivic factors in his analysis of the Chopin mazurka above. This is something Kofi Agawu has pointed out too, especially given how Lerdahl and Jackendoff cite examples of performances of the Mozart symphony that choose one reading over another – but do not cite a performance that refrains from making such a choice (Agawu (1994): 98-99). The point being that a surface-oriented approach to musical structure will always create the space for ambiguity to arise, and also be disambiguated, in a way that does not happen in a generative approach to musical structure.

Further evidence for how ambiguity arises as a significant feature only in surface-oriented approaches to musical structure can be found in another, more recent, proposal by Agawu, viz. in his discussion of "retroactive" versus "prospective" prolongation (Agawu (2008b): 122-124). These terms are defined in the context of the progression X-Y-Z, where Y is a hierarchically-inferior sonority, and can be understood either as prolonging the earlier X sonority "prospectively", or the following Z sonority

“retroactively”. Although Agawu situates the discussion of the two types of prolongation within a larger discussion of how surfaces might be generated, he goes on to say that the two possible descriptions of how Y might be functioning in this progression:

“...begin to seem confused, but that is precisely the point. That is, the domains of both a prospective prolongation of [sonority X] (by means of a suffix) and a retroactive prolongation of [sonority Z] (by means of a prefix) overlap. Such is the nature of the continuity and internal dependence of the elements of tonal-functional material.” (Agawu (2008b): 123)

I read Agawu’s depiction of “confusion” in the above quote as referring to the progression X-Y-Z being inherently ambiguous, in the way its two “domains of prolongation” overlap. But, importantly, there is no overlap between them in the context of structure generation in itself, since in generating a surface Y will be taken *either* to prolong X *or* Z – there can be no ambiguity, no overlapping domains of prolongation, here, or the surface will never be generated. It is only when *analyzing* such a surface that any such ambiguity in the function of Y will become apparent. Therefore, the fact that Agawu makes this statement within the context of a discussion of how surfaces might be generated, suggests that he is conflating generation (from the background) with analysis/interpretation (from the surface). In other words, the distinction between prospective and retroactive prolongation is really a distinction that lies in the realm of interpretation, governed by the pragmatics of musical discourse.

I would like to discuss one final example of a surface-oriented, analytical (as opposed to generative) approach to musical structure, in order to show not only the problematic privileging of ambiguity this leads to within discussions of musical structure, but also how this leads to problematic evaluations of the very project that is Schenkerian music theory. This can be found in a recent essay by the music theorist Poundie Burstein, on the connections between Schenker’s ideas and the scientific method. Burstein begins his essay by making a striking claim:

“It has been suggested that Schenkerian analysis has – or should have – a scientific or logical basis. For instance, it often is claimed, either explicitly or implicitly, that Schenkerian analysis can demonstrate whether a composition is tonally unified, or that Schenkerian analysis can show how a work is understood by skilled listeners. But if this were the case, then the method developed by Heinrich Schenker would be either ineffective or – at best – inefficient.” (Burstein (2010): 1)

He goes on to infer from this that:

“The notated compositions examined by the typical Schenkerian analysis, however, admit a variety of possible valid realizations in performances that are intended for a variety of types of audiences. Consequently, the best a Schenkerian analysis can do is to propose a persuasive and effective way of interpreting a composition. As I shall argue, this is in itself a worthy goal.” (Burstein (2010): 1)

Burstein’s very emphasis here on Schenkerian *analysis* (rather than generation) already reveals the problematic surface-oriented take on Schenkerian theory that I just critiqued. That he should think that *this* way of thinking about Schenkerian theory has been the basis for certain scientific enterprises in music theory also reveals a misunderstanding of at least the scientific project that is generative musical grammar.

However, what is more instructive for our present purposes is how Burstein understands the practice of Schenkerian analysis, as illustrated by his discussion of how one might analyze the main theme from the finale of Beethoven’s Op. 31 No. 2 “Tempest” piano sonata. Burstein argues that most people would analyze this passage as possessing something akin to the structure shown in Example 1.2-39, a claim he stakes on having discussed this passage with numerous students and colleagues.⁷⁷ (I have labeled this an “uninterpretable analysis”, for reasons that will be discussed soon.) Even if one disagrees with the details of the analysis shown in the example, Burstein argues that most people would agree more with such an analysis rather than Schenker’s own analysis of this passage, published in *Der freie Satz*.

I will discuss Schenker’s analysis in more detail in chapter 2.2 (it is discussed there as Example 2.2-8), so for our present purposes all that matters is how the analysis presented in Example 1.2-39 differs from Schenker’s analysis. Example 1.2-39 takes the first twelve measures of the passage as prolonging tonic harmony. This prolonged tonic harmony supports the scale degree 3 *Kopftön* of the passage, i.e. F5, which is first attained on the downbeat of measure 8 after an initial ascent (or *Anstieg*) from measure 1, and is then prolonged to the downbeat of measure 12 (as shown by the horizontal beam connecting

⁷⁷ I was one of these students myself, and Example 1.2-39 is my own analysis of the passage, which I completed as part of a workshop on Schenkerian analysis conducted by Burstein.

Example 1.2-39. Beethoven, Piano Sonata #17 "Tempest", Op. 31 No. 2/iii: 'Uninterpretable' analysis

mm. 1 3 8 10 12 14 15/16

^ 3 ^ 3 ^ 2 ^ 1

i **V** **i**

Detailed description: The image shows a musical score for the third movement of Beethoven's Piano Sonata No. 17, Op. 31 No. 2. The score is written for piano and consists of two staves: a treble clef staff on top and a bass clef staff on the bottom. The key signature has one flat (B-flat). The tempo is marked 'mm.' (measures). The score is divided into measures 1, 3, 8, 10, 12, 14, and 15/16. Above the treble staff, there are accents (^) over measures 3, 8, 10, 12, 14, and 15/16. A dashed line connects the beginning of measure 3 to the beginning of measure 8. A slur covers measures 9 and 10. Roman numerals are placed below the staves: 'i' below measure 1, 'V' below measure 15, and 'i' below measure 16. The notes in the treble staff include a half note G4 in measure 1, a quarter note G4 in measure 3, a quarter note A4 in measure 8, a quarter note B4 in measure 10, a quarter note C5 in measure 12, a quarter note D5 in measure 14, and a quarter note E5 in measure 15/16. The bass staff has a whole note G3 in measure 1, a whole note F3 in measure 3, a whole note E3 in measure 8, a whole note D3 in measure 10, a whole note C3 in measure 12, a whole note B2 in measure 14, and a whole note A2 in measure 15/16.

the two F5s in mm. 8-12), before finally descending to scale degree 1 in the *Urlinie* to close the passage in mm. 13-15, supported by the prolonged V – I progression of these measures. In his analysis, Schenker recognizes the F5 of measure 8 to be the *Kopfton* too, also attained by an initial ascent – but he only takes the initial tonic harmony to be prolonged up to this point, i.e. only in the first *eight* measures of the passage, after which he believes *subdominant* harmony is prolonged (for the next three measures). This means that according to Schenker, the F5 in measure 12 (and its supporting harmony) does not continue the initial prolongation of tonic harmony in mm. 1-8 into mm. 9-12. Instead, he analyzes this sonority as an *Ausfaltung* (i.e. unfolding) of the following cadential 6-4 harmony of bar 13.

In other words, the disagreement between Schenker, on the one hand, and almost everyone else on the other (at least according to Poundie Burstein) is about what is prolonged in mm. 9-11. Whereas Schenker thinks that these measures prolong IV, leading to a prolongation of V in mm. 12-14, and then the final I of measure 15, almost everyone else would hear mm. 9-11 as prolonging I, thus continuing the prolongation of tonic in mm. 1-8, and leading to further tonic prolongation in measure 12, before the theme ends with the V – I progression of mm. 13-15.

Now, if we take Burstein's claim to be true – viz. that almost everyone would analyze this Beethoven theme *not* in the way Schenker analyzed it, the question arises as to why they would do so. One could resort to the old cliché that the analysis presented in Example 1.2-39 is how most people 'hear' the Beethoven theme. Burstein frequently uses the verb "hear" too, to describe the action that leads people to arrive at an analysis of the Beethoven theme similar to the one shown in Example 1.2-39. This is suspect, if for no other reason than the fact that we clearly do not know what it even means to say that someone 'hears' a passage in a certain way. Surely it does not mean that a person can be shown to actually hear the Beethoven theme in the manner of Example 1.2-39, if, say, his/her ears and auditory cortex were to be observed in real time while they listen to the theme (perhaps using the experimental paradigms of electroencephalography (i.e. EEG) or functional magnetic resonance imaging (i.e. fMRI). How would such a study even demonstrate, on the basis of something like cortical firing patterns, that the person hears tonic, and not subdominant, harmony as what is being prolonged in mm. 9-11 of the

Beethoven passage? Reasons such as these are why I have defended the idea that Schenkerian theory is not a theory of perception, and a Schenkerian analysis not a description of how people hear certain musical passages. Instead, analyses like the one presented in Example 1.2-39 depict how one *understands* or *conceives* of certain musical passages.

But there are different ways of understanding a musical passage too. One form of understanding is to recognize, intuitively, that a passage is grammatical, or that it conforms to how passages are supposed to be structured in an idiom. This is the kind of understanding that generative theory ascribes to those who have native competence in an idiom (e.g. a child proficient in piano performance, or a connoisseur of Classical music who grew up collecting records and going to concerts). However, another, rather different, form of understanding is involved in recognizing that a given grammatical passage has a *specific structure* (in which a specific sonority, for example, is deemed to be hierarchically superior to another), since this is the kind of explicit, ‘reasoned’ understanding of a musical passage displayed by someone who is taught to analyze musical passages according to certain explicit analytical methodologies (e.g. a music theorist). Ideally these two forms of understanding should converge, at least if one subscribes to the generative perspective on these issues – a music-theoretic understanding of a passage should ideally correspond to how people with native musical competence in that idiom actually understand that passage. But sometimes this is hard to establish experimentally. That is, a very different, and probably more difficult, experiment is required to establish whether people understand a certain sonority as being hierarchically superior in a passage, as opposed to the easier experiment in which one has to distinguish ungrammatical passages from grammatical ones – e.g. by deciding whether a given passage ‘sounds’ deviant or not.

In the case of the Beethoven Tempest finale theme, Schenker’s analysis, and the analysis presented in Example 1.2-39, are therefore instances of the second type of understanding described above, i.e. the explicit understanding of a passage displayed by a music theorist, and as revealed through a Schenkerian voice-leading graph. Ideally, one of these analyses should also exemplify the first type of understanding, i.e. the intuitive understanding of the Beethoven theme that someone with native fluency

in the Classical tonal idiom would have – but this is hard to demonstrate empirically. In this sense, the contrast between Schenker’s analysis of the theme, and the analysis presented in Example 1.2-39 is no different from the way two linguists might disagree about whether to analyze “the cat” as a noun phrase or a determiner phrase. In such a case, one hopes that at least one of the analyses reflects the intuitions of native English speakers. But again, this is hard to demonstrate empirically – as opposed to the much easier demonstration of the ungrammaticality of the phrase “*cat the” (relative to “the cat”), which one could do by just asking a native English speaker to decide which of the two phrases sounds deviant.

Given the divergent analyses of the Beethoven theme proposed by Schenker and Example 1.2-39 – and the difficulties inherent in deciding empirically which analysis corresponds to the intuitions of those with native fluency in the Classical tonal idiom – Poundie Burstein concludes that analysis, and particularly Schenkerian analysis, is subjective and has no basis in science, as evident from his statement that “unless one can provide experimental documentation, it seems fraudulent to argue that one’s own analytic interpretation is something that others are actually hearing subliminally” (Burstein (2010): 7), and moreover that:

“Appealing to Science might bolster [the] claim that one’s analysis is not simply a matter of personal opinion, but rather something that has cognitive backing, or objectively logical support. As far as most Schenkerian analyses are concerned, however, I doubt whether finding such scientific support is entirely possible or even desirable. What I find attractive about Schenkerian analysis is that it offers a powerful model that allows one to effectively relate subjective interpretations of nuances in a tonal composition, and for me, this is reason enough to recommend it as a useful analytic tool.” (Burstein (2010): 12)

This conclusion is problematic for several reasons. For one, it puts too much faith in the role of “experimental documentation”, especially of the perceptual type, in justifying the scientific basis for certain music-theoretic projects. The history of theoretical (as opposed to *experimental*) science shows that one can make scientific assertions about various phenomena, even without any immediate experimental confirmation for such assertions, as long as they provide lucid, coherent accounts of observed data – especially since such assertions might be difficult to justify experimentally for a variety

of methodological and technological reasons.⁷⁸ Moreover, competing assertions have been made about certain kinds of data in the theoretical sciences, again in the absence of empirical justification for one assertion over the other, without these assertions being deemed unscientific or subjective as a result.⁷⁹ Of course, a theoretical assertion about a certain phenomenon could be *wrong*, but that does not necessarily have any bearing on whether it is scientific or not.

The biggest problem with Burstein's assertion though is that it mistakes generative descriptions of musical structure for interpretative ones, and therefore *ex hypothesi* scientific descriptions of music for subjective ones. And it is musical ambiguity that lies at the bottom of all this too. To understand this, consider the fact that Schenker's analysis and the one proposed by Example 1.2-39 are descriptions of two separate S-structures, based on certain ideas about structure generation in tonal music. To this extent they can both be considered psychological descriptions of how the mind conceives of certain tonal structures, and therefore scientific descriptions of some aspect of tonality. (One or both of these descriptions could be wrong too – but, again, this has no bearing on their scientific 'status'.)

But when seen in purely analytical terms, divorced from their generative context – and especially when seen as *conflicting* analyses of the Beethoven Tempest finale's main theme – an issue of ambiguity appears to arise. That is, the conflict between the analyses proposed by Schenker and Example 1.2-39 suggests that the theme is ambiguous, meaning that our preference for one analysis over the other (i.e. how we choose to *disambiguate* the theme) is a matter of *interpretation*. And since this is governed by

⁷⁸ Three examples from theoretical physics in this regard are (1) the postulation of the elementary particle known as the "Higgs boson" in 1964, whose existence was not empirically confirmed until 2012, and that too only tentatively. (2) The postulation of another particle known as the "neutrino", first proposed by Wolfgang Pauli in 1930, and later by Enrico Fermi in 1933, the latter's paper proposing this particle being rejected by the journal *Nature* for being too speculative, even though the existence of the neutrino was later demonstrated empirically, but not until 1956. (3) The Big Bang model for the early development of the universe, which was first proposed by Georges Lemaître in 1927, but which was not confirmed empirically, that too only controversially, until two years later in Edwin Hubble's work on the phenomenon of redshift, and which received substantial confirmation only in 1964, with the discovery of cosmic microwave background radiation.

⁷⁹ An example of this is the debate between the Big Bang versus "Steady State" theories of the origins of the universe, neither of which was conclusively confirmed – without either being deemed unscientific or subjective – until the former proposal was validated with discoveries such as the one mentioned in the last footnote. In fact, the physicist Fred Hoyle, one of the developers of the Steady State theory, continued to assert the veracity of that proposal over the Big Bang one until his death in 2001.

extra-grammatical factors, it might come across as being a subjective phenomenon – but only when it is divorced from its proper grammatical, and therefore scientific, context.

The above discussion highlights how the generative approach to musical structure inherent in Schenkerian theory is one that focuses (or should focus) on grammatical description rather than extra-grammatical interpretation, akin to the Chomskyan approach to similar issues in language. And this is what makes this approach a scientific one too. But this does not mean that such an approach cannot at least countenance interpretation in an intelligent way. In fact, being able to do so is rather important for a generative theory since the issue of interpretation reveals how external conditions on language affect the workings of the grammatical system. As discussed before, conceptual necessity requires us to say that grammar exists only to meet external conditions imposed on it (by semantics and phonology, in the case of language) – so exploring these conditions is a legitimate enterprise within generative theory, even if this will not tell us whether a structure that is generated according to these conditions *will* be interpreted or not, for that is of course a performance, and not a competence, issue.

To this extent, one could understand Schenker's analysis of the above Beethoven theme as illustrating how certain external conditions affect the generation of the theme's structure – so long as one accepts that this structure is generated specifically to meet these conditions, as Minimalism necessitates. For example, if meter (and rhythm more generally) is understood as an aspect of musical phonology (just as speech rhythm, or prosody, is considered an aspect of linguistic phonology), then metrical well-formedness conditions may act as external, phonological, conditions on musical S-structure generation, which ensure that generated S-structures correctly map to musical phonology. (Perhaps at some musical equivalent to a PF level of representation.) In this light, one could argue that Schenker analyzes the Beethoven theme as prolonging tonic harmony only to the end of measure 8 (and not measure 12 as Example 1.2-39 suggests) because this allows the structure to meet certain external conditions on metrical well-formedness, which is what allows this theme to be interpreted by musical phonology. This is

essentially the argument I will make when we revisit Schenker's analysis of the theme in chapter 2.2, within a discussion on the connections between musical grammar and musical rhythm.

On these grounds, one could also argue that the analysis of the Beethoven Tempest finale theme presented in Example 1.2-39 is actually *uninterpretable*, because it fails to satisfy these very external conditions. Moreover, this analysis might be considered to fail other external conditions – perhaps a semantic one that requires that S-structures only have one moment of climax, which the two *Kopftons* in measures 8 and 12 of the analysis prevents from happening, and which might therefore prevent it from being interpreted by the external semantic systems (presumably at some musical LF level of structure). All of which would make Schenker's competing analysis of the theme a better one. On the other hand, one could argue that Example 1.2-39's analysis is better because it is more efficient, in the way it understands the first 12 measures of the Beethoven theme as prolonging just one harmony – and which accords with the Minimalist preference for structural economy.⁸⁰ But the important point here is that these arguments are not subjective, and are fully consistent with a scientific investigation of musical grammar and the musical mind.

One could argue that Carl Schachter's invocation of motivic factors in his above discussion of the Chopin mazurka is, similarly, a discussion of the external (in this case motivic) conditions that govern whether a

⁸⁰ One needs to be careful when invoking efficiency as a reason for preferring one structural analysis to another though. This can be seen in the difficulties inherent in analyzing the structures linguists call “garden path sentences”. An example of such a sentence is the famous “The horse raced past the barn fell” (Bever (1970): 316). The most efficient way to understand this sentence is from left to right, one word at a time, so that we first understand a noun phrase “the horse”, followed by what appears to be a verb phrase “raced past the barn”. But this leads to trouble when we then encounter the word “fell”. The correct analysis would be to understand “fell” as the actual verb phrase of this sentence, which has a long-distance connection to the initial noun phrase “the horse” – with “raced past the barn” really being a subordinate clause with a null complementizer. Re-reading the sentence with the complementizer inserted demonstrates this clearly: “the horse *that* raced past the barn fell”. So, the correct analysis is actually one that appears inefficient, because of the long-distance connection it requires between “horse” and “fell”, but this is the only way we can avoid being confounded by the final verb of such a sentence, after being initially ‘led up the garden path’. Moreover, this analysis seems inefficient only from a surface-oriented perspective. In a hierarchical, generative account “horse” and “fell” are close to each other, which they must be for “horse” to receive nominative Case from “fell”, after which “horse” is *moved* to its surface position far away from “fell”. So, surface-oriented readings are not necessarily the most efficient, nor are they necessarily even the correct, analyses of a structure – which is something anti-generative linguists, and especially music theorists, often forget, even when efficiency forms a central part of their theorizing (as in Tymoczko (2011)).

generated musical S-structure is interpretable – thus making the one analysis Schachter prefers out of the three possible ones he discusses also the only analysis that is interpretable. (Again, regardless of whether this analysis is actually interpreted or not, which is subject to various performance factors.) For this reason, it seems that there is a good reason to explore even the external conditions that affect structure generation within a grammatical theory of musical structure.

And this is essentially what one does when discussing how certain ambiguous musical passages can be disambiguated. So, it is possible that this is what Kofi Agawu was hinting at when he said, in response to Carl Schachter’s views on musical ambiguity, that “the clarity/ambiguity dialectic therefore urgently demands explication”. In this light, Agawu’s ideas on this topic are worth examining more, in order to understand how issues of interpretation may influence Schenkerian descriptions of grammatical structure in music – despite what I believe is Agawu’s incorrect evaluation of the role of ambiguity in music, as I discussed in this section. Agawu has discussed the role of interpretation within a theory of grammar under the various labels of “Schenkerian poetics” and “introversive semiosis”, so these terms will be worth exploring in greater detail – and I will turn to this exploration in chapter 2.1.

1.2.5. Conclusion: Schenker, Minimalism and the Simplicity of Generative Theory

In this chapter, I have tried to demonstrate the possibilities for a Minimalist approach to musical structure, and how Schenkerian theory already seems to imply such an approach, in its descriptions of how Western Classical tonal passages are structured. Sometimes making this connection requires revisiting aspects of Schenkerian theory, such as through the lens provided by Allan Keiler, and sometimes it requires supplementing or extending traditional Schenkerian theory in new ways, e.g. by introducing a Principles and Parameters perspective to it, and extending it to the study of idioms like Rock, and as we shall see in the next chapter, Indian music.

In the process, I have also attempted to demonstrate in this chapter how anti-Schenkerian approaches are often incorrect in their rejection of Schenkerian theory, and how the claims they make about music-language connections are often inaccurate – Lerdahl and Jackendoff’s assertions to this

effect being a particularly good case in point. Essentially, there seem to be many points of convergence between music and language, and between generative music and linguistic theory, and these can be observed by comparing the Schenkerian generative program in music theory with the Chomskyan Minimalist program in linguistics – much of which requires the several fairly technical discussions we have pursued in this chapter. From these discussions, we have drawn many important conclusions about the nature of generative theories of music and language too.

Of these conclusions, there is one that seems to be of overarching significance, which involves the role of *simplicity* in generative theories of music and language. The importance of simplicity has arisen in many forms in our discussions on generative musical and linguistic grammar in this chapter. It is inherent in the very name of the Minimalist Program. It forms the basis for the Cartesian and Humboldtian belief, and also the Schenkerian belief, as we have seen, in the infinitely *creative* nature of music and language, which only a simple (but elegant) model of the mind as proposed by generative theory can describe adequately, enhanced by the ability to perform recursive and transformational computations.

Simplicity furthers the search for descriptive adequacy by *underspecifying* musical and linguistic structure as well, which not only allows for the (parameterized) diversity of languages and musical idioms we see in the world, but which also allows individual minds to acquire competence in any of them. Moreover, simplicity allows generative theory to explain how minds can acquire such musical and linguistic competence by means of simple and elegant sets of principles and parameters – and this enables generative theory to aim for explanatory adequacy as well.

All of this leads to a picture of music and language that is significantly abstract – i.e. generative theory describes and explains musical and linguistic structure in terms of structures and principles that are often far removed from the surfaces structures we see across the world's musical idioms and languages. Abstract entities pose certain problems for scientific explanation. One has to not only justify their existence, but also find ways to explain how they arise or are governed by the laws of nature. In other words, this poses problems for those who wish to explain how music and language evolved, or how they

are ‘caused’ in the course of our musical and linguistic behavior. But the simplicity inherent in generative approaches to music-theoretic and linguistic explanation again saves the day here. By simplifying the architecture of the linguistic and musical mind, generative theory reduces the quantity of abstract entities and principles that demand explanation. Furthermore, by problematizing the role of adaptation in the evolution of music and language, and by taking musical and linguistic behavior to be *freely* creative, generative theory simplifies the role of causation in explaining music and language too.

Finally, simplicity makes generative theory streamline its explanation of music and language to only those entities and principles that are *conceptually necessary*, and it also forces generative theory to describe music and language only in terms of processes that are *economical*, both computationally and representationally. Which means that simplicity is what lies behind the more recent focus on general principles of economy in generative theory, which therefore allows it to understand music and language in ways that transcend explanatory adequacy.

Now concepts such as creativity, freedom, underspecification, conceptual necessity, and economy are part and parcel of the Minimalist Program in linguistics, and I have tried to show in this chapter how they are inherent in the Schenkerian approach to music too. Given how simplicity lies at the basis of this, we can now see how closely aligned Schenkerianism and Chomskyan Minimalism are too – and not because of some pre-theoretical commitment to Occam’s razor, but because music and language seem to be intrinsically simple in their nature and origins as well. In my opinion, this is what really justifies the two identity theses for music and language proposed in the last chapter.

The above conclusion might seem startling; especially in the face of the fairly complicated, technical descriptions of C_{HM} and C_{HL} we have entertained in this chapter. But I wish to end this chapter by conveying how all of this *is* essentially quite simple – albeit by no means superficial. In order to defend the complicated claims I have made, and will continue to make, in this dissertation, some fairly complicated arguments needed to be made, and some fairly complicated data had to be explored. But ultimately most of these claims should be *intuitive* (to use another term that has some significance within

generative theory). Scientific theories in general are marked by their simplicity and elegance, and the situation is no different for a scientific theory of music or language. So, to this extent, generative grammar, as a scientific theory of music and language, should be simple and elegant – and should therefore make intuitive sense even to those who are not familiar with all its technical intricacies and complications, or who are not aware of all the data that justifies this approach. This was certainly what drew me to Schenkerian theory, during my years as a student of cognitive science, prior to my becoming a professional music theorist. And this should also be evident from the tremendous, and unparalleled, explanatory power that generative grammar brings to a discussion about music and language – and not only in terms of their internal computational structure, but also in terms of their meaning and articulation, as we will see in further chapters.

Jonah Katz and David Pesetsky have discussed the explanatory scope of different types of theories of music and language, with the greatest scope being accorded to what they call a “Type 4 theory”, i.e. a theory that accounts for properties that are both unique to music, and which are also shared with other cognitive systems like language (Katz and Pesetsky (2011): 5). This dissertation certainly strives to develop a Type 4 theory based on Schenkerian and Chomskyan ideas, and it is my belief that Schenkerian theory, when put in its proper scientific context, should make intuitive sense to a linguist, who does not know music, and likewise Minimalism should make intuitive sense to a (Schenkerian) music theorist, who does not know linguistics – i.e. the validity of these theories lies not merely in their ability to explain the data, but rather in their formal simplicity and elegance that a theorist from another domain can recognize without being aware of the data the theory purports to explain. Therefore, I sincerely hope that the rather technical points I have made in this chapter do not obscure the big picture that is a Minimalist approach to musical grammar – viz. that such an approach really provides us with an intuitively simple, although extraordinarily powerful, way of thinking about music and language, as aspects of human nature.

Chapter 1.3

Minimalist Musical Grammar: (ii) Lexicon¹

The previous chapter discussed the computational aspects of musical grammar, including the inputs to and outputs from this grammar. One of the issues that proved to be particularly problematic in this chapter was the issue of a musical lexicon, i.e. whether music has ‘words’ like language that act as inputs to the grammatical system. I suggested that harmonic structures like chords might be taken to instantiate such a musical lexicon, given a particularly melodic, Schenkerian approach to harmony. However, this leaves unresolved whether other, primarily monophonic, musical idioms also have chord-like inputs to their grammar, even if chords are understood melodically to suit the monophonic textures of these idioms, and in light of the above arguments about the melodic nature of harmony. Without resolving this matter, we cannot argue for the universality of harmony, which leaves the lexical nature of chords undecided.

In this chapter I will argue that it is indeed the case that a musical idiom that has often been described as being monophonic, viz. the Classical music of North India, is actually built upon chord-like inputs. (This does not mean that Indian music, or indeed any other musical idiom, is made up of *triadic* structures in the way Western tonal music is, but only that these idioms possess structures that are chord-like in their grammatical *function* – a matter I will discuss in some depth in this chapter.) I will show that the *rāgas* of North Indian music are best described as having chord-like properties of the kind found in Western tonal music, but only when we understand chords in a specific way – namely, as being linear, *melodic* entities, which, as we saw in the last chapter, is exactly how Heinrich Schenker described chordal harmony (and which is why his theory is ideal for the kind of Minimalist, universal generative grammar of music being discussed in this dissertation).

¹ The material in this chapter was presented in a shorter version at the 55th Annual Meeting of the Society for Ethnomusicology in Los Angeles, under the title “Motivic Hierarchies, Grammar and Meaning in North Indian Rāga Music”. I would like to thank Richard Cohn, Matthew Rahaim, Hans Utter, and Richard Widdess for their helpful comments and questions during the meeting, and Kofi Agawu and Ustad Zakir Hussain for discussions that led up to this paper. A revised version will also be presented at the 3rd International Conference on Analytical Approaches to World Music in London, under the title “A Grammatical Isomorphism between Western Tonal and North Indian Classical Music.”

1.3.1. *Rāga* as neither “Scale” nor “Tune”

Before we can embark on this journey, we first have to understand what a *rāga* is, and this is by no means a simple affair. In fact, as the central term of the musical culture of the Indian subcontinent, the term “*rāga*” symbolizes many things. It invokes basic technical ideas about pitch structure in Indian music theory, but it also appeals to the rich diversity of personal and stylistic expression that it makes possible in Indian culture. Therefore, defining “*rāga*” becomes a high stakes affair – and not one that is immediately amenable to a universal grammatical perspective. This is why only partial or cryptic definitions of the term are normally proposed in Indian music scholarship, epitomized by Harold Powers’ famous statement in the New Grove dictionary, “a *rāga* is not a tune, nor a ‘modal’ scale, but rather a continuum with scale and tune as its extremes” (Powers (1980a)).

In this section, I will examine why labels such as “scale” or “tune” seem to be unsuitable for describing *rāga*. However, in the next section I will argue that the unsuitability of these labels extends to descriptions of *Western* music as well, particularly to the much-theorized common-practice musical canon that spans the masterworks of Bach to Brahms. This suggests that the problem of defining “*rāga*” is a much more universal problem about the adequacy of descriptions for any complex musical idiom, not just Indian music. As a consequence of this, I will argue that North Indian *rāga* music is best described in idiom-general terms, as just an instance of a *universal* human capacity for making music.

Let us begin by revisiting Powers’ definition of *rāga*. Powers’ invocation of the two conceptual poles of “scale” and “tune” in his definition has a historical basis. Indian theorists have been interested in descriptions of the scale-like properties of Indian music for centuries, but the modern relevance of “scale” as a descriptive term in Indian music is usually associated with the *rāga* classificatory scheme of the early 20th-century theorist Vishnu Narayan Bhatkhande (1860-1936). Bhatkhande was a proficient *sitār* player, and had devoted a large part of his life to the first hand study of the performance practice of many important *gharānās* (i.e. ‘schools’) of North Indian Classical performance. In the process, he had collected a multitude of traditional *rāga* compositions, and from these, in addition to his dialogues with

several master musicians over the years, he believed he had inferred the essential structure of most of the important *rāgas*. So, he commenced a project of classifying the important *rāgas* of North Indian music into a rational and comprehensive system using this data, which he published in his multivolume *Hindustānī Sangīta Paddhatī* (Bhatkhande (1909-32)). This work has come to have a lasting influence on contemporary Indian musical thought. As Harold Powers says in his Grove dictionary article on Bhatkhande:

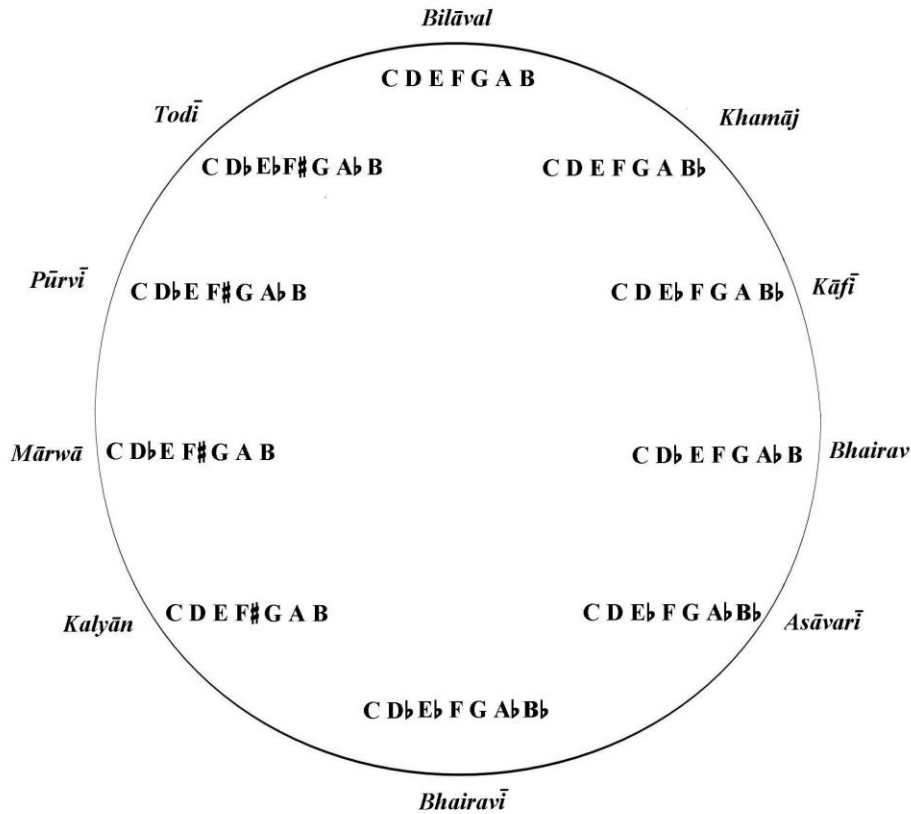
“His position with respect to raga in Indian music resembles Rameau’s with respect to harmony in European music, in that even those who most vigorously rejected his hypotheses have done so in terms he set.”

The most significant theoretical construct Bhatkhande developed was that of “*thāt*” – heptatonic scales constructed from the seven scale degrees in an octave, but with only one version of each scale degree (i.e. flat, natural or sharp) allowed in each *thāt*. Bhatkhande described ten such *thāts*, shown in Example 1.3-1, and believed that each *rāga* in the North Indian Classical tradition could be accommodated in exactly one of the ten *thāts*. Bhatkhande’s inspiration for the *thāt* system came partially from his experience as a *sitār* player, since “*thāt*” often refers to the setting of the movable frets of a *sitār* (movable, so that a *sitār* player can fret the different versions of a scale degree on the fingerboard of the instrument). There was also a historical precedent for his development of the *thāt* system since several attempts had been made in extant Indian music theory to develop a classificatory scheme for *rāgas*, one of which, viz. the South Indian Classical *mela* system of the 17th-century theorist Venkatamākhī, had actually influenced Bhatkhande in his theorizing, as he had learned of it during a trip to South India in 1904 (Powers (1992): 12).

However, despite this influence, Bhatkhande was in no way beholden to the *mela* system, and did not hesitate to reject historical ideas when he felt they contradicted the actual practice of music in the present age. As Harold Powers says:

“In the matter of reconciling the transmitted body of *sangīta-śāstra* [i.e. the canonical body of doctrine about Indian music] with the transmitted practice of *śāstriya-sangīta* [i.e. the tradition of musical performance], Bhatkhande’s stance had been essentially that of a Western positivist historian: he tried to

Example 1.3-1. The *thāt* system of *rāga* classification by Vishnu Narayan Bhatkhande



understand the practice in the light of the theory, induced general rules from the particulars of the practice, and saw no reason to deal with traditional theory that seemed on the face of it irrelevant to modern practice.” (Powers (1992): 37)

This attitude of Bhatkhande’s probably had something to do with his general outlook on Indian music, and the world of ideas in general, as an upper-middle class, university-educated, English-speaking lawyer – whose connection to Classical music was that of someone who traveled in Classical music circles, and was a highly cultivated listener, but who in matters of actually performing music (and thus having a vested interest in sustaining a deep-rooted performance tradition) was merely an amateur. Therefore, when it came to describing *rāga* structure he took the descriptive power of his *thāt* theory to be strong enough as to not warrant any additional reference to historical authority (such as that of ancient treatises).

A good example of this can be seen in one of the main problems with *thāt* theory, viz. its inability to clearly categorize some *rāgas* that have more complicated scale structures. *Rāga Bihāg*, which we shall explore in some detail later in this chapter, has a scale structure that uses both the natural and sharp version of scale degree 4, so it cannot be directly accommodated into any specific *thāt*. In the face of this, one could argue that the problem here lies with the *thāt* system itself – that the system should be modified to accommodate problematic *rāgas* like *Bihāg* more adequately. But Bhatkhande took the *thāt* system to be adequate and chose to solve the problem in a different way, viz. by asserting that one version of scale degree 4 is more important in *Bihāg* (which implies that *Bihāg* is essentially heptatonic) and can therefore be accommodated into the specific *thāt* that contains the more important version of the 4th scale degree.

Now the justification for this approach was not purely theoretical, despite the theoretical orientation of his system. He believed, quite correctly, that the natural-4th scale degree is more important than the sharp version in *Bihāg*, and therefore assigned it to the *Bilāval thāt* as opposed to the *Kalyān thāt*; two *thāts* that differ only in the variety of scale degree 4 present in them – and this decision was based on an examination of how the two versions of the 4th scale degree are actually treated in performances of the *rāga* by stalwart musicians. However, the fact that he still held the *thāt* concept to be sacrosanct in his system, a concept that had no direct basis in historical musical thinking and was inferred from observations of musical practice, reveals the strikingly positivist aspects of Bhatkhande’s thought.

Bhatkhande’s positivist insistence on the veracity of a *thāt*-based system of *rāga* classification might have been justified in the instances where *rāgas* with problematic structures (e.g. ones that possess multiple versions of a scale degree) were accommodated within a *thāt* more or less convincingly. But such accommodation is not always possible. For example, one could argue that a pentatonic *rāga* can *never* be assigned to a specific *thāt* because its relationship with the other *rāgas* in a *thāt* is one of having a common origin or of sharing important notes – not one of (heptatonic) *thāt* membership. Following this line of thought, one could also claim that grouping certain *rāgas* in the same *thāt* implies that they have historical similarities, not just pitch or pitch-intervallic similarities, when such historical similarities might not exist in reality. So, grouping them into the same *thāt* would contradict actual musical practice

and history. Since the names Bhatkhande gave the *thāts* are also the names of 10 prominent *rāgas* in the North Indian tradition, one could be led to believe that the *rāga* after which a *thāt* is named is the ‘parent’ *rāga* of that *thāt*, and the other *rāgas* assigned to the same *thāt* were all derived from it throughout the course of musical history. This criticism would be unfair since Bhatkhande never made any such assertions about the origin of *rāgas* within a *thāt*; but it is the kind of critique that has been leveled against Bhatkhande, and that is often leveled by the more historically-conscious against a system that is consciously ahistorical and positivistic.

This is why “scale”, in its guise as *thāt*, has often been considered inadequate (as is implicit in Powers’ above definition) for describing *rāga*. This is also why more historically oriented or less positivistic thinkers have often subjected Bhatkhande’s system to the kinds of attacks mentioned in the previous paragraph. The most notable of these attacks came from the famed vocalist and pedagogue Omkarnath Thakur (1897-1967), who laid out his views on *rāga* structure in his multivolume *Sangītānjālī* (Thakur (1938-62)). Thakur rejected outright Bhatkhande’s reliance on *thāt* membership as a criterion for grouping *rāgas*, and focused instead on the pitch properties of individual *rāgas* and their (alleged) shared history with other *rāgas* – a much more diachronic approach compared to that of Bhatkhande. Thakur also fleshed out his ideas about how *rāgas* are related as a tool for teaching *rāgas*, given his inherent interest in pedagogy. This reveals Thakur’s practical approach to *rāga* theory, as opposed to the theoretical perspective endorsed by Bhatkhande.²

In the first volume of his text, Thakur discusses six basic pentatonic *rāgas* in which the student is expected to learn some simple compositions and improvisation exercises. These *rāgas* are *Bhoopāli*, *Hamsadhvanī*, *Durgā*, *Sārang*, *Tilang* and *Bhinna-Śadja*. They are marked by their different intervallic structures, chosen to help the student train his/her ear to the structural differences between them, a task

² In fact Bhatkhande’s rather theoretical approach even to music pedagogy is revealed in his widely influential text *Kramik Pustak Mālikā* (Bhatkhande (1913-37)), which is a large compendium of vocal compositions that he had collected over the years, arranged by *rāga*, in which the only connection between theory and practice is a brief description of the *rāga* for each group of compositions at the beginning of that group, described in terms of concepts developed in the *Hindustānī Sangīta Paddhatī*.

aided by the pentatonic simplicity of the six *rāgas*. (As Harold Powers points out (Powers (1992): 20), this makes the choice of these *rāgas* striking, since students learning North Indian Classical music rarely begin their instruction with any of these *rāgas* – particularly the last one, which is still a relatively obscure *rāga* in terms of popularity, rarely performed even by masters.) As the student progresses, s/he learns more ‘complicated’ *rāgas* by either filling in the missing scale steps of the learned pentatonic *rāgas* to acquire hexa-, and heptatonic *rāgas*, or by altering one or more pitches of a learned *rāga* to create a new one. In this manner, the student learns new *rāgas* by focusing on the pitches that differentiate one *rāga* from another, also learning how *rāgas* are related in the process. So, Thakur’s approach to understanding the similarities and differences between *rāgas* (and therefore classifying them) was a more practical one, independent of abstractions such as Bhatkhande’s *thāts*. Thakur also tried to establish links between those groups of *rāgas* that he considered related within his system and various melodic types listed in ancient treatises like the *Nāṭyaśāstra* and the *Sangītratnākara*, which therefore endowed his system of *rāga* comparison with (in his opinion) a certain historical foundation.

The preceding discussion attempts to show how Omkarnath Thakur’s critique of Vishnu Narayan Bhatkhande’s approach reveals the limitations of describing *rāga* structure in terms of scale-like constructs such as *thāt*. However, it was suggested above that “tune” is also inadequate when describing *rāga* structure, and that this construct forms the opposite extreme of a continuum of descriptions of *rāga* structure, the truth lying somewhere in the middle according to Harold Powers. I think the best way to understand this assertion is by situating it within our comparison of Bhatkhande and Thakur’s music theories. And just as the merits of Thakur’s system helped reveal the inadequacies of a scalar description of *rāga*, as found in Bhatkhande’s approach, we will see that the problems inherent in treating *rāga* as a tune are revealed in some of the *de*-merits of Thakur’s system compared with that of Bhatkhande.

Both Thakur and Bhatkhande conceived of their theories as systems for distinguishing *rāgas*. Thakur did this in order to trace the lineage of, and teach students how to perform, specific *rāgas*; Bhatkhande did this in order to classify specific *rāgas* convincingly. But to distinguish between *rāgas* one

has to ascertain what counts as a similarity or a difference between them first. Only when such a determination has been made can *rāgas* that possess these similarities or differences be assigned to similar or different categories, and only then can the existence, or lack, of a historical relationship between them be asserted.

In traditional *rāga* theory, the “*vādi*” tone has often been taken as one of the indicators of similarity or difference between *rāgas*. This term is usually thought to refer to a sonant or prominent tone in a *rāga* (Jairazbhoy (1972): 64-65) – so two *rāgas* with different *vādis* would normally be considered different. Now, the use of the term *vādi* has actually been the subject of much debate, but the debate usually revolves around whether or not this term is synonymous with some other historical term in Indian music (such as *amśa* or *nyāsa*), or whether a specific pitch class in a *rāga* qualifies as a *vādi* tone or not – the general (and rather vague) definition of the term just mentioned usually goes uncontested. However, what makes a tone prominent is a highly complex issue, which definitely impacts our ability to decide whether a certain *rāga* pitch class qualifies as a *vādi* or not and therefore whether that *rāga* is different from another *rāga* – so, an explanation for why a *vādi* is a prominent tone is an important question that needs a definite answer.

In the interests of brevity I will only focus on one aspect of this question here. This is the fact that any tone (*vādi* or not) can be prominent in two different ways – either as a tone that is *markedly* prominent or whose prominence is *unmarked*, to borrow two terms from structuralist theory. A tone has its prominence marked when it is somewhat unusual or derivative – so a marked prominent tone in the context of *rāga* classification is a tone that distinguishes a *rāga* from another, usually because it is not shared between the two. An example of such a tone might be a highly dissonant pitch, which is striking to hear, and thus might help distinguish the *rāga* in which it appears from other *rāgas*. On the other hand, when the prominence of a tone is unmarked, this could be because the tone is very important for the *rāga*, but not as a distinguishing characteristic of the *rāga*. So, such a tone could be a structurally important tone that is actually *shared* between *rāgas*, because without it *rāga* phrases in any of those *rāgas* would

not be well formed. The best example of a tone whose prominence is unmarked is obviously the tonic pitch of a *rāga*, which is shared by all *rāgas* and is indispensable to all of them for structural reasons.³

From this discussion it might seem that the prominence of the *vādi* tone in a *rāga* is for the former reason, i.e. it is a tone that is prominent because it is marked – which is why the *vādi* is usually taken as a distinguishing feature of a *rāga*. However, this is exactly where one can find a significant point of difference between Bhatkhande and Thakur’s systems. As Harold Powers says:

“The note assigned [in Bhatkhande’s system] as a *rāga*’s most important note, its *vādi*, would always be the most audibly prominent one, with the *samvādi* [a note of secondary importance often paired with the *vādi*, and situated in the opposite half of the octave from the *vādi*] assigned four or five scale degrees away. Thakur’s objection was that those audibly prominent notes were of little or no use in distinguishing one *rāga* from another – certain pairings of audibly prominent notes occur in many different *rāgas* – and thus that may carry relatively little musical meaning.” (Powers (1992): 28)

For example, Thakur argues that Bhatkhande’s choice of scale degrees 3 and 7 as the *vādi* and *samvādi* of *rāga Yaman* is misplaced because the absence of these notes in a performance of the *rāga* would not ruin the identity of the *rāga*. That is, one would be able to tell this *rāga* apart from other related *rāgas* even if these notes are omitted in performance. On the other hand, the scale degrees (natural) 2 and 5 are crucial

³ Nazir Jairazbhoy (1972) discusses the two above ways of thinking about *vādi* prominence, but in slightly different terms. Citing Victor Zuckerkandl, he notes that the tonic could be considered the *vādi* in every *rāga* because of its “inherent dynamic function” (p. 68), which makes the tonic the perfect pitch for resolution. (This is a variation on the definition of *vādi* as an unmarked prominent pitch.) However, Jairazbhoy goes on to say that other factors like rhythm endow pitches with an “induced dynamic function” that overrides the importance of the tonic in favor of pitch structures that are symmetrical, either for rhythmic reasons or because of the melodic intervals in them. On these grounds, pitches that have an important role in defining such symmetrical structures (e.g. pitches that occur at the ends of two tetrachords that divide the octave into two symmetric halves) would be considered more important than the tonic. Therefore, such pitches would have greater claim to being the *vādi*. Since these symmetrical structures are obviously marked due to rhythmic or melodic-intervallic factors, the tones Jairazbhoy accords *vādi* status under this argument would be ones whose prominence is marked. Given the number of pages he devotes to exploring such *vādi* tones in different *rāgas*, and especially the detailed descriptions he gives of the symmetrical pitch structures in which such *vādi* tones occur, one gets the feeling that Jairazbhoy ultimately favors the ‘marked’ explanation for *vādi* prominence in a *rāga* (see especially p. 78).

Jairazbhoy’s description of the symmetry within *rāga* pitch structures is fascinating, especially the way in which he illustrates how these structures form statement-answer pairs in *rāga* phrases (p. 69). However, his bias towards arguments of markedness prevent him from explaining the ultimately *asymmetric* (and unmarked) nature of tonality – i.e. pitches that are prominent for unmarked reasons tend to form asymmetric, hierarchical relations with other pitches in a *rāga*. Which is why an asymmetry exists between the *vādi* and the *samvādi* pitches too – the *vādi* usually being the more prominent of the pair. Jairazbhoy recognizes that the pitches that ‘bookend’ two octave-dividing tetrachords normally form a *vādi*-*samvādi* pair, but he does not give a reason for why only one of them can be the *vādi*. This is not surprising, since in his symmetry-oriented approach, there is no way of distinguishing between the *asymmetrically*-related *vādi* and *samvādi* pair. There is also no way of explaining why the pitch classes that bookend these symmetric structures should be the most prominent ones in this approach, as opposed to the other pitch classes in these structures.

for establishing the *rāga*'s identity because without them *Yaman* could easily be confused with *Pūriyā*, which has flat-2 instead of natural-2 (Thakur (1938-62): vol. 2: 153-154, quoted Powers (1992): 29).

It is clear from the above that Thakur conceived of note prominence in *rāgas* in terms of markedness, i.e. the prominent note of a *rāga* is prominent *because* it distinguishes that *rāga* from other *rāgas*. It is not as clear how Bhatkhande conceived of note prominence in a *rāga*, but there are some interesting points that support his conception of note prominence in *unmarked* terms. For example, there are instances where Bhatkhande considers the *vādi* to be the prominent note of a *rāga* partly because it is the ending note of a *rāga* performance, though this is not always the case (Jairazbhoy (1972): 66, also note 6). Now, the ending note of a performance generally has unmarked structural importance as part of what brings closure to a musical phrase – hence, the structural importance of the final in Gregorian chant, and the tonic in Western common-practice tonal music. So, the ending note status of a *vādi* justifies its structural importance in a *rāga*. The fact that Bhatkhande's *vādi* is not always the ending tone of a *rāga* should not come as a surprise though, since there is at least one tone that is always structurally more important than the *vādi*, viz. the tonic (assuming that the *vādi* of a given *rāga* is not the tonic itself of course), which is therefore the tone that the *rāga* should end on more frequently than the *vādi*.

Bhatkhande also believed that the *vādi* is the most frequently occurring tone in a *rāga*. Jairazbhoy disputes this by showing that in Bhatkhande's own notations of *rāga* compositions the *vādi* occurs less frequently than other tones (Jairazbhoy (1972): 66). Ironically, this just goes to cement the *vādi*'s role as an unmarked, structurally important tone in a *rāga* because such tones often appear only infrequently in musical surfaces. In generative music theory, structurally important tones are often elaborated with less structurally-important tones (e.g. neighbor and passing tones) in the surface of a musical passage in a process called *prolongation* (a concept we explored in the previous chapters) – and these elaborating tones often appear *more* frequently than the structural tones they prolong because through prolongation a structural tone maintains its structural 'control' of the passage without literally being present all the time in the passage (Forte and Gilbert (1982): 142-144). (Since these tones elaborate a structurally more important tone by being dissonances that resolve to the more consonant structural tone, these elaborating

tones are therefore more marked in the surface of a musical passage too.) In Jairazbhoy’s analysis of the frequency with which different tones appear in Bhatkhande’s notation of *rāga Yaman*, Bhatkhande’s *vādi* scale degree 3 appears less frequently than scale degrees 2 and 5 (i.e. the pitch classes deemed prominent in this *rāga* by Omkarnath Thakur). (Scale degree 3 appears 70 times, whereas 2 and 5 appear 83 and 74 times respectively (Jairazbhoy (1972): 80).) Following the prolongational argument above, this just shows why Bhatkhande’s conception of the *vādi* seems to be that of a tone that is prominent in an unmarked way, whereas Thakur’s definition of *vādi* seems to be that of a tone that is prominent in a marked manner.⁴

On the basis of how they treat note prominence in *rāgas*, one could argue that the difference between Bhatkhande and Thakur lies in the fact that the former stressed the similarities between *rāgas* whereas the latter stressed the differences between them. (Though, of course, Bhatkhande’s very attempt to classify *rāgas* was clearly an attempt to distinguish them too (by putting them in separate *thāts*), and Thakur’s attempt to relate *rāgas* historically was certainly aimed at stressing the similarities between *rāgas* as

⁴ The reality of the matter is a bit more complicated than has been described above. Thakur actually did not put much emphasis on the term *vādi* at all, instead focusing on the term *amśa*, which was Thakur’s term for the prominent note of a *rāga*. So, Bhatkhande’s *vādi* and Thakur’s *amśa* seem to be just different ways of referring to the same prominent note of a *rāga*. However, the actual pitch classes in a *rāga* that instantiate these terms are different for Bhatkhande and Thakur (cf. the diametrically opposed views of prominence, i.e. marked vs. unmarked, seen in the two authors’ systems). And this difference is critical for understanding why Thakur chose the term *amśa* over Bhatkhande’s *vādi*. *Amśa* literally means “part of” – so that an *amśa* tone ends up being the essential *part* of a *rāga*, the tone that distinguishes the *rāga* from all others by its marked prominence in the *rāga*. We have seen that this seems to be a rather different definition of note prominence than that inherent in Bhatkhande’s “*vādi*”. Instead, Thakur seems to have thought that “*vādi*” should only be used to denote a tone that stands in a specific intervallic relationship with another tone (i.e. the *samvādi*) in the opposite half of the octave. Obviously, there will be cases in which the prominent note of a *rāga* also stands in an intervallic relationship with the *samvādi* tone, in which case the same tone will be the *amśa* and the *vādi* for that *rāga*. However, Thakur thought these were exceptional cases. He also thought that Bhatkhande treated these cases as exemplifying a rule rather than exceptions because the latter misunderstood one particular passage in the ancient *Nāṭyaśāstra* treatise, where *amśa* is equated with *vādi* (Powers (1992): 29-35). The *Nāṭyaśāstra* discusses three important tones in a *rāga*, viz. the *graha* (the beginning tone of a *rāga*), the *nyāsa* (the cadential tone), and finally the aforementioned *amśa*. We have seen that Bhatkhande’s notion of prominence allows him to use one label “*vādi*” for both the *amśa* and the cadential *nyāsa* tones. We have also seen how Thakur’s different conception of note prominence makes him assert a very different role for the *amśa* that cannot be accorded to the *vādi* or *nyāsa* tones – which is why he thinks Bhatkhande misreads the aforementioned passage from the *Nāṭyaśāstra*. Now, Thakur based his conception of the *vādi*, as having a lesser, intervallic, significance on his reading of the *Nāṭyaśāstra*’s treatment of this term too, whereas Bhatkhande was not particularly interested in basing his arguments on ancient treatises. So Thakur could be right in his assertion that Bhatkhande’s definition of “*vādi*” is based on a misreading of the *Nāṭyaśāstra*. But this depends in the first place on the *Nāṭyaśāstra*’s conception of *rāga* note prominence – and this is a matter that is open to speculation.

well.) Now if Thakur's critique of Bhatkhande shows why the notion of "scale" is problematic for *rāga* description, an opposite Bhatkhande-ian critique of Thakur's 'difference-oriented' system can show us why the notion of "tune" is equally problematic for *rāga* description. This is because "tune" is a concept that depends on notions of difference. A tune is a tune because it has certain characteristic features that help *distinguish* it from all other tunes – features such as a particular shape, characteristic motives or pitches, a certain register and tempo etc. It is because of these distinguishing features that tunes 'pop out' at us when they are played, and what often makes them so memorable. But this is precisely why "tune" is a bad way of characterizing *rāga*, for *rāgas* have features that are shared between them, which Thakur himself stressed in his theory.

Now Thakur never developed a theory that was based on a tune-like construct, in the manner in which Bhatkhande's system was based on the scale-like construct *thāt*. However, his emphasis on the identity of *rāgas*, on the features that make one *rāga different* from another, shows the same weaknesses in his system that it would have had if he had actually devised a tune-centered theoretical system for *rāga* description and classification. This not only goes to show why both "scale" and "tune" are problematic notions for *rāga* description, but also why the truth about *rāgas* seems to lie somewhere in the continuum between scale and tune, as Harold Powers pointed out in his definition of *rāga*.

Bhatkhande's system engaged with and tried to describe the tradition of North Indian Classical *rāga* music, but was not *of* that tradition – it neither developed out of, nor attempted to situate itself within, the tradition of historical Indian music theory. So, we could argue that "scale" in its guise as *thāt* becomes an *etic* concept in Indian music. We could also say that Omkarnath Thakur's insistence on creating his system from *within* the traditions of Indian performance practice and music history amounts to an *emic* critique of Bhatkhande – an approach that eventually prevented him from taking an equally *etic* stand as Bhatkhande, which might have happened if he had developed a theory based on the opposite, yet equally problematic, concept of "tune".

In this light, it might be that the unsuitability of Western labels such as “scale” or “tune” to describe *rāga* really just reflects the limitations of an etic approach to describing *rāga* structure, especially of the kind that seeks a universal lexicon for music, as this chapter does. Additionally, it might be that the continuum between scale and tune suggested by Powers is an emic one, and this is why it is the appropriate locus for finding a satisfying definition of *rāga*.

1.3.2. Melodic Structure and Hierarchy in North Indian vs. Western tonal music

However, the above conclusions would be valid only if concepts like “scale” and “tune” fare better in describing *Western* music. But this does not seem to be the case, as we saw in the last chapter’s discussion of the “theme and variations” second movement of Beethoven’s Kreutzer Sonata. Regarding that piece, we could say that the main theme of this movement is just a tune, but this would not explain the *similarities* between the theme and its variations in the movement, which is why the variations are variations of *that* specific theme and no other. On the other hand, reducing the theme to a mere instance of the F-major scale does not help either because this does not explain the essential *differences* between the theme and its variations, which make the variations *variations* of that theme and not exact restatements of it. But is this not the exact problem that impedes attempts to define “*rāga*”? After all, calling a *rāga* a tune does not help because this does not capture the similarities between different *rāgas*, or the similarities among the tunes we hear in multiple performances of the same *rāga*; and calling it a scale does not help either because this does not capture the differences between *rāgas* that share the same scale, or the different ways in which that *rāga* can be played by different performers.⁵

So, the problem of defining *rāga* and explaining its structure seems to be quite akin to the problem of explaining melodic structure in Western music. This shows us that the limitations in using “scale” to describe *rāga* is not a problem inherent in applying etic approaches to Indian music, but rather

⁵ The parallels between *rāga* structure and a “theme and variations” model in Western tonal music, such as the one present in the Beethoven Kreutzer movement, are especially relevant in a Schenkerian context, since it is the latter model that is often taken as a starting point for understanding the Schenkerian approach toward tonal structure in Western music. For example, one of the classic texts in Schenkerian pedagogy (Forte and Gilbert (1982)) takes precisely this approach toward describing tonality in its very first chapter.

one of applying the *wrong* etic approach, as exemplified by the problematic theory of *thāts*, as a result of which it continues to be difficult to describe *rāga* structure. But descriptions of Western musical structure have had more success, as we have seen. In the case of the Beethoven Kreutzer example, we know quite clearly that the similarities between the theme and its variations are the result of a common, abstract harmonic structure. Moreover, this similarity is realized in the form of shared, harmonically derived and hierarchically structured, pitches present in the melodies of the theme and its variations – structural motives as I have called them. And such a description of pitch structure in Western tonal music is indeed implicit in the Schenkerian approach to studying the grammar of Western tonality. So, we know from this approach that such hierarchies exist in Western tonal music.

Now, Schenker's theory attempted to describe the structure of the entire idiom of Western common-practice tonality. I will not attempt anything nearly as ambitious in this chapter. But I hope to show in the subsequent pages that adapting some Schenkerian ideas to North Indian Classical music can go a long way in demystifying the concept of *rāga*.

In an important study done in the eighties, three psychologists (Castellano, Bharucha and Krumhansl (1984)) showed that not only do pitch hierarchies exist in North Indian Classical music, they are actually understood by experienced listeners in this idiom. In contrast, American listeners unfamiliar with Indian music who were tested in the same study were unable to understand these tonal hierarchies in their proper grammatical context, hearing the *rāga* music played to them primarily in terms of the salience of certain pitches instead. So, the existence of tonal hierarchies in Indian music is very much a part of the cultural knowledge of *rāga* music in India. Harold Powers described such tonal hierarchies in South Indian music in his landmark dissertation on the subject as well (Powers (1959)), but tonal hierarchies are commonly described in North Indian music theory too. For example, the hierarchical superiority of the tonic and the dominant scale degrees, the *Sa* and the *Pa*, are implied in all Indian music, not least because a *tānpura* drone is usually tuned to these pitches.

Moreover, we saw in section 1.3.1. that the *vādi* can be seen as a structurally prominent pitch class in *rāga* music, under a certain reading of Bhatkhande’s approach to note prominence in *rāgas*. So, we could take the notion of the *vādi* as evidence for tonal hierarchies in North Indian music too. This might raise some eyebrows (despite the fact that Castellano et al.’s psychological results seem to confirm it) because there is often a lack of consensus over which pitch class constitutes the *vādi* of a *rāga* in Indian music theory even if we do define *vādi* unambiguously as the most prominent *unmarked* pitch class in a *rāga*, second only to the tonic. The lack of consensus over which pitch class in a *rāga* is the *vādi* of that *rāga* might suggest that the term is subjective and not a matter of acoustical or psychological fact, as tonal hierarchies in Western music are often taken to be. A few things can be said in this regard. First, if there is lack of consensus over the *vādi* of a *rāga*, the disagreement is far more restricted than is often believed to be the case. There are only a few pitch classes that can really be the *vādi* and they usually seem to be pitch classes that are fairly consonant with respect to the tonic (as is the case in Western music). The tonic *Sa* itself is the obvious candidate being the most consonant with itself; probably for this very reason it is often, unsurprisingly, considered the *vādi* of a *rāga* too. However, the *Sa* is not always considered to be the *vādi* since hierarchically superior pitch classes often tend to be *slightly* dissonant with respect to the tonic, as this generates just enough tonal tension for the music to move forward without losing structural coherence, before finally *coming to rest* on the tonic. In Western tonal music, this is seen in scale degrees 5 and/or 3 (rather than the tonic itself) being prolonged for long periods of time before a passage comes to rest on the tonic at a perfect cadence. And *vādi* pitches tend to behave in just this way – which supports their being treated as superior pitches in a tonal hierarchy. So, pitch classes that are slightly dissonant with respect to the tonic, such as the dominant, mediant and submediant (i.e. *Pa*, *Ga* and *Dha* in Indian solfege) are taken to be the *vādi* much more frequently than highly dissonant pitch classes such as the raised 4th (*teevra Ma*) and flat 2nd (*komal Re*).⁶ Also, the

⁶ We saw earlier that Omkarnath Thakur considered the 2nd scale degree to be a prominent pitch in a *rāga*, but he was thinking of note prominence in more marked terms, as is captured by the term “*amśa*” better than the term “*vādi*”.

majority of *rāga* phrases prolong the *vādi* with more dissonant tones before the phrase comes to rest on *Sa*, just as happens with hierarchically superior pitch classes in Western tonal music.

Therefore, the choice of *vādi* seems to be restricted to a small number of slightly dissonant, but generally consonant pitch classes. Even then, there can be disagreement over which *specific* pitch should be the *vādi* of a *rāga*, and this might prompt the evaluation of the *vādi* as a culturally relative, emic concept, in contrast to tonal hierarchies in Western music. But are tonal hierarchies in the latter tradition any more specific? If tonal hierarchies in Western music are determined on the basis of harmonic structure or bass lines, as they usually are, there is still frequent disagreement over which of two, generally consonant, pitches should be considered hierarchically superior to the other. This is precisely why Schenkerians can disagree over how to construe the Fundamental Line (i.e. Schenker's *Urlinie*) of a melody – should scale degree 3 be taken as hierarchically superior or scale degree 5? As an example, Schenker himself read the harmonically unambiguous first sonority of the opening of Mozart's A major piano sonata K. 331, with its explicit bass line, as a prolongation of scale degree 5 (i.e. E) (Schenker (1979): figs. 123 & 140), whereas Lerdahl and Jackendoff find good reason to read it as a prolongation of scale degree 3 (i.e. C#) (Lerdahl and Jackendoff (1983): 112-117). Though I admit that a more rigorous theory of pitch hierarchy in Indian music (including a more rigorous definition of “*vādi*”) is sorely needed, the above shows us that the *a priori* indeterminacy of tonal hierarchy is by no means any more characteristic of Indian music than it is of Western music.

Finally, there is a good reason for why there should *not* be absolute tonal hierarchies in *rāga* music, i.e. for why they must be decided on a case-by-case basis by a well-informed music analyst. This has to do with the fact that the hierarchical status of a pitch can change depending on context. Whereas pitch X might be superior to pitch Y in the tonal hierarchy of one musical passage, it might be inferior to it in another (e.g. the generally superior tonic becomes inferior to the leading tone in a cadential 6-4 progression in Western tonal music). In fact, a pitch can even prolong *itself* because of the recursive nature of pitch structures. Therefore, the very structural (and very etic) properties of pitches makes it

impossible for tonal hierarchies to be absolute.⁷ (An important point to note here for our subsequent discussion is that not every occurrence of a pitch class that is taken to be the *vādi* of a *rāga* will actually be the *vādi* because of the different contexts in which pitches occur. Only when a pitch in a musical passage gets endowed with the specific unmarked, structural importance accorded to the *vādi* tone in a *rāga* should it really be considered an ‘authentic’ *vādi*.)

In the above light, it seems clear that *rāgas* have tonal hierarchies. But they also have structural motives, since hierarchically superior pitches often form motives within the melodic fabric of a *rāga*. Because of their hierarchical pitch origin and their common existence in a number of *rāgas*, such motives play structural roles in *rāga* melodies. Moreover, when more dissonant pitches prolong these structural motives, they form characteristic *rāga* motives called *chalan* or *pakad* that help identify a *rāga* – in the way that dissonant pitches help distinguish one *rāga* from another by their marked prominence, as discussed earlier. For example, when the motive *Ga-Sa* is prolonged by the intermediate supertonic pitch *Re* we get one of the characteristic phrases of *rāga Bihāg*. Example 1.3-2 illustrates this better.

In this example, I have described how a characteristic, complete phrase in *rāga Bihāg* might be generated. Following our discussion of the phrase structure of the Kreutzer theme in the last chapter, I am assuming here that a musical phrase is generated by joining harmonies (and their melodic realization as structural motives) together in a hierarchical fashion, which can be illustrated by means of a tree diagram. Also, following Allan Keiler’s views on pitch organization within Schenkerian theory, a description of how such a phrase is generated must begin by illustrating how the two hierarchically most superior structures in the phrase are joined together. (In Example 1.2-5 in the last chapter, these two structures are the structural motive in measure 5 (that includes the *Kopfton*) and the structural motive in measure 8 that

⁷ However, this does not mean that a structural analysis of a musical passage is an ‘anything goes’ affair. Even though analysts have to make choices in deciding which pitches are superior or inferior in a tonal hierarchy, these choices are constrained by what reveals the structure of a passage best – what best reveals the harmonic organization or melodic trajectory of a passage. So, even if a highly dissonant pitch is locally stressed in the melody for aesthetic reasons, e.g. when it is a high point, that by itself will not make it a hierarchically superior pitch. Whether it is superior or not will be determined by the global context of the passage – if it is superior it will be further prolonged by non-adjacent pitches; if it is hierarchically inferior, it will prolong them instead.

Example 1.3-2. *Rāga Bihāg*: Phrase generation

(some \circ , i.e. *Kopfton* = Western Classical tonal music
Vādi = North Indian Classical music)

The musical score consists of seven staves, labeled A₁ through B₄. Each staff shows a sequence of notes and rests. Above the notes, there are various markings: a circled 'o' (representing Vādi) and a plain 'o' (representing Kopfton). Above some notes, there are numbers 3 and 1, indicating fingerings. Above one note, there is a circled 2, indicating a trill. Brackets are used to group notes into phrases. A large upward-sloping line above the staves indicates the increasing complexity of the phrase. A large bracket spans the entire piece.

ends the phrase, both being supported by stable tonic harmonies.) In the case of North Indian *rāga* music, the same generative procedure can be described, starting with the two hierarchically most superior

structures in the *rāga*, viz. the structural motives involving the tonic and the second most prominent (in an unmarked sense) pitch class in a *rāga*, which from the above discussion I am taking to be the *vādi* of the *rāga*. (This being scale degree 3, or *Ga*, in the case of *rāga Bihāg*.) Since North Indian Classical music and Western Classical music do not have any obvious geographical or historical relation, one way of explaining the convergence in their generative procedures is to postulate, as I have done throughout this dissertation, a *universal* grammar, a cognitive faculty that all humans possess innately, which allows musicians to generate similar musical phrases across cultures. As a result, the Kreutzer theme and a phrase in *rāga Bihāg* might be generated in very similar ways, both involving a descent to a cadential tonic pitch from the second most structurally prominent pitch in the idiom, i.e. the *Kopfton* in Western tonal music, and the *vādi* in North Indian Classical music. This is illustrated in level A₁ of Example 1.3-2, with the structurally prominent pitches in it notated (according to Schenkerian convention) with white note heads. (Note that this level exemplifies the concept of “Headline” I introduced in the last chapter.) The structural motive involving the *Kopfton* in Western tonal music often involves an ascent *up* to that pitch, which Schenker has described in his notion of the *Anstieg*. This initial ascent seems to be true of phrase structure in North Indian *rāga* music too, since a phrase in this idiom rarely begins with the *vādi* as the first note of a phrase (remember from the discussion in section 1.3.1 that the *vādi* is considered distinct from the *graha* tone even in Indian music theory, the latter being the pitch class that starts a *rāga*). As is more often the case, a phrase in North Indian music begins with a prolongation of the tonic, which serves to ‘hammer’ in the tonic in the mind of the listener, and then slowly progresses to the *vādi* as the phrase unfolds, finally returning to the tonic again to complete the phrase. This shared ascent seen in both Western and North Indian music is the reason for the bracket in level A₁ of the figure.

Level A₂ of Example 1.3-2 just makes the choice of *Kopfton* or *vādi* explicit by choosing scale degree 3 to play this role, and level A₃ completes the descent to the final tonic by prolonging the final tonic motive with scale degree 2. In Western tonal music, this scale degree would normally be harmonized by a dominant harmony, leading to a harmonic I – V – I Schenkerian *Ursatz*. However, the descent to scale degree 1 via 2 would only occur if scale degree 2 is actually present among the scale

degrees in an idiom, which is certainly not true of various pentatonic idioms, and of many Indian *rāgas* as well. For this reason, scale degree 2 is optional in the descent illustrated in level A₃, which is why it is shown in parentheses. Scale degree 2 is present in *rāga Bihāg* though, and the descent 3-2-1 is an important characteristic progression in this *rāga*. This is made explicit in level B₁ of the example, where scale degree 2 is represented as an actual (and not optional, parenthetical) pitch in a canonical phrase in this *rāga*.

Since levels A₁ to A₃ of Example 1.3-2 show a generative process that is common to North Indian and Western tonal music, I am proposing these levels as universal aspects of a generative musical grammar. However, from level B₁ onwards we enter the *idiom specific* aspects of North Indian music, and specifically of *rāga Bihāg*, since the actual pitches selected in the derivation so far are representative of the phrase structure of this *rāga*. Level B₁ itself adds a scale degree 5 to the structural motive of the *vādi*, since this is a characteristic progression in *Bihāg* phrases too. However, even though this level is an idiomatic generative level in North Indian music, it could be shared between *rāgas* that share scale degree 3 as the *vādi*, and which also share the order in which pitches appear in the temporal unfolding of the phrase. Level B₂ now derives a structure from level B₁ that is more distinctly *Bihāg* sounding because it introduces a motive within the larger structural *vādi* motive in which the *vādi* is prolonged by the natural-4 scale degree – leading to the characteristic *Bihāg* motive 3 – natural-4 – 3. As was mentioned earlier, when dissonant (and thus marked) pitches prolong a structural motive, we get a marked motive that can distinguish a *rāga*, which is known as a *chalan* in Indian music theory. So, by prolonging the *vādi* in a characteristically *Bihāg* sounding manner, the natural-4th scale degree introduces a *chalan* motive into the phrase, and converts it from being an instance of a more universal phrase type into an idiomatic *Bihāg* phrase. This is shown in level B₂ with scale degree 4 represented by an eighth note, which is the standard Schenkerian notation for a neighbor-note figure, since the dissonant, marked, scale degree 4 in *rāga Bihāg* characteristically elaborates the *vādi* scale degree 3 by an upper neighbor motion. Level B₂ also prolongs the initial tonic of the phrase with the leading tone, which yields another characteristic, marked motive in *rāga Bihāg*.

Levels B₃ and B₄ complete the derivation of a characteristic *Bihāg* phrase. This happens first by elaborating the scale degree 5 introduced in level B₁ with a scale degree #4 lower neighbor (remember that *rāga Bihāg* is noted for using both versions of scale degree 4 in the manner shown in level B₃). Then, scale degree 5 is elaborated further with a descending pattern made up of scale degrees 7 and 6 (which is often used to elaborate scale degree 5 when the large structural progression from the opening tonic to the *vādi* is realized as a *descent* to the *vādi* from scale degree 8 rather than an ascent from scale degree 1 – a phenomenon that occurs quite frequently in this *rāga*). Finally, the first and third pitches of the phrase in level B₃, scale degrees 7 and 3, are elaborated themselves by scale degrees 5 and natural-4 respectively. An important point to note here is that scale degrees that have an important structural status at a deeper level of structure, such as scale degree 5 in level B₁, can also have a less structural status at a shallower level of structure, as when scale degree 5 ends up prolonging scale degree 7 in level B₄, even though this latter scale degree is structurally weaker than the scale degree 5 in B₁. This just repeats the point made earlier about how it is impossible to determine the structural status of a pitch in a tonal hierarchy *a priori*, given the way structural pitches can themselves elaborate structurally weaker pitches at shallower levels of structure, and often even elaborate themselves given the recursive nature of the computations in musical grammar.

In the above manner, Example 1.3-2 shows how all the pitches in *rāga Bihāg* can be derived by a recursive, generative procedure to generate a characteristic phrase in the *rāga*. Even though I have de-emphasized the role of scales in Indian music, I have depicted the characteristic ascending and descending scales (i.e. *aroha* and *avroha*) of *rāga Bihāg* in the next example, Example 1.3-3, just to show how we can visualize the scale-like quality of *rāgas* by arranging in a stepwise way all the pitches of the *rāga* that were derived by the generative procedure in Example 1.3-2. However, as should hopefully be clear by now, these ascending and descending scalar representations of *rāga Bihāg* are merely by-products of a generative process in which the structural motives of the *rāga* are joined together – the *rāga* itself is not derived from a pre-existing scale, especially given the pitfalls of describing *rāga* structure in

Example 1.3-3. *Rāga Bihāg*: Scalar and motivic structure

Aroha / Avroha (Ascent/Descent)

Ni Sa Ga Ma Pa Ni Sa Sa Ni Dha Pa Ma Pa Ga Ma Ga Re Sa

Ni (7̂) motives Sa (1̂) motives

Pa (5̂) motives

Ga (3̂) motives

Weak #4 Bilāval ang

scalar terms that we explored previously. This point becomes even clearer when we examine the characteristic descent of *rāga Bihāg* shown in Example 1.3-3, which is not a straight scalar descent at all, but rather a *vakra* (i.e. zigzag) one, which betrays the presence of some of the marked motives that characterize this *rāga* – and which were themselves derived in Example 1.3-2 (for example, the scale degree 3 – natural-4 – 3 neighbor motive in the latter half of the descent, which was derived in level B₂ of Example 1.3-2). Several *rāgas* in North Indian music are described as having *vakra* structures in the literature, which just reveals their motivic rather than scalar structure. The rest of Example 1.3-3 illustrates the characteristic *chalan* motives of *rāga Bihāg*, which arise by prolonging the structural motives in the *rāga* with dissonant, marked pitches. These motives are derived from the hierarchically superior, structurally important pitches in the *rāga* of course, in this case *Sa* and *Pa*, and the *vādi Ga*. The leading tone *Ni* has a fairly important status too as the *samvādi* pitch of *rāga Bihāg*, but it is normally

inferior to *Sa*, *Pa*, and *Ga*. (For ease of reference, I have named these motives according to the most structurally superior pitch in it, analogous to the way that structural motives in Western music have one primary pitch within them that belongs to the *Urlinie*.)

Example 1.3-3 reveals an important fact about structural motives, which is that they are often shared between *rāgas*, something that was the focus of Omkarnath Thakur's views about *rāga* structure. At the bottom right of the image we see that the interval *Dha-Ga* (or A-E), which often appears in *Bihāg* phrases, is a *chalan* motive in *rāga Bilāval*. However, since *Bilāval* is the historically older *rāga*, the motive is named the *Bilāval* “*ang*” after it, *ang* being the Hindi version of the Sanskrit term *amśa*, which we have encountered before, as the pitch (or group of pitches) that identifies a *rāga*. Since *ang/amśa* means “part of”, the *Bilāval ang* is therefore the pitch/pitch-group that is the characteristic part of *Bilāval* and helps identify it – so that we can say that a part of the parent *Bilāval* appears in *rāga Bihāg*. This explains the strong relation of *Bihāg* to *Bilāval*, which is why Vishnu Narayan Bhatkhande placed *Bihāg* in the *Bilāval thāt* in the theory of *thāts* discussed earlier.⁸ It also explains why the sharp-4th scale degree *teevra Ma* in the bottom left *Ga*-motive in Example 1.3-3 must be played weakly, since its parent *Bilāval* has no sharp-4th scale degree among its pitches. (Incidentally, this de-emphasis of the sharp-4 in *Bihāg* helps distinguish it from *rāga Shuddha Kalyān*, in which the sharp-4th is emphasized, in the descent from *Pa* to *Ga*.)

The fact that structural motives are shared between *rāgas* reveals their diachronic aspect too, which is that *rāgas* have a history – they evolve in time and often from each other. In the context of *rāga Kāmod* in Example 1.3-4, this diachronic fact is made even clearer, since we see that various characteristic motives in the *rāga* arise from parts of other parent *rāgas*. Nazir Jairazbhoy observed this too when he tried to demonstrate how one *rāga* gives rise to another by introducing alternate notes in its

⁸ Note that the term “*ang*” is being used here as a structure that can help relate one *rāga* to another, even though it was previously described as a distinguishing feature among *rāgas*. There is no contradiction here; musical structures can have relative degrees of markedness, so that a pitch (or pitch group) that is marked enough to distinguish two *rāgas* that possess it from all other *rāgas*, might be still be *unmarked* enough to be shared between those two *rāgas*.

Example 1.3-4. Rāga Kāmōd: Ang structure

Aroha / Avroha (Ascent/Descent)

Malhār Ang *Hamūr Ang* *Kalyān Ang*

Kāmōd Ang (shared with Shyām Kalyān)

structure (Jairazbhoy (1995): 102-121). However, it is important to remember that structural motives are derived at different levels of the generative process, and therefore have varying degrees of hierarchical importance – the *Ni* motives in Example 1.3-3 are hierarchically inferior to the *Sa* motives since *Ni* itself prolongs *Sa*. Therefore, the same motive might be derived at different levels of structure in different *rāgas*, implying that it will not have the same structural status in both *rāgas*, even though it is diachronically shared between them. Therefore, one has to be careful when assessing the relationship between *rāgas* based on their shared *angs*, and must situate this comparison within a study of motivic hierarchies and generative processes in *rāga* phrase structure. Jairazbhoy did not do this himself, and was therefore unable to show, in my opinion, how *rāgas* can be different in spite of having shared *angs*, while simultaneously being quite similar to other *rāgas* despite not sharing many pitch classes or *angs* with them. (Incidentally, the widely accepted relationship between *rāga Bihāg* and *Bilāval* is actually based on the fact that their common *ang* arises at a rather deep level of generation in both *rāgas*, since it involves the *vādi Ga* of *Bihāg* which is also important in *Bilāval* – which therefore reveals the strong structural relationship between these two *rāgas*.)

Notice an interesting detail about the derivation of the *Bihāg* phrase in Example 1.3-2. There we saw an example of the fact that *rāgas* have hierarchically ordered pitches, and that these pitches form structural motives that, when prolonged in characteristic ways, often help identify the *rāga*. However, as these motives were derived at different levels of structure, they actually serve to prolong each other too – i.e. these motives have hierarchical relationships amongst *themselves*. So, we can say that phrases in *rāga* music are nothing but sequences of structural motives in hierarchical, grammatical relationships with each other. Since this is exactly how phrase structure was described in Western tonal music in the previous chapter’s discussion of the musical lexicon, my claim about *rāga* phrase structure has important implications for a universal musical grammar. Most importantly, it helps support the claim that music has a lexicon, which has universal attributes, and which can be understood in the context of the specific, melodic, approach to harmony implicit in Schenkerian theory. Therefore, I would like to examine how phrase structures arise in actual performances of *Bihāg* now, in order to show that the theoretical points and examples presented in the past few pages are not merely speculative assertions, but demonstrable facts about the similarities between Western and Indian music – and possibly all music – and which can be revealed by analyzing actual musical practices in the North Indian Classical idiom.

1.3.3. Motivic Hierarchies in *Rāga Bihāg*: Some Analytical Examples

Let us examine some excerpts from four performances of *rāga Bihāg*, in order to see how the hierarchical relationships between a small set of structural motives yield the variety of *Bihāg* phrases seen in them. All the performances we will examine are from the first few minutes of the introductory unmetered *alāp* section of these performances. The lack of a distinct meter will allow us to examine the pure pitch relations in this *rāga*, ‘uncontaminated’ by rhythmic, and other, factors.⁹

⁹ The recordings for these four pieces can be accessed from the following links:

http://www.princeton.edu/~mukherji/Bihag_BismillahKhan.wav
http://www.princeton.edu/~mukherji/Bihag_BuddhadevDasGupta.wav
http://www.princeton.edu/~mukherji/Bihag_HariprasadChaurasia.wav
http://www.princeton.edu/~mukherji/Bihag_VilayatKhan.wav

Rather than leave the best for the last, I will start with what I feel is the grandest of the four performances, one by the *shahnāi* legend Ustad Bismillah Khan, a transcription of which is given in Example 1.3-5A. (Only the maestro's performance has been transcribed here, not those of his accompanists, especially since they, for the most part, just play the tonic drone in the performance.) As the figure illustrates, the maestro commences his performance with two complete *Bihāg* phrases, represented by the two tree diagrams in the example – both of which correspond very closely with the phrase generated in Example 1.3-2. As was described in that example, the large-scale structure of both phrases is an ascent to the scale degree 3 *vādi Ga* from the initial tonic, which is followed by a descent back to the tonic. In Example 1.3-5A, both trees reveal this specific large-scale structure, as can be seen in the highest branches of the two trees, depicted with darker lines in the image. Also, just as in Example 1.3-2, both trees are generated by merging the two most hierarchically superior motives in the phrase, which in this case are the motives arising from scale degrees 3 and 1, shown by the brackets beneath the two staves in the image.

Let us now look at the structure of these two phrases in greater detail starting with the first phrase, in the upper staff. The slurs beneath them demarcate each structural motive from which this phrase is generated. At the very left of the phrase, we see that the first hierarchical relation between two structural motives is between a scale degree 7 *Ni* motive that resolves to the following, hierarchically superior tonic *Sa* motive, both *Ni* and *Sa* being represented with white note heads here. This relationship is represented by the two branches of the tree located at the letter a in the phrase. Since the *Sa* motive is made up of that single pitch, there is no slur underneath it; but the preceding *Ni* motive is prolonged by the scale degree 2 *Re*, shown to the left of the *Ni* with a black note head. In this manner, we see that the two leftmost branches of the tree realize the scale degree progression 7 – 1, with 7 prolonging 1, which is exactly what happens in the beginning of level B₂ in Example 1.3-2. As is typical in North Indian

(Note that the pitch of the recordings has been altered to match that of the second recording, in this case to a tonic (*Sa*) of C, for ease in comparing the performances and their transcriptions.)

Example 1.3-5. Rāga Bihāg: Tree diagrams of motivic hierarchies in actual performances: A. Ustad Bismillah Khan

Ust. Bismillah Khan

Classical music, the first *Sa* motive serves to cement the tonic of the passage, and thus acts as the initial structural tonic that begins the ascent to the *vādi* in the deepest levels of structure in this phrase. Therefore, the branch that represents this tonic is the dark line that extends all the way to the right to meet the *vādi*'s branch toward the end of the phrase.

Following the first pair of structural motives, another scale degree 7 – 1 pair of structural motives appears next in the phrase, joined at the letter b. This time both motives are elaborated to a greater degree than before by means of prolonging pitches; the *Ni* motive is actually elaborated by the tonic, notated with a black note head to the left of the *Ni*. Usually, the tonic is structurally superior to scale degree 7, but here it is structurally inferior and prolongs the 7, much as happens in the cadential 6-4 progression in Western tonal music – which emphasizes again the undesirability of deciding the hierarchical strength of pitches *a priori* in a musical idiom. The *Ni* motive resolves again to the structurally superior *Sa* motive that follows it, as the branches of the tree indicate. The *Sa* motive is itself elaborated by the pitch sequence B3-C4-E4-D4-C4. Now compare this pitch sequence with the first three notes in level B₃ of Example 1.3-2 – we see that those three notes are B, C and E, with the C being the structural tonic that is prolonged first by the B and then by the E. Notice how the elaboration of the *Sa* motive under discussion in Example 1.3-5A itself contains these very pitches, in the same order as they appear in level B₃ of Example 1.3-2. After the melody rises to the E4 in the *Sa* motive it descends back to C4 via D4, which is the characteristic, closing *chalan* motive in *rāga Bihāg* – and which also makes the *Sa* motive, seen in its entirety, resemble an entire *Bihāg* phrase, as illustrated in the very deepest levels of derivation in Example 1.3-2.

Considering all of these facts, one could argue that the *Sa* motive is a complete *Bihāg* phrase in itself, not just a single motive in a larger structural progression. Such a view would be consistent with a wider view in music theory that rebels against the kinds of large-scale hierarchical models of musical phrase structure preferred in Schenkerian theory, and exemplified by the two, large-scale tree structures in Example 1.3-5A. In fact, the argument for treating the *Sa* motive as a complete phrase in itself could even be applied to the next *Sa* motive in the performance, the one represented by the right branch at the letter c. This motive resembles a complete *Bihāg* phrase even more than the previous *Sa* motive does, since it

introduces more of the structural pitches of the *Bihāg* phrase derived in Example 1.3-2, and in the same order as they appear in that example too – for instance, compare this motive with level B₃ in Example 1.3-2, and notice how in both structures the initial tonic is prolonged by scale degree 7, after which the melody rises to scale degree 5 via scale degree 3 and then descends to 3 again before finally coming to rest on the tonic via a descent through scale degree 2.

The problem with this argument is that it is too ‘myopic’ – it does not consider the wider grammatical context in which these *Sa* motives occur. (Remember the point raised in footnote 7 about how one must consider the wider context of a passage to decide the grammatical status of a particular pitch structure.) We know from the derivation in Example 1.3-2 that the initial *Ni* motive that starts a *Bihāg* phrase resolves to a subsequent *Sa* motive, which is often the structural tonic that begins the progression towards the *vādi* in the deeper structure of the phrase. If we treat the *Sa* motives at letters b and c in Example 1.3-5A as independent phrases in themselves, then the preceding *Ni* motives at letters b and c will remain unresolved. A better solution would be to treat the entire *Ni* – *Sa* motive pairs of branches at letters b and c as complete *Bihāg* phrases. What we would have then are a succession of phrases at letters a, b and c, each phrase adding more pitch material to the previous one, thus endowing each successive phrase with a richer musical surface that increasingly approximates the canonical *Bihāg* surface shown in level B₄ of Example 1.3-2.¹⁰

However, this leads to a problem when evaluating the grammatical status of the motive at letter d. I see this motive as being a *Ga* motive that elaborates scale degree 3 on the way to scale degree 5 – as the third pitch in level B₃ of Example 1.3-2 does. This is why the E4 in the motive, shown with a white note head, is taken as the structural pitch of this motive. But from Example 1.3-2 we see that this pitch is preceded by the initial structural tonic of the phrase, my hypothesis being that this pitch (and the motive that arises from it) prolongs the initial tonic – which is why the motive at letter d branches off of the branch representing the initial tonic. However, if we consider the first three motives of the phrase, at

¹⁰ Incidentally, such a model of phrase structure, in which a musician allows a phrase to grow ‘organically’ by embedding more pitch material in successive reiterations of the phrase, resembles the models that the 19th century German music theorist Adolf Bernhard Marx proposed for phrase structure in Western tonal music.

letters a, b and c, to be independent phrases in themselves, then the *Ga* motive at letter d will not have an earlier tonic tree branch to branch off of. One might want to consider one of the two C4 pitches at the beginning of the *Ga* motive, shown with black note heads, as the initial tonic of the phrase for the E4 to branch off of – but this will not work either because for these pitches to act as the initial tonic, their role as initial structural tonic has to be cemented in the fabric of the piece first, which is normally done, as we have seen, by prolonging the tonic with a *Ni* motive. (The grace notes attached to these two pitches cannot be considered *Ni* motives either, since they are not independent pitches themselves. They are just the starting points for *glissandi* that the performer uses to sound tonic pitches in the melody – a standard procedure in Indian music, often used to mimic the fluid quality of Indian vocal music textures.) And since the maestro establishes the structural tonic at the beginning of the performance with a clearly articulated *Ni* pitch, it seems kind of nitpicky to force a ‘structural tonic’ reading on to the C4 pitches in the *Ga* motive at letter d.

In light of the above arguments, I read this whole passage as one large structure, in which the structural tonic motive is cemented right at the beginning of the phrase, and then prolonged by means of two more, right branching tonic motives at letters b and c, before the passage ascends to the *vādi* motive. In the process, the initial structural tonic is prolonged by the *Ga* motive at letter d, and the *vādi* motive itself is left-prolonged by the scale degree 5 *Pa* motive at letter e. Finally, the *Pa* and *Ga* motives at letter e are prolonged locally with the scale degree 4 neighbor note figures too, which hammers in the *Bihāg* flavor of the phrase beyond any doubt – and all of which happens in a manner that is consistent with what happens in levels B₁ to B₃ of Example 1.3-2.

Since the structural tonic that commences the progression to the *vādi* is cemented right at the beginning of the phrase at letter a, the two tonic motives at b and c just serve to cement this tonic further by branching off of the branch that represents the initial tonic. So, the initial tonic motive is being prolonged by subsequent structures that happen to be tonic motives themselves. This leads to a phenomenon in which a pitch structure is prolonged by self-similar pitch structures – an event that Lerdahl and Jackendoff call “strong prolongation” (Lerdahl and Jackendoff (1983): 182). Strong

prolongation of the *Sa* is a particularly important ingredient in Indian music, a remarkable example of which we will see in Example 1.3-5D. It also occurs after the *vādi* has been established in the first phrase of Example 1.3-5A. Rather than descend directly to the final tonic, via scale degree 2 (as occurs in every level of Example 1.3-2), Ustad Bismillah Khan does a sort of ‘summary’ of what has been accomplished so far in the phrase; after reaching the *vādi*, he goes back to the tonic and retraces the path back up to the *vādi* again, though he omits a lot of the rich pitch elaboration that took place earlier in the phrase. As a consequence we get a repeat of the initial scale degree 1 to 3 progression, which is represented by the branches at letter f in the example. (Also, the scale degree 3 motive is left-prolonged by a *Pa* motive once again.) By retracing the 1 – 3 progression, the maestro serves to strongly prolong the *vādi* motive with another *Ga* motive, which is why the *Ga* motive at letter f branches rightwards from the *vādi* motive.

The phrase descends to the cadential tonic after the ‘recap’ of the 1 – 3 progression has been accomplished. Bismillah Khan brings the phrase to an end in doing this, which is hammered in by the long pause that follows the end of the phrase, in which all the musicians play the tonic loudly and in unison. The fact that the phrase has ended is confirmed when the maestro commences a second phrase by following the same procedure he used in performing the first one – he returns to scale degree 7 and proceeds to cement the structural tonic (this time for the second phrase) all over again – as can be seen in the lower tree diagram in Example 1.3-5A.

What is most remarkable about the maestro’s performance of the first *Bihāg* phrase in my opinion is the elegance and beauty with which he executes it. He uses an abundance of dynamic contrasts (as shown in the transcription) to shape the phrase, which is unusual in Indian musical performance practice. This yields particularly ravishing results when he performs the highly dissonant #4 scale degree in the neighbor note elaboration of scale degree 5, and continues when he contrasts the #4 with the natural-4th scale degree to elaborate the *vādi* of the phrase. By means of these dynamic contrasts, combined with the nuanced yet languid, and indeed almost tranquil, manner in which he presents the melody, the maestro shows us what simple beauty a thoughtful, well-executed North Indian *rāga* can have. Yet this simplicity masks the dazzling complexity of the maestro’s phrase development; in a highly sophisticated display of

craftsmanship he elaborates the canonical phrase structure of Example 1.3-2 with motives that sound almost like complete phrases in themselves, by recursively embedding them within the larger structure of the phrase – a process that mirrors the ways in which Western composers often embed complete phrase-like structures (such as the canonical I – V – I progression) within a larger, yet self-similar, structure.

The lower stave of Example 1.3-5A shows the second phrase in Bismillah Khan's *rāga Bihāg* performance. Since the overall structure of this phrase is so similar to the previous one (which can be seen by just looking at the general contour of the two tree diagrams), this phrase will not require too much explanation. However, there are a few things that are still worth commenting on here. First, notice how the initial scale degree 7 *Ni* motive, which serves to prolong and thus cement the initial structural tonic of the phrase, is itself strongly prolonged by two more *Ni* motives that branch rightwards off of it. This makes sense given that we have already heard a complete *Bihāg* phrase now, so the performer can afford to pay greater attention to the details of the phrase – especially by elaborating the less structurally important pitches of this *rāga* such as *Ni*. In the manner that he did with the first phrase, the maestro elaborates the three *Ni* motives individually by embedding pitch sequences within them that resemble complete *Bihāg* phrases themselves.

After this prolongation of *Ni* we reach the initial structural tonic of the phrase. This time Bismillah Khan does not spend too much time prolonging the structural tonic motive, as he did at letters b and c in the previous phrase (presumably because he had already done exactly this in the first phrase). Instead, the phrase ascends directly into a prolongation of *Pa*, which we expect (from both the previous phrase and Example 1.3-2) to lead to a left-branching prolongation of the *vādi*. This is precisely what happens too – but not before the maestro lingers on the *Pa* with the #4th scale degree, played *ppp* and with such delicacy that it never fails to bring tears to my eyes in spite of the many times I have heard this passage. In fact, Khansaheb's gorgeous use of dynamics here almost evokes the image of a piper piping in the distance – which adds to the bucolic and serene ambience of the performance.

Subsequent to this passage, we arrive on the *vādi* again, elaborated by the upper neighbor scale degree 4 *Ma* pitch, and once more the maestro strongly prolongs the *vādi* by means of another 'recap'

passage before ending the phrase with a descent to the structural tonic. There is one more interesting event that happens here though, at the place marked with an asterisk in the example. There we see that the maestro sounds the *Ga* pitch with a downward glissando from the scale degree 6 *Dha*. Given its fleeting appearance, this pitch event might seem to be otherwise unremarkable. However, compare this structure with the *Ga* motives shown in Example 1.3-3. There we see that the pitch structure A-E is a characteristic *ang* of *rāga Bilāval*. As noted before, *rāga Bihāg* has strong affinities to *Bilāval*; so by playing these pitches Ustad Bismillah Khan reveals the *Bilāval* aspects of *Bihāg*. However, since this sequence is not characteristic of *Bihāg* itself, as *Dha* does not have the same structural importance in *Bihāg* as it does in *Bilāval*, it should not be over-emphasized in a performance of the former *rāga*, which might explain its fleeting appearance in the maestro's performance. In fact, one of things that makes Bismillah Khan's performance of *Bihāg* so exquisite is the fact that he avoids too many nonstructural pitches in his performance (compared with some of the later performances we will examine), which makes his rendition of *Bihāg* a much purer, traditional one. This adds to the aforementioned simplicity of the performance, with its greater emphasis on the natural-4th scale degree characteristic of *Bihāg*, but which also brings out the *Bilāval* flavor of *Bihāg*.

Example 1.3-5B shows a transcription of the first few minutes of the *alāp* from a performance of *rāga Bihāg* by *sarod* maestro Pandit Buddhadev Das Gupta.¹¹ Here again we have two complete phrases to commence the performance. Moreover, notice how similar these two phrases are in their general appearance to the phrases from Ustad Bismillah Khan's performance. This is especially striking given the important stylistic differences between the two performers. Whereas Bismillah Khan belongs to a essentially self-styled *shahnāi* performance tradition from the city of Benaras, Buddhadev Das Gupta belongs to the Senia Shahjahanpur *gharānā* of *sarod* performance from the city of Shahjahanpur, which

¹¹ This recording is the one from the *Raga Guide* (Boer et al. (1999): 48-49), for which the authors provide a transcription. I have re-transcribed it here to reveal its hierarchical, structural aspects.

Example 1.3-5. *Rāga Bihāg*: Tree diagrams of motivic hierarchies in actual performances: B. Pandit Buddhadev Das Gupta

Pt. Buddhadev Das Gupta

Middleground: $\hat{1}$ $\hat{3}$ $\hat{1}$

Background: $\hat{3}$ $\hat{1}$

Middleground: $\hat{1}$ $\hat{3}$ $\hat{1}$

Background: $\hat{3}$ $\hat{1}$

traces its roots to the famed Mughal court musician Tansen and to *rebab* players from Afghanistan (Chib (2004): 44).

Looking at the details of the performance, we see that Pandit Das Gupta starts the performance with the usual *Ni* motive prolongation with which we have come to expect *Bihāg* performances to open. In this case, the *Ni* motive is prolonged with a G3 pitch. Following this, the *Ni* motive itself ends up prolonging the following *Sa* motive, which is the initial structural tonic of the phrase that will begin the ascent to the *vādi*. Instead of prolonging the initial tonic further with more scale degree 7 – 1 progressions as Ustad Bismillah Khan did, the maestro proceeds to do something a little novel. He ascends to the *Ga*, which is elaborated with the characteristic upper neighbor natural-4, and then descends back to the tonic, which is the characteristic closing motive in *Bihāg*.

So, we have a passage so far that cements the initial tonic, rises to the 3rd scale degree and then descends back to tonic again. This seems like a complete *Bihāg* phrase in itself, just as the initial motivic structures in Bismillah Khan's performance seemed to be – which is why I have shown it in the example as having a 'middleground' structure that arises from joining a scale degree 3 motive with a scale degree 1 motive, represented by the brackets under the staff. However, as was also the case in Khansaheb's performance, reading Pandit Buddhadev Das Gupta's performance so far as a complete *Bihāg* phrase ignores the wider context of the phrase. The motive that follows this middleground 3 – 1 progression is the motive made up of the pitches G3, B3, C4 and E4, which is the *Ga* motive that prolongs the initial structural tonic, in the progression that leads to the *vādi* from that initial tonic. For the phrase to be grammatically correct this *Ga* motive should be able to attach to the structural tonic – but if we read the passage that precedes the *Ga* motive as a complete phrase in itself, then the *Ga* motive will not have a structural tonic motive to attach to. Therefore, just as in the Bismillah Khan excerpt, this passage should be seen as a coherent whole, in which the first *Sa* motive, prolonged by the preceding *Ni* motive acts as structural tonic, which is then right-prolonged by another *Sa* motive, and then the *Ga* motive just mentioned, before it progresses to the *vādi*. The *vādi* motive is itself prolonged, as usual, with a *Pa* motive elaborated with the lower neighbor scale degree #4, and then more locally with another *Ga* motive

that is elaborated by the natural-4th scale degree to realize the characteristic *Bihāg* motive of 3 – natural-4 – 3. (This is illustrated by the lowest branch that attaches to the *vādi* motive in the example.) Subsequently, the phrase descends back to the tonic via scale degree 2 to end the phrase in the expected, *Bihāg*-specific manner.

One more point needs to be made about the initial middleground 3 – 1 progression in this phrase. If this progression is taken as a complete phrase in itself, then the final tonic of the progression, marked by the scale degree 1 symbol, should be the hierarchically most superior pitch in the progression – in the way that the final, cadential tonic of the whole tree is the most superior event and hence gets the tallest branch in the tree too. In this light, it is this last tonic in the middleground 3 – 1 progression that should get the tallest branch in the tree, not the first tonic that was prolonged by the *Ni* motive that starts the phrase. But remember that this middleground progression should not be seen as an independent phrase by itself. Therefore it is not grammatically isomorphic with the entire phrase seen as a whole. So, the hierarchical structure of this progression is not the same as that of the whole phrase. The initial tonic of this progression is superior to the final one in the hierarchy of the phrase, and therefore gets the taller branch as the structural tonic motive of the entire phrase.

I do feel that this middleground progression is more structurally ambiguous than the motives in the Bismillah Khan example though, and I think this has to do with the quicker pace of Pandit Buddhadev Das Gupta's performance. Since Panditji does not linger on the initial tonic as much as Bismillah Khan does, or even as is typical in Indian Classical music, he introduces the later parts of the characteristic *Bihāg* progression quite quickly in his presentation of the phrase – which makes the phrase sound rather 'front-heavy'. This might owe to the pedagogical nature of his performance. Since Buddhadev Das Gupta made this recording for the *Rāga Guide* textbook on Indian music, his performances here are less elaborate and tend to 'get to the point' of a *rāga* rather more quickly than happens in a standard performance. In my opinion, this leads to the relatively quicker unfolding of the *Bihāg* phrase in his performance, which makes structurally weak motives appear to be complete phrases in themselves when they actually are not.

In the modern world of radio and television broadcasts of *rāga* performances, performers have often been led into taking the same approach to *rāga* development as Pandit Buddhadev Das Gupta does – an approach in which the elaborate, unmetred unfolding of a *rāga* in the *alāp* is replaced by a quick overview of the *rāga* before the performer begins to play a metered composition accompanied by a *tabla*. Such shortened, front-heavy *alāps*, called *auchār*, have become increasingly common as a symptom of the recorded age in Indian Classical music.

Moving on to the second phrase in Pandit Buddhadev Das Gupta’s performance, we see that he gets straight to the initial structural tonic, without elaborating it with the expected left-prolonging *Ni* motive, probably because that was already done in the first phrase. After this initial tonic has been sounded, the performer strongly prolongs it by means of two more tonic motives, both of which are left prolonged by a *Ni* motive this time – and as we can see, Panditji makes each successive iteration of the *Ni* motive more elaborate here, just as Ustad Bismillah Khan did at the beginning of the second *Bihāg* phrase in his performance. Subsequently, the phrase rises to the *Ga*, which is elaborated by a natural *Ma* as we have come to expect by now, before rising to the #4-inflected *Pa*. This *Pa* motive left-prolongs the following *vādi* motive, thus completing the ascent and allowing the phrase to end by descending to the final tonic two notes later. (Notice how Pandit Das Gupta does not do a ‘recap’ of the ascent in the way Bismillah Khan does after reaching the *vādi*; probably to stay in line with the short, expository character of the performance.)

The third example we will look at is from a performance by the flute maestro Pandit Hariprasad Chaurasia, shown in Example 1.3-5C. This example explores the structure of one complete *Bihāg* phrase, rather than two as was the case in the previous examples. Therefore the tree diagram here extends across both staves in the image, joined at the arrow signs.

Pandit Hariprasad Chaurasia belongs to yet another *gharānā* in North Indian Classical music, the

Example 1.3-5. Rāga Bihāg: Tree diagrams of motivic hierarchies in actual performances: C. Pandit Hariprasād Chaurasia

Pt. Hariprasād Chaurasia

(tonic prolongation)

(ś prolongation)

3

1

famous Maihar *gharānā* of the legendary pedagogue Ustad Allaudin Khan, who counted Pandit Ravi Shankar, Ustad Ali Akbar Khan, and Pandit Nikhil Banerjee amongst his disciples. In fact, Pandit Hariprasad Chaurasia's primary tutor was Allaudin Khan's daughter, the celebrated *sitār* and *surbahār* maestro Annapurna Devi. So, we can see that Pandit Chaurasia was schooled in a different style of *rāga* performance from Ustad Bismillah Khan and Pandit Buddhadev Das Gupta, in addition to being a specialist on a different instrument, and thus a different performance practice – yet his performance shows striking similarities to the previous performances. This suggests strong similarities between these performances at a deeper level of structure, since there are a few surface differences between Pandit Hariprasad Chaurasia's performance and the previous ones.

For one, Panditji gives us an example of a phenomenon, briefly alluded to earlier, where the unfolding of the *Bihāg* phrase proceeds by a *descent* from the initial structural tonic (i.e. scale degree 8) to the *vādi*, rather than by an ascent to that tone, before the *vādi* descends to the final tonic to end the phrase. Therefore, the initial tonic of the phrase is the C5, represented by the third slur in the figure, which descends to the *vādi* E4 in the second stave. But in spite of this difference, the large-scale structure of this phrase is identical to those in the previous examples. The maestro begins with a *Ni* motive, which is strongly prolonged by another *Ni* motive before it resolves to the structural tonic that begins the descent to the *vādi* – all of which corresponds to the canonical manner of commencing a performance in *Bihāg* that we have seen previously.

Pandit Hariprasad Chaurasia adds an interesting touch to the characteristic *Bihāg* opening here. He prolongs the second *Ni* motive by a glissando down to G4, represented by the triple grace note figure that follows the second *Ni*. While playing this figure, he gently introduces the non-structural, marked, flattened-seventh (or *komal Ni*) pitch B-flat to the phrase, which gives it a very idiosyncratic flavor not seen in the earlier performances. In this manner we see how imaginative performers can elaborate the shared (*ex hypothesi* universal) deep (or more accurately D-) structures in musical phrases with idiosyncratic gestures that make their performances stand out. Pandit Chaurasia has to be careful not to overemphasize the flat-7 pitch here though, because doing so would undermine the importance of the

natural-7th scale degree – whose importance, as the *samvādi* of this *rāga*, we have seen time and again. This might explain the fleeting appearance of the flat-7 within the triple grace note figure.

After establishing the structural tonic, Panditji prolongs this with another triple grace note figure. Now we have come to expect an ascent to *Ga* to follow after the initial structural tonic has been established, the *Ga* right-prolonging the tonic on its upward path to *Pa*, which then left-prolongs the *vādi*. (Refer to level B₃ of Example 1.3-2 to see how this happens in a canonical *Bihāg* phrase.) However, since Pandit Hariprasad Chaurasia progresses to the *vādi* via a descent from scale degree 8, the path to the *Pa* is not mediated by an ascent through *Ga*, but by a descent via *Ni* instead. So, the structural progression from initial tonic to *vādi*, which used to be an ascent like this: (7) – 1 – (4) – 3 – 5 – *vādi* is now a descent like this: (7) – 8 – 7 – (6) – 5 – *vādi*, with the prolonging pitches shown in parentheses. One can get an idea of how this descent is achieved by examining how scale degree 5 is prolonged by a descending scale degree 7 – 6 figure in level B₄ of Example 1.3-2. Pandit Hariprasad Chaurasia follows exactly this trajectory in his performance after the initial structural tonic has been established. He first right-prolongs the tonic by a *Ni* motive, which is the last slurred structure in the top stave of Example 1.3-5C. Then, in the first slurred structure in the lower stave, he prolongs a *Pa* motive, which eventually left-prolongs the *vādi* in the characteristic manner, the *vādi* arriving later in the lower stave at the scale degree 3 sign. The *vādi* is itself prolonged more locally by a left-branching 3 – natural-4 – 3 motive, which is the characteristic *chalan* motive of *Bihāg* as we have seen before.

The prolongation of the *Ni* and *Pa* motives prior to the arrival of the *vādi* are worth commenting on. As is shown in the example, the *Ni* motive is prolonged by an elaborate grace note figure marked “tonic prolongation”. Despite its elaborate appearance, all that this grace note figure does is prolong the tonic pitch, much as a double neighbor note figure around the tonic in Western tonal music often serves to prolong that tonic itself. In fact, the structure of this grace note quite resembles a double neighbor note figure around the tonic, because it essentially involves a rapid alternation between the C5 pitch and the neighboring B4 and D5 pitches (with the B4 itself alternating with the G4 in the first half of the figure). So, in the end, this complete figure just prolongs the tonic pitch C5, which then prolongs the *Ni* pitch that

heads this motive, in the way a tonic prolongs the leading tone in a cadential 6-4 progression in Western tonal music. The following grace note figure, in the beginning of the lower staff, has a very similar explanation as the last one – it is nothing but a neighbor-note like prolongation of the scale degree 5 *Pa*, which is to be expected anyway since we know that scale degree 5 is prolonged by the lower neighbor scale degree #4 in a characteristic *Bihāg* progression.

After we reach the *vādi*, Pandit Hariprasad Chaurasia does a recap of the progression so far, just as Ustad Bismillah Khan did in his performance, though Pandit Chaurasia’s recap involves ascending back up to scale degree 8 and then descending back down to the *vādi* again. This recap gives us a mini-phrase that starts with a repeat prolongation of scale degree 8 by a left-branching *Ni* motive, followed by a descent to *Pa* that, as we know, left-prolongs the *vādi* again – all of these pitches are shown by the white note heads that appear after the first descent to the *vādi*. As we reach the end of this recap, the second appearance of the *vādi* now strongly right-prolongs its first appearance, after which the usual descent to the tonic via the scale degree 2 *Re* occurs, which closes the phrase.

The last example we will look at is from a performance by the *sitār* maestro Ustad Vilayat Khan, shown in Example 1.3-5D. Yet again we have a performer from a completely different performance tradition, viz. the Imdadkhani *gharānā* of *sitār* performance, named after Vilayat Khansaheb’s illustrious grandfather, the 19th century *sitār* virtuoso Ustad Imdad Khan. For one, this tradition is a specialist string, and particularly *sitār*, tradition, which endows it with a very different performance practice compared to the flute and *shahnāi* traditions we have explored earlier. Moreover, the Imdadkhani tradition is noted for its unique *sitār* technique, which sets its exponents apart from even the celebrated *sitār* exponents of the Maihar tradition, such as Pandit Ravi Shankar and Nikhil Banerjee. The identifying feature of the Imdadkhani string technique is the way it involves bending the strings of an instrument to mimic the characteristics of vocal (*gāyaki*) style – which leads Imdadkhani *sitār* players to use very intricate grace note patterns in their playing to match the subtle inflections of the human voice. As the celebrated doyen

Example 1.3-5. *Rāga Bihāg*: Tree diagrams of motivic hierarchies in actual performances: D. Ustad Vilayat Khan

Ust. Vilayat Khan

Ĥ — = Ĥ prolongation)

Ĥ* — = Ĥ prolongation w/ Ĥ substituted by Ĥ)

The image displays four staves of musical notation, each representing a different performance of the Rāga Bihāg. Above each staff is a tree diagram that illustrates the motivic hierarchy of the piece. The notation includes notes, rests, and ornaments, with specific symbols like Ĥ and Ĥ* marking key moments. The tree diagrams are branching structures that map the relationships between different motifs throughout the performance.

Example 1.3-5. D. (contd.)

The image displays three systems of musical notation for a piece in D major. Each system consists of a single staff with a treble clef and a common time signature. The notation includes various rhythmic values, slurs, and fingerings. The first system shows a melodic line with slurs and fingerings (1, 2, 3, 4, 5, 6, 7). The second system continues the melody with slurs and fingerings (1, 2, 3, 4, 5, 6, 7). The third system concludes the piece with slurs and fingerings (1, 2, 3, 4, 5, 6, 7). The notation is presented in a clear, professional layout, suitable for a music textbook or score.

of this tradition in the 20th century, Ustad Vilayat Khan is particularly known for the way he lingers on pitches, elaborating them with extremely detailed grace note patterns, played with perfect intonation. (In fact, given his tendency to sound a pitch repeatedly, I have omitted these repeated notes from the transcription, except in places where these repetitions are crucial for the structure of the passage, such as the repeated E5s at the end of the bottom stave on the first page of Example 1.3-5D.)

In spite of these characteristics, Ustad Vilayat Khan's performance of *Bihāg* is again remarkably similar to the previous performances in its essential structural attributes. Before we explore these though, it is worth noting the rather different scope of this performance. For one, it is a much longer performance, and so the one phrase we will examine takes the maestro over five minutes to finish. Also, the wider scope of the performance allows him to linger on specific notes that have structural importance with great abandon, particularly the initial tonic – which he does by extended, and highly intricate, grace note play of the kind mentioned above. Just one quick glance at the complexity of the transcription in Example 1.3-5D confirms this.

The broad structure of this performance is an initial establishment of tonic, followed by a descent to the *vādi* via the *Pa*, and then the final descent to the cadential tonic via the *Re*. So, in its broad architecture this phrase is identical to the one performed by Pandit Hariprasad Chaurasia. However, in developing this phrase, Vilayat Khan spends the vast majority of his time just elaborating the initial structural tonic – hence the long dotted branch that extends from this initial tonic (which is the second slurred figure in the top stave of the example) all the way to the *vādi* branch in the last stave of the second page of the example. This allows him to show off his immense skill in elaborating the tonic with one intricate grace note figure after another. In the process, we have one tonic prolongation after another, all of which strongly prolong the initial structural tonic – hence the forest of tonic motives that right branch off of the initial tonic all the way into the second page of the example. This just goes to demonstrate why strong prolongation is such an important grammatical phenomenon in Indian Classical music.

Khansaheb begins his performance with the standard prolongation of the initial structural tonic with a prolonged *Ni* motive. Even here we see the maestro displaying the characteristic string-bending

technique of his *gharānā*. He plays the structural tonic by preceding it with a bent-string grace note figure that prolongs scale degree 7, just as the grace note figures in Pandit Hariprasad Chaurasia's performance prolonged the tonic and dominant scale degrees. (This is what the notation "7 ____" implies.) By means of these bent-string grace note figures, Vilayat Khan is able to prolong scale degree 7 in several imaginative ways, which allows him to play with the structural scale degree 7 – 1 progression for such a long time in the performance – which can be seen in the numerous instances of the "7 ____" sign in the transcription.

Notice that many of these scale degree 7 prolongations involve a grace note figure that actually begins on the 5th, and not the 7th, scale degree; e.g. the grace note figures towards the end of the first stave. This 5th scale degree was frequently present in some of the *Ni* prolonging motives in Pandit Buddhadev Das Gupta's performance too. Why not consider these motives scale degree 5 *Pa* motives then, rather than *Ni* motives? The answer has to do with the grammatical derivation of the passage. Scale degree 7 is the necessary pitch in this motive, which is why it was introduced earlier than the 5th scale degree in this motive in the derivation of the *Bihāg* phrase in Example 1.3-2. (Scale degree 7 was introduced in level B₂, whereas scale degree 5 was introduced in level B₄.) The necessity of the 7th scale degree vis-à-vis the 5th degree can also be seen from the way Ustad Bismillah Khan always uses it in his performance, to prolong tonic, whereas he *never* uses scale degree 5. This is why the scale degree 7 prolongations really prolong that scale degree and not scale degree 5.

A striking feature of the grace note figures the maestro uses to prolong scale degree 7 is the way he bends the *sitār* string up from scale degree 7, passes the tonic, and bends up to scale degree 2 before returning to the tonic again. This figure appears again and again in the transcription, particularly towards the end of the first stave and the beginning of the second. We have already seen that this pitch process helps prolong *Sa*, but it also allows Khansaheb to play around with the nonstructural supertonic pitch *Re*. This phenomenon is worth describing in greater detail. For one, the repeated sounding of *Re* gives Ustad Vilayat Khan's performance a certain idiosyncratic sound, much in the way the flat-7th scale degree did for Pandit Hariprasad Chaurasia – since *Re* is normally avoided in ascending progressions in *rāga Bihāg*,

and is usually used only to realize the descending scale degree 3 – 2 – 1 progression that ends a phrase in this *rāga*. Secondly, bending the string over multiple scale steps is technically challenging since it is hard to accomplish with perfect intonation. So, these multiple bent-string figures, which often span a wide scalar interval, allow the maestro to show off both his musical creativity and his technical skill. (In fact, most string players in the Imdadkhani tradition, and especially Vilayat Khan, are known for their ability to dazzle audiences with their virtuoso string bending, across unbelievably wide intervals of usually a perfect fourth or more, often at high speeds and in perfect tune.)

Finally, the second scale degree has a special relevance in *sitār* technique, when combined with the leading tone to form a double neighbor pair around the tonic. In closing sections of a *rāga* performance, *sitār* players, and sometimes other instrumentalists, play a sixteenth-note neighbor figure at high speeds with scale degrees 7 and 2 like this: 7-1-1-1, 2-1-1-1, 7-1-1-1, 2-1-1-1. The tonic pitches here are often struck on some special drone strings on the *sitār* called *chikāri*, which are often referred to as “rhythm strings” (e.g. Wade (1979): 95), because of the rhythmic character of the sixteenth-note figures played on them. When these figures are played by means of a rapid alternation between the *chikāri* and main melody strings (that sound the 7th and 2nd scale degrees) it leads to a bright, jangle of sounds, which is referred to by the onomatopoeic term “*jhāla*” – which when played at a fast tempo and a loud dynamic brings a *rāga* performance to a brilliant climax and a rapid, thrilling close.

In this light, scale degree 2 has a special importance for *sitār* players, especially in the Imdadkhani tradition, and when combined with scale degree 7. Therefore, in passages where scale degree 7 is being prolonged, it can be substituted with scale degree 2 – as an alternate way of prolonging 7. In Ustad Vilayat Khan’s performance of *Bihāg*, several of the tonic-prolonging motives in the earlier part of the phrase begin with *Re* rather than *Ni*. For the reasons just mentioned, I see these motives as essentially being *Ni* motives (given the role of these motives in prolonging tonic), but with scale degree 7 substituted with 2 instead. I have labeled these with the “7 * ___” sign. To my mind, these ‘alternate *Ni*’ motives not only help prolong the tonic in an imaginative way because of the 7th scale degree’s substitution with 2, but

also evoke the characteristic *sitār* stroke pattern that will eventually arrive towards the end of the performance to bring it to a close.

With the aid of the regular and alternate *Ni* motives, Vilayat Khansaheb prolongs the initial structural tonic all the way to the beginning of the fourth stave on page one of Example 1.3-5D. Towards the end of the third stave, he also begins to insert other pitches from the *Bihāg* ‘scale’ into these tonic-prolonging motives, often giving these motives a life of their own, just as Ustad Bismillah Khan, Pandit Buddhadev Das Gupta and Pandit Hariprasad Chaurasia did in their performances. The order in which these new pitches are inserted into the *Ni* motives are very similar to the way in which they were inserted by Pandit Hariprasad Chaurasia in his performance, given the similarity between the large-scale structures of the *Bihāg* phrases played by the two masters. (In this regard, compare the first motive in Example 1.3-5C with the penultimate motive in the third stave of Example 1.3-5D).

Beginning with the second pair of branches in the fourth stave of Example 1.3-5D, Vilayat Khansaheb begins to bend the melody string all the way up to scale degree 3, and not just scale degree 2. In so doing, the phrase begins to sound more like a canonical *Bihāg* phrase, because after bending the string to *Ga*, the maestro brings it back down to *Sa*, thus sounding the characteristic 3 – 2 – 1 motive of *Bihāg*, even if it is sounded within a *Ni* – *Sa* structural motive progression, as was the case in the first few motives of Bismillah Khan and Buddhadev Das Gupta’s performances (consider especially the initial ‘middleground’ motive of Pandit Das Gupta’s performance in this regard). Ustad Vilayat Khan plays with the descending scale degree 3 – 2 – 1 structure all the way to the end of the second stave on page two of the example. We also see a particularly rich prolongation of the *Ni* right at the beginning of this stave, illustrated with the long “7 _____” sign. The white note head shows that this is a long *Ni* motive, which rises to *Re* via *Sa*, as has been typical in this performance. But rather than resolve immediately to the following *Sa* motive, as *Ni* motives do here, Vilayat Khan further prolongs this *Ni* motive by some very intricate grace note figures, which trace out the *Ni*-prolonging neighbor note progression consisting of scale degrees 7 (B4) – 5 (G4) – 7 (B4) – 2 (D5) – 7 (B4). Only after this long prolongation of *Ni* has been accomplished does it finally resolve to *Sa*.

At the beginning of the final stave of this example, Ustad Vilayat Khan finally moves on to the next step of the progression towards the *vādi*. Since the large-scale structure of his phrase is a descent to the *vādi*, in the manner of Pandit Hariprasad Chaurasia, he will ultimately arrive at that pitch via a descent through *Pa*. But what Khansaheb actually does is very interesting – he descends straight to the 3rd scale degree, and then proceeds to progress *upwards* to *Pa*, in the way Ustad Bismillah Khan and Pandit Buddhadev Das Gupta did – thus combining the upward trajectory of their *Bihāg* performances with the downward trajectory of Pandit Hariprasad Chaurasia’s performance. As the first pair of branches in the bottom stave shows, the scale degree 3 that Khansaheb descends to (which itself right-prolongs the initial structural tonic as usual) is itself prolonged by a natural-4th scale degree grace note figure. Interestingly, this grace note figure has a fleeting #4 pitch embedded in it – an unusual, side-by-side juxtaposition of the two 4th scale degrees. Ustad Vilayat Khan commonly uses these kinds of juxtapositions in his performances, since his bent-string, *gayaki* way of playing the *sitār* lends itself well to executing subtle semitonal inflections of the kind found in Indian vocal textures.

Following the *Ga* motive, the phrase ascends to *Pa*. The usual lower neighbor #4 scale degree, which in this case is realized as a grace note figure that rapidly alternates between G4 and F#4, elaborates the *Pa*. Finally, after the long and intricate journey that the maestro has taken us through, we arrive at the *vādi*, which is left-prolonged by the preceding *Pa* motive – but not before the *vādi* motive is itself prolonged by a natural-4th grace note figure, which is the standard upper neighbor elaboration of the *vādi* in *rāga Bihāg*. This grace note figure again juxtaposes a fleeting #4th pitch within the natural-4th context of this figure.

After finally reaching the *vādi*, Ustad Vilayat Khan does a quick recap of the preceding phrase in the manner of Ustad Bismillah Khan and Pandit Hariprasad Chaurasia, though his recap is much more condensed given the amount of material that precedes it. This recap ultimately prolongs the *vādi* again, which strongly prolongs the preceding, structural *vādi* motive in the process. We have now come to expect that this strong prolongation of the *vādi* will allow the phrase to finally descend to the cadential

tonic. And this is exactly what the maestro proceeds to do, via the 2nd scale degree *Re*, to finally end this phrase – over five minutes after he started playing it.

With the above examples, I have tried to show how a basic D-structure underlies the varied performances of a *rāga* by four of the most distinguished musicians in the North Indian Classical idiom. I have also tried to show what a strong isomorphism exists between this structure and the phrase structures of Western tonal music, suggesting that these examples might serve as evidence for a universal musical grammar. In this context, notice a curious fact about the four performances analyzed in Example 1.3-5A-D. Not only are they generated from a common set of *rāga* phrases derived from structural *Sa* and *Ga* motives, they are comprised of generally left-branching tree structures too. That is, they are made up of phrases in which the rightmost motive is the highest in the hierarchy, with the other motives branching leftward from it. This is not particularly surprising, since the rightmost motive in most *rāga* phrases (and certainly in the four phrases just examined) is a tonic *Sa*-motive. This motive is the most superior in the motivic hierarchy because it resolves the tension generated by the previous, slightly dissonant, motives. As we have seen in previous chapters, this is exactly how phrase grammar in Western tonal music seems to work too, since Western phrases generally branch leftward from the cadential tonic harmony, which Allan Keiler calls “tonic completion”.

As we also saw in the last chapter, Keiler describes most Western phrases as starting with a *rightward* prolongation of an initial tonic too. This generally seems to be the case in the four *Bihāg* excerpts we have looked at, all of which prolong the initial structural tonic before commencing the progression towards the *vādi*. However, in all the examples, the initial tonic is first prolonged by a *Ni*-motive, which branches left from the *following Sa* motive. This disparity between Western and Indian music might be a result of the fact that the drone already sounds the tonic in Indian music, which preempts the need for the performer to start a *rāga* performances with the tonic as the first sounded note.

1.3.4. Conclusion: *Rāga* motives as an instance of a (harmonic) musical lexicon

Two striking features of *rāga* structure seem to emerge from the above discussion. First, it seems that we can now answer the question asked at the beginning of this chapter about what a *rāga* is, by saying “a *rāga* is neither a tune nor a scale, but a sequence of structural motives in hierarchical, grammatical relationships with each other.” And as we have discussed earlier, this reveals an intriguing similarity between *rāga* structure, and how tonal structure in Western Classical music was described in the last chapter. In this regard, we have come full circle, since the conclusion seems to be that both idioms are generated from a shared generative grammar of tonal hierarchies and harmonically-grounded, structural motives. Moreover, we have now seen that *rāga* motives are susceptible to prolongation, just like the chords of Western Classical music, and this even helps distinguish one *rāga* from another. So, what might have appeared to have been an emic question of defining a cultural construct such as *rāga* seems to be an etic question that we have now provided one answer to, by framing it in Schenkerian grammatical terms. The major implication of this is that a common harmonic lexicon seems to exist in both Western and Indian musical traditions, if indeed harmonic structures constitute a lexicon in the proper sense of the term – and if harmony is understood in the *Stufe*-based Schenkerian sense described in the last chapter.¹² Therefore, the question of whether music has a universal lexicon (without which it cannot have a universal grammar) seems to have a positive answer.

This answer needs to be qualified with three caveats though, so I will end by briefly discussing them. The first caveat has to do with the claim that North Indian *rāga* music can be understood in terms of a theory

¹² Recall in this regard the argument from the last chapter that just because grammatical functions are based on abstract harmonic foundations, it does not mean that “grammatical function” is equivalent to “harmonic function” (in the Riemannian sense). Rather, it is related to the Schenkerian idea of the scale-step or *Stufe*, as described in the last chapter. A *Stufe* is an abstract harmonic entity, but more tied to specific pitch classes than Riemann functions, since the latter can be realized by sonorities that do not even share the same scale-step (e.g. II and IV are different scale-steps, but are both accorded subdominant “S” function in the Riemannian system). I am asserting here that structural *rāga* motives are equivalent to scale-steps (and therefore have a harmonic foundation), but not harmonic functions in the Riemannian sense – especially since Riemann functions are more style specific too; it is not clear that the S function has any role in North Indian music (though the T (tonic) and D (dominant) functions might). Also, the notion that scale-steps II and IV both exemplify S function might be true of Western common-practice tonality, but is not necessarily true of Rock music, as was argued in section 1.2.4 of the last chapter.

of musical structure that is primarily harmonic in nature, because of the way I treat *rāga* motives as being analogous to (arpeggiated) chords in Western music. This claim might seem surprising in the light of the widespread belief that Indian music is purely linear or melodic.

However, by saying that Indian music is harmonic, I am not claiming that it is made up of vertical, triadic structures. This is because of the Schenkerian harmonic theory I am applying here, which should not be mistaken for a theory of vertical chords, such as those developed by Jean-Philippe Rameau or Hugo Riemann. As the preceding discussion illustrated, Schenkerian grammatical theory can be construed as a primarily melodic theory – a theory of structural melodic motives – that only get their grammatical *function* from an abstract, harmonic foundation. (Recall in this regard the discussion in the last chapter of the monophonic cello theme from Beethoven’s A major Cello Sonata, Op. 69, which emphasized this very point.)

So, the reason for my focus on the role of harmony in Indian music is really to bring out the abstract, grammatical properties that certain pitch structures have in that idiom – it is not meant to describe the actual textures of Indian music, which are clearly not made up of vertical sonorities, as is the case in most Western music. In this sense, what we are dealing with here is a *cognitive* theory of harmony, a theory that deals with how we *understand* the abstract grammatical properties of music, as embodied in its harmonic structure, even in a melodic idiom like North Indian music. Without this harmonic structure, a musical passage has no (hierarchical) grammar, but merely proceeds from note to note in first-order Markov chain fashion (see Richard Cohn’s discussion of the “overdetermined triad” (Cohn (1997)) in this regard). Therefore, primarily melodic theories of Western music (such as Narmour (1990): 96-280, and Huron (2001)) have often downplayed the hierarchical aspects of music in order to focus on issues of note-to-note voice leading or perceptual grouping. Even some Schenkerians who have taken a more melodic approach to musical structure have often focused on the processes by which note-to-note melodic motion takes place, thus giving greater emphasis to the role of voice leading in music over the more traditional Schenkerian emphasis on hierarchical harmonic structure. A good example of this is the theory of melodic ‘forces’ developed by the Schenkerian Steve Larson ((1997-98, 2004, 2012),

Larson and VanHandel (2005)) in which musical structure is described more in terms of the tendency of one melodic pitch to move to another due to principles of musical ‘gravity’, ‘magnetism’ and ‘inertia’ (which determine, respectively, whether a pitch moves down, up or stays where it is) rather than in terms of the harmonies that support those pitches. Even Fred Lerdahl, in his generative approach to musical structure (Lerdahl (2001)), acknowledges hierarchy only in the harmonic domain, since melodic structure in his theory is governed by principles of melodic *attraction*, i.e. the relationship between proximate pitches in a musical passage (not pitches connected by means of non-adjacent, hierarchical properties), a thesis that is very similar to Larson’s. Finally, Joseph Straus argues that it is precisely the inability to distinguish harmony and melody that leads to difficulties in developing a Schenkerian generative grammar for post tonal music (Straus (1987): 5-7).¹³

For the above reasons, I am proposing a Schenkerian ‘harmonic’ theory of grammar for Indian music precisely to account for the hierarchical aspects of Indian musical structure, without which we cannot account for grammatical function in *rāgas*. So in taking this approach I am implicitly arguing against prior attempts to theorize about Indian music in purely melodic terms (such as Cooper (1977) and Larson (2010)), which do not account for the truly *grammatical* (and especially recursive) properties of Indian music, because they do not consider the hierarchical, harmonic aspects of *rāga* structure in their approaches. Moreover, I am also arguing against those approaches that focus on the kinds of building blocks that are used to construct *rāga* phrases, but which have no immediate connection to the grammar of these phrases, being instead formulas or schemas that musicians acquire through rote learning or convention (as seen for example in Morris (2001, 2006) and Zadeh (2012)). By avoiding a discussion of the connection between grammar and the building blocks of *rāga* phrases, these approaches fail to shed light on the hierarchical, harmonic aspects of North Indian music, which make the generation of musical structures in both Indian and Western music possible in the first place.

¹³ Straus says that melodies prolong “harmonic intervals” in tonal music, a notion that is strikingly similar to the idea of a structural motive being proposed here.

This is not to say that the harmonic structures of Indian music and Western tonal music are identical. Just because harmonic lexicons might be universal, it does not mean that they are realized identically in every musical idiom, just like different languages have different words even if these words share certain universal lexical properties. Therefore, I am not claiming, for one, that harmony in Indian music is triadic. This leads to a significant *difference* between Indian and Western music. This is seen especially in the fact that the latter can have melodies in which one pitch from a triad *substitutes* for another pitch in a melody, which is why pitches can be implied in melodies in Western tonal music, as described in the last chapter. Not being triadic, such substitution does not normally happen in Indian music. The only exception to this that I am aware of is octave substitution, in which a pitch in a melody can be replaced by a pitch one or more octaves above or below without ruining the identity of the melody.

There might be another exception to this. Melodies in Indian music are often played in one half-octave, followed by a similar melody played in the other half-octave. Usually this amounts to performing a melody in the lower tetrachord (the *purvāṅga*) and then another melody in the upper tetrachord (the *uttarāṅga*), which is how a *sthāyi-antarā* phrase pair is usually performed. One could argue that the two melodies thus performed are really a theme and its variation, even though the pitches that make up these melodies are completely different, since they belong to different tetrachords. If this is really the case, then one could say that the second melody arises by substituting the first one with pitches from the other tetrachord – which would explain why the *vādi* and *samvādi* are considered analogues of each other in the two tetrachords of the *rāga* scale. (In other words, the *samvādi* pitch substitutes the *vādi* when the melody moves to the other tetrachord.) This does not mean that the two melodies can be played against each other contrapuntally though, as the different voices of a triadic chord progression can be in Western tonal music – so in this respect, the difference between Indian and Western music still obtains.

The second caveat to my claim about the universality of the harmonic musical lexicon has to do with the special role of musical meaning in this theory. One could argue that by focusing on universal lexicons and grammar, universal psychologically-grounded tonal hierarchies, and the cross-cultural similarities

between disparate musical cultures, I am ignoring or devaluing the culturally-entrenched modes of expression that are found in the way great individual performers and *gharānās* of *rāga* performance communicate through their music. This point is relevant since the structure of the lexicon is of utmost importance in determining how meaning arises in music and language, an issue that will concern us in the second half of this dissertation.

However, I do not think that I am ignoring how musicians express themselves in idiosyncratic ways in specific musical idioms – in fact, I believe I am doing the exact opposite of this, because only by isolating that which is shared between performers, styles and idioms, can we reveal the ways in which great musicians depart from these ideas and practices, and thus make celebrated contributions to a musical culture – a point emphasized by the music theorist Leonard Meyer as well (Meyer (1998): 3). It is precisely in this vein that I tried to show how four great *rāga* performers display their genius through the imaginative ways in which they ‘play’ with conventional *rāga* motives in Example 1.3-5A-D. And this is precisely the Schenkerian project in musical meaning too – as Kofi Agawu says (Agawu (forthcoming)), it is in the generation of diverse musical surfaces from common harmonic deep structures that we can distinguish “the conventional from the exceptional, the routine from the inspired”.

My final caveat to the above conclusions about the universal grammar and lexicon of music has to do with the idea that this chapter takes an etic attitude toward studying Indian music, as evident from my use of Schenkerian concepts to explore North Indian *rāga* structure. But I take issue with the very notion of “eticness”, complete with all of its postcolonial and Orientalist baggage. This is because the ability to understand hierarchical structures in music, and the ability to combine them to generate rich, often highly recursive, grammatical phrases is no more outside of the cultural experience of Indian music just because it was described by the Austrian Schenker. In fact, as we have seen in earlier chapters, the Minimalist Program argues that these abilities are uniquely human *cognitive* abilities that are grounded in human biology, and which distinguish us from other, non-human, animals. So, such cognitive processes form an intrinsic part of human nature – which is what even makes purportedly ‘emic’ phenomena possible in the

first place. By talking about motivic hierarchies, harmony, grammar and the lexicon in both Western and North Indian *rāga* music, all we are doing is revealing that music, following Longfellow, is the “universal language of humankind”, whether it is revealed through a masterwork by Beethoven or by Bismillah Khan playing *rāga Bihāg*.

Part II

Minimalist Music Theory and (Semantic/Phonetic) Interpretation

Chapter 2.1

Musical LF, Meaning and Agawu's "Schenkerian Poetics"

In this chapter, I turn to a consideration of musical meaning, and how this might be accounted for within a Minimalist approach to musical grammar. It is worth noting right at the outset, though, that musical meaning has often appeared to be a more intractable concept for music scholars than has musical grammar. It resists systematic attempts to define it, theorize upon it and build a consensus on how to go about searching for it in works of music. As Kofi Agawu says in a recent paper on the subject (Agawu (forthcoming)), "This vastness [of the scholarly literature on musical meaning] could be taken as a sign of the pertinence of the subject; it also dramatizes the absence of common points of reference for scholars interested in developing a universal theory of musical meaning."

Part of this intractability might have to do with the commonly held assumption that it is in the semantic sphere that music and language seem to differ the most. Linguistic meaning is taken to be capable of reference and of verification, which allows the linkage of meaning with notions of truth – but none of this applies to music in the strict sense. So, language models of meaning have only made the task of explaining musical meaning harder, unlike language models of grammar, which have been very fruitful in describing aspects of musical syntax. This, in turn, either leads thinkers to reject the similarity between music and language on the semantic front, or leads them to find other ways in which music might approximate language, in terms of how it relates to the world and is thus capable of some form of 'reference' or 'truth'. Given the difficulty of such a task, it is understandable why it has been difficult for scholars interested in musical meaning to find "common points of reference", to quote Agawu's above sentiment.

Now if one subsumes a discussion of meaning under a theory of grammar, then the inherent similarities between musical and linguistic grammar, as described in Part I of this dissertation, might allow one to find points of convergence between musical and linguistic meaning too. But the typical position in both discussions of music and language is to assert that meaning and grammar are distinct, and

that a grammatical theory cannot account for the meaningful aspects of either music or language. In my opinion, this in itself has been the first stumbling block that has impeded a clearer understanding of meaning in both music and language.

In the case of language scholarship, the view that grammatical theory cannot account for linguistic meaning is ubiquitous in discussions of meaning by non-linguists, and even linguists who do not subscribe to the Chomskyan approach to language. An example of this can be seen in a statement by the noted computer scientist Marvin Minsky, who says that:

“Chomsky seems almost entirely concerned with the formal syntax of sentences, to the nearly total exclusion of how words are actually used to represent and communicate ideas from one person to another. He thus ignores any models indicating that syntax is only an accessory to language. For example, no one has any trouble in understanding the story implied by the three-word utterance “thief, careless, prison” although it uses no syntax at all.” (quoted in Brockman (1995): 178)

A more complicated example can be found in the philosopher Michael Dummett’s rejection of the Chomskyan view of language in favor of one that views language as a communicative activity. Dummett says that for successful communication to occur the same *theory of meaning* must inform a speaker’s and a hearer’s communicative activity. However, he argues that in the Chomskyan view of language:

“The speaker’s knowledge of his language is identified with his knowledge of such a theory [of meaning]. Since the speaker obviously has no explicit knowledge of more than a few fragments of such a theory, this must be classed as unconscious knowledge; and, if anyone objects to the notion of unconscious knowledge, he is to be pacified by the replacement of the word “knowledge” by “cognition”. This is not offered as the outcome of conceptual analysis, but as an empirical scientific hypothesis, specifically a psychological hypothesis. It is guarded from attack as embodying a mythological conception of an immaterial psychic mechanism by the identification of psychology as neurophysiology on an abstract level: the theory of meaning is an abstract description of a structure within the brain. Since a subject cannot be expected to have even an abstract awareness of much that goes on within his brain, it is unsurprising that his knowledge of the theory should be unconscious. On the face of it, this account is open to the criticisms that Frege and Husserl brought against psychologism: meaning becomes private and hence no longer in principle communicable. This is to say that faith is required if we are to believe that we communicate with one another. The hearer must presuppose that he is interpreting the speaker as the speaker intends: but the speaker’s intention and the hearer’s interpretation are, at best, constituted by inner states of each respectively, not accessible to themselves, let alone to the other.” (Dummett (1996): 176-177)

In the Chomskyan perspective, knowledge of language is of course knowledge of grammar – which is an innate, and (as Dummett correctly says) unconscious, knowledge of the principles and parameters on

which one's native I-language is founded. Which means that in this perspective meaning is subsumed under a psychological, grammatical theory of a person's language faculty, as Dummett again notes correctly. But according to Dummett, this makes the Chomskyan perspective suspect, because a theory of meaning has to be explicit – and therefore presumably independent of the kind of grammatical theory Chomskyan linguists propose – for it to be able to account for how successful communication is possible. In other words, a theory of grammar says nothing about meaning in language, insofar as language is understood as a meaningful, communicative activity.

Dummett adds that the above is especially true if we believe that language is, as Dummett puts it, a “vehicle for thought”, which involves believing further that communication really involves a hearer's grasping the sense of the speaker's words. This is itself a mistaken belief because it depends on something mental, viz. the speaker's thoughts, represented by words, which are then communicated to the hearer – in a way that makes communication an act of faith:

“Communication is possible between two individuals, according to this picture, when they have the same thing in their heads: the business of the philosophy of language is to say what it is that they have in their heads that makes their utterances meaningful. The standard answer is that what they have in their heads is a theory of meaning, usually construed as a theory determining the truth-conditions of sentences of the language. ... This natural disposition was what was attacked by Wittgenstein, ... when he wrote that we have to make a radical break with the idea that language always functions in one way, always serves the same purpose: to convey thoughts – whether these thoughts are about houses, pains, good and evil, or whatever and that we are so accustomed to communication through language, in conversation, that it seems to us as if the whole point of communication lay in someone else's grasping the sense of my words, which is something mental, and, so to speak, absorbing it into his own mind.” (Dummett (1996): 186)

As I hope to show in subsequent sections, the above thoughts expressed by Minsky and Dummett are not without issue – in the very least they misunderstand the role of meaning within generative grammatical theory. But for the time being, all that is worth noting is that this type of criticism, i.e. against the attempt to subsume meaning within a theory of grammar, is arguably the majority position in language

scholarship, and is particularly prevalent in non-linguistic approaches to language, as in the examples from computer science and philosophy above.¹

Moreover, these ‘anti-grammatical’ positions all seem to depend on the belief that language is used for communication, and that this should figure in a theory of linguistic meaning. That is, a theory of grammar is not sufficient to deal with the issue of meaning in language because for this one has to contend with the extra-grammatical communicative uses of language. This assumes of course that linguistic meaning is something external to grammar itself – it is something that must be understood in the context of phenomena *external* to grammar, for example, by reference to objects in an external world. For scholars who take this attitude, meaning is therefore a form of *external* or *extrinsic* signification – it occurs by signifying realities external to the grammatical structure of language itself. Therefore, theories of meaning that understand meaning in this way are *externalist* theories of meaning.

2.1.1. Internalism vs. Externalism in Musical Meaning

As is the case in language scholarship, such externalist theories of meaning dominate musical scholarship too, which can be seen in the diversity of research paradigms within music in which it makes its appearance. Here are three examples, from an eminent music theorist/semiologist, a music psychologist, and a philosopher of music:

“If one considers that the sense of music resides in the play of its forms, everything is semantic. If, on the other hand (and this is my view), one admits that it is possible to distinguish between intrinsic and

¹ Another example from the world of philosophy is John Searle’s well-known Chinese Room thought experiment (discussed in Searle (1997)), in which someone who can form Chinese sentences by manipulating some symbols according to a rulebook for doing so will appear to be thinking in Chinese, even if he doesn’t know Chinese – so long as native Chinese speakers cannot reliably distinguish his sentences from those of an actual Chinese speaker. Searle conceived of this thought experiment to refute Alan Turing’s eponymous test for human intelligence, in which a machine is said to be capable of thinking like a human if it can form sentences that other humans cannot reliably distinguish from those produced by a human – Searle’s point being that all this demonstrates is that the machine can manipulate symbols in human fashion but cannot actually think like a human, akin to the person in his Chinese Room thought experiment. But what this also demonstrates is that Searle believes that thinking in a language involves knowing a language – and for Searle this amounts to more than being able to form grammatical sentences in that language. In other words, knowing how to combine symbols grammatically is not the same as knowing what those symbols *mean* in a language – knowledge of a language’s grammar is not equivalent to knowledge of the language itself, which is exactly the opposite of what Chomskyan linguists believe to be the case.

extrinsic significations, then the term ‘musical semantics’ can be reserved for the exploration of the latter category.” (Jean-Jacques Nattiez (1989): 33)

“The principle issue in musical semantics is whether we can establish analogous equivalences between music and some essentially non-musical phenomenon. Does music cause psychological effects which can be caused by other means, and does it do this in a systematic way? The first position we might consider is that music simply has no semantics. On this view, music is psychologically self-contained, a separate species of psychological activity for which unique modes of representation have been developed. ... However, I believe that the available evidence forces us to accept that there is some ‘leakage’. Musical experience *is* translated in other representational modes. ... For a great number of us, then, music has extra-musical meaning, however intangible.” (John Sloboda (1985): 58-59)

“To understand the music is to recognize the gesture which it enacts. ... To understand musical meaning, therefore, is to understand how the cultivated ear can discern, in what it hears, the occasions for sympathy. ... Someone who described the first chord of [Wagner’s opera] *Tristan* as an inversion of a G# minor triad with an added major sixth would be *wrong*, even though that description correctly identifies all the pitches of all the tones. He would be wrong because what we hear lies outside the musical possibilities of such a chord, outside what is implied in the idea of a minor triad. And someone who described the last movement of *The Jupiter Symphony* as morose and life-negating would also be wrong, however clearly his judgment was founded in theory. For this description is of something that cannot be heard in the music by a musically cultivated person.” (Roger Scruton (1987): 170-175)

Kofi Agawu’s description of the vastness of the literature on musical meaning, stated at the beginning of this chapter, really points to the diversity of externalist music theories of the kind Nattiez, Sloboda and Scruton advocate. For instance, one such theory proposed by Lawrence Kramer (2001), suggests that meanings in music arise from people with certain cultural identities making music in certain historical contexts. This approach has been the driving force behind more hermeneutic approaches to musical meaning, especially those that deal with cultural constructs of gender (McClary (1991)), race (Baraka (1963)), sexual orientation (Brett (1994)) and disability (Rodgers (2006), Straus (2006)). In contrast, research into entities known as musical *topoi*, advanced by Leonard Ratner (1985) and developed by many others (notably his students Wye Allanbrook (1992, 2008) and Kofi Agawu (2008a)), sees musical meaning in terms of characteristic gestures, or “topics”, that are “associated with various feelings and affections; others [having] a picturesque flavor” (Ratner (1985): 9). In this way, topics “point to the external world and thus facilitate recognition of the origin of a work’s constituent gestures in an ‘outside’ world” (Agawu (forthcoming)). This way of theorizing musical meaning has been especially effective; Robert Hatten (2004, 2009) has developed the idea of topical musical meaning into an elaborate theory of

expressive gestures in music, and Raymond Monelle (2006) has developed these ideas in the context of discussions about musical style and temporality.

Other approaches show a convergence of interest in the conceptual aspect of musical gestures inherent in topic theory with the psychological aspects of musical meaning. For example, Michael Spitzer (2004, 2008) has argued how the metaphorical aspects of musical thought can reveal the ways in which musical meaning is culturally founded. Lawrence Zbikowski (2002, 2008) has examined this metaphorical aspect of musical meaning in terms of its relation to musical grammar, influenced by similar projects in the anti-generative “cognitive linguistics” paradigm of language scholarship associated with the work of George Lakoff (1990) and others.

The conceptual aspects of musical meaning show how a theory of meaning can be externalistic without being referential, i.e. without relating music to the outside world in some way. So, externalist theories can also be externalist if they find meaning in an ‘inside’ world, a psychological world, such as the world of emotions or concepts, so long as that world lies outside the notes of the music. Roger Scruton’s statement quoted above is a good case in point. He argues that the notes of music themselves have semantic implications, that chords have ‘musical possibilities’, yet he chooses to locate these implications outside of the notes themselves – in the ‘idea’ of a chord and more generally in the gestures enacted by musical structures, which a cultivated listener can recognize. The long tradition of studying the relation between music and emotion, seen in the affect-centered parts of topic theory and in the famous work of Derycke Cooke (1959), takes this path, and has continued in much of the cognitive or behavioral approaches to musical meaning (summarized well, for example, in Sloboda and Juslin (2001), and also in Koelsch (2010) and Krumhansl (1997)). The theories of Leonard Meyer (1956, 1973) are particularly relevant in this regard, since he makes an explicit distinction between internalist and externalist meaning (calling them “embodied” and “designated” meaning, respectively). Even though much of Meyer’s work focuses on the embodied aspects of meaning, he eventually seeks to relate these to the emotional aspects of the musical experience, thus externalizing the source of musical meaning. Eric Clarke (2005) takes a similar behavioral approach, substituting “culture” with the more scientific term

“environment”, to see how musical meanings arise from our association with various environmental stimuli, particularly auditory ones.

Despite the diversity of theories and ideas advocated by the above authors, one thing is common to all of them – a belief in a realm of musical reality that transcends the mere structures of music of the kind musical grammarians talk about. Under this approach it is only when we start talking about this reality – the cultural, conceptual or perceptual realms of music – do we start talking about musical meaning.

The above illustrates the prevalence and popularity of the externalist approach to meaning in both language and music, much of the latter being in fact influenced by the former. Now externalist theories of meaning can also be considered *functionalist* theories because they do not see linguistic or musical signs as being ends in themselves. Rather, externalist theories consider signs as serving a certain function, namely the function of signifying a world external to musical and linguistic structure.

But in contrast to such externalist theories, one can imagine certain *internalist* theories of meaning too, in which linguistic or musical signs are ends in themselves – i.e. in this approach to the subject, meaning is inherent in the very structures of music and language, of the kind studied by theories of musical and linguistic grammar. Moreover, since meaning thus described lies within the internal *form* of musical and linguistic grammar, the internalist theory of meaning that results from such an approach is a *formal*, as opposed to a functional, theory of musical or linguistic meaning.

2.1.2. Logical Form and Internalism in Linguistic Meaning

As might be evident now, the grammatical approach to language found in Minimalist linguistics takes exactly such a formal, internalist attitude toward meaning in language. An example discussed by the linguist Massimo Piatelli-Palmarini (in the foreward to Uriagereka (1998): xxx) will help illustrate this:

(1a) They saw that he was trying to escape.

(1b) They saw him trying to escape.

(2a) They suspected that he was trying to escape.

(2b) *They suspected him trying to escape.

Both (1a) and (1b) are related in some way, just as (2a) and (2b) seem very different – in fact, (2b) is simply ungrammatical, which is why it is preceded by an asterisk. As Piatelli-Palmarini explains:

“The similarity between (1a) and (1b), on the one hand, and the dissimilarity between (2a) and (2b), on the other, are somehow determined by the meaning of the verbs *see* and *suspect*: witness the fact that *suspect* is naturally followed by a sentence (*suspect that ...*), whereas *see* is naturally followed by the name of an object (*see a cat*) ... The contrast between (1b) and (2b) teases apart these two intuitions. Yet it would be pointless to sit down and think very hard about the *concepts* of seeing and suspecting, or about the social “use” of these concepts in our everyday life, in order to derive an explanation for these simple facts. ... the difference between the two groups of verbs lies primarily in their ability to combine with, and select, other elements in a sentence. The relevant difference, in other words, is syntactic, not conceptual.”

In other words, we can talk about the meaning of language within the context of grammar – in fact we can even say that grammatical generation contains a semantic level of structure, where semantic issues, such as the above, are dealt with, in order to generate well-formed sentences like (1b) and to avoid ill-formed sentences like (2b). Most importantly, we can do this without making any reference to what the world is like, or to the various cultural or individual differences among people in the world – in fact, to find meaning in these sentences there does not have to be a world external to the sentence at all. And this constitutes an internalist alternative to the predominantly externalist approaches to meaning prevalent in language scholarship.

As noted at the beginning of the chapter, the advantage of such an approach from a musical perspective is that the grammatical similarities between music and language might allow such a grammar-internal approach toward meaning to relate musical and linguistic meaning in suggestive ways too. And has been a recurring theme throughout this dissertation, such an internalist theory of musical meaning seems to be inherent in Schenkerian music theory, so that a Minimalist interpretation of Schenkerian theory might reveal an identity between music and language even in the domain of meaning – in a manner suited to the identity theses for music and language this dissertation has been defending. So, let us now attempt to

understand the Minimalist approach to meaning in greater depth, in order to understand subsequently what this implies for an internalist theory of musical meaning in particular, and more generally for a theory of musical *and* linguistic meaning within a joint Minimalist approach to musical and linguistic grammar.

Let us begin by examining why the Minimalist approach to linguistic meaning takes an internalist attitude toward meaning, despite the fact that an externalist attitude is more common in language scholarship, albeit primarily in non-linguistic circles. That is, why do Minimalist linguists privilege grammatical approaches to linguistic meaning, when this prevents one from accounting for the use of language as a communicative activity (as Michael Dummett argued above)?

Part of the answer to this has to do with the centrality of grammar *per se* in Minimalist linguistics, and the emphasis on conceiving significant aspects of language in grammatical terms – which, as argued in Part I of the dissertation, is something Minimalists defend on a variety of philosophical, scientific, and methodological grounds. But the other part of the answer to this has to do with the problems inherent in developing any systematic theory of language in which communication has any role to play. That is, the *function* of language in communication, which is the focus of functionalist theories of language (and which includes most externalist approaches to linguistic meaning), is so prone to the vagaries of human social behavior that it is impossible to develop a systematic theory of these functions. This is why generative linguists include all of these vagaries in the *performance* aspect of human language, which, unlike *competence*, is not the focus of a generative theory of language. And this is why such a functionalist attitude towards *music*, as inherent in much ethnomusicological research, seemed a futile perspective to take in the quest for developing a systematic theory of music *and* language, or even just a cross-idiomatic theory of music taken by itself, as I discussed in the first part of chapter 1.1.

But the matter is not as simple as this. Language clearly seems to have properties that pertain to its meaning, and which might have to do with its social functions – properties that music does not seem to possess, as a result of which so many thinkers have felt the need to see language and music as being distinct. For example, there is a reason why Lerdahl and Jackendoff assert, as we saw in chapter 1.1, that,

“whatever music may “mean,” it is in no sense comparable to linguistic meaning; there are no musical phenomena comparable to sense and reference in language, or to such semantic judgments as synonymy, analyticity, and entailment” (Lerdahl and Jackendoff (1983): 5). This implies that it is not enough to say that musical and linguistic meaning can be understood as related simply because they seem to be grammatically-related, or because an extra-grammatical theory of music or language suffers from certain problems.

However, there is good reason to believe that the case for language having certain semantic properties that music does not have, as implicit in Lerdahl and Jackendoff’s above statement – making music and language distinct systems in the process – is overstated. Take for example the common belief, repeated above by Lerdahl and Jackendoff, that language possesses the property of reference. The strongest evidence for this belief comes from the existence of what philosophers call *proper names*. John Stuart Mill, in his *System of Logic*, arguably the classic treatment of proper names, says that:

“Proper names are not connotative: they denote the individuals who are called by them; but they do not indicate or imply any attributes as belonging to those individuals. When we name a child by the name Paul, or a dog by the name Caesar, these names are simply marks used to enable those individuals to be made subjects of discourse.” (Mill (1974): 33)

So, according to this view, proper names have no meaning of their own, and exist only to refer to “the individuals who are called by them”. Which makes the existence of proper names the clearest illustration of the referential aspect of language.

But many philosophers since Mill have attacked this view of proper names. For example, Bertrand Russell, developing the ideas of Gottlob Frege, proposed a *descriptive* theory of naming (most famously in his paper *On Denoting* (Russell (1905)) – according to which proper names are really just abbreviated definite descriptions, i.e. they have a meaning that is identical to certain other garden variety descriptions. An example is the proper name “Led Zeppelin”, which has a meaning identical to the description “the band that recorded *Houses of the Holy*”. Russell also distinguished between two different kinds of definite descriptions, viz. those that designate specific objects, such as the one in the preceding sentence, and those that do not, such as “the band that recorded albums for Atlantic Records in the

1970s”, which could designate the band Led Zeppelin, but also bands such as Yes, Genesis, and Bad Company. In this view, proper names are therefore abbreviated definite descriptions of the former type, since “Led Zeppelin” is an abbreviated definite description that presumably designates a specific object, i.e. a specific group of individuals that included Jimmy Page, Robert Plant, John Paul Jones, and John Bonham.

The idea that proper names are really abbreviated definite descriptions receives more support from the fact that they can be used descriptively in sentences like, “The Led Zeppelin that so heavily influenced the development of popular music in the seventies became quite a different band in the eighties”. This sentence also demonstrates that one can use a determiner in front of a proper name, just as one would in a regular, garden-variety definite description such as “The people that so heavily influenced the development of popular music in the seventies turned to different pursuits in the eighties”. This is especially true of certain languages (although not in English), where determiners are frequently used before proper names even when they are the subject of a sentence, e.g. in the German sentence “Der Ludwig ist ein Künstler”(i.e. “Ludwig is an artist” in English, where “The Ludwig is an artist” would be ungrammatical).

What this implies is that a proper name *can* be considered to have a meaning, i.e. the same meaning as any definite description that fits it, when we take a descriptive, Russell-ian approach towards proper names. This creates the possibility that a proper name will refer to any entity in the real world that fits its description. But this is not something inherent in the name itself – which means that the descriptive theory of naming also creates the possibility that names do *not* necessarily refer to anything at all, despite having meaning. Which then brings into question the very phenomenon of reference in language, since proper names are supposed to be the prime examples of this phenomenon. This is a conclusion that can be drawn, for example, from critiques of the descriptive theory of names proposed by philosophers like Peter Strawson and Saul Kripke. Strawson says for example that it is the use that a proper name is put to that allows it to refer to a specific object, and this should not be confused with its meaning, i.e. the definite description it is associated with (Strawson (2004): 5-7). So, it is only when we use the proper name “John

Paul Jones” to talk about a specific musician, do we mean (i.e. invoke the description) “the bassist for Led Zeppelin”. That is, the specific reference of “John Paul Jones” is not inherent in the name itself, but only arises from the use to which the name is put. If this name were to be used in another way, e.g. to refer to one of the leading protagonists in the American Revolution (i.e. the naval captain John Paul Jones), the description “the bassist for *Led Zeppelin*” would still be meaningful, but would not refer to this individual.

This means that rather than being a definite description that designates a specific object, a proper name really just presupposes the existence of such an object. Strawson argues in consequence that Russell seems to be confusing meaning for reference in the descriptive approach to naming, and that proper names by themselves do not have a reference:

“The source of Russell’s mistake was that he thought that referring or mentioning, if it occurred at all, must be meaning. ... Because Russell confused meaning with mentioning, he thought that if there were any expressions having a uniquely referring use, which were what they seemed (i.e. logical subjects) and not something else in disguise, their meaning must be the particular object which they were used to refer to. Hence the troublesome mythology of the logically proper name. But if some one asks me the meaning of the expression “this” – once Russell’s favourite candidate for this status – I do not hand him the object I have just used the expression to refer to, adding at the same time that the meaning of the word changes every time it is used. Nor do I hand him all the objects it ever has been, or might be, used to refer to. I explain and illustrate the conventions governing the use of the expression. This is giving the meaning of the expression. It is quite different from giving (in any sense of giving) the object to which it refers; for the expression itself does not refer to anything; though it can be used, on different occasions, to refer to innumerable things.” (Strawson (2004): 7)

Saul Kripke develops this thesis in a different way in his classic text *Naming and Necessity* (Kripke (1980)). Kripke argues that if the meaning of a proper name is its description, then people should know this description, or rather most or all of the descriptions that constitute the name’s meaning, when they use the name. However, people often do not know these descriptions, and sometimes associate incorrect descriptions with the names of things too. For example, the meaning of the name “Ferdinand Magellan” is identical for many people to the description “the first person to have circumnavigated the globe” – indeed, this may be the only meaning that “Ferdinand Magellan” has for many people. But this is actually false, since the first person to actually circumnavigate the globe was not Magellan but one of his captains, Juan Sebastián Elcano. According to the descriptive theory of names, this means that when most people use the

name “Ferdinand Magellan”, they are actually referring to a different individual than they think they are (viz. Elcano) – but Kripke argues that this seems to be a rather counterintuitive claim. What Kripke’s argument therefore suggests is that the descriptive theory unnecessarily mandates a connection between a speaker’s beliefs about an object and his/her name for that object. What this also implies is that such a necessary connection might not even exist. So, even if “Ferdinand Magellan” does mean “the first person to have circumnavigated the globe” in some hypothetical world, this is not *necessarily* true, as is the case in our own world. Kripke argues that having such a necessary connection is important though, to allow us to speak about things Ferdinand Magellan *might* have done, in a hypothetical world. Since the descriptive theory of naming is not able to countenance this, it just ends up problematizing the very notion that language is capable of reference.²

In light of the above, a better proposal might be to claim that if proper names are to genuinely illustrate reference in language, they should necessarily refer to specific objects, in all possible worlds – i.e. they should *rigidly designate* certain objects – but without this depending on the meaning of names, and especially people’s knowledge of these meanings. Such a proposal can be found in an alternative to the descriptive approach to naming and reference, known as the *causal* theory of reference, and which is largely based on suggestions made by Kripke. According to this approach, proper names are said to rigidly designate certain objects, not because of the meaning of the name, but because of an act – an “initial baptism” (Kripke (1980): 96) – that ‘fixes’ the referent of the name. Subsequent uses of the name to refer to that object are then said to be causally linked to the initial baptism (e.g. a person who does not know that “Led Zeppelin” refers to a band, might hear an interview with the band members, in which the origins of the band’s name is discussed, which then allows the person to use “Led Zeppelin” to successfully refer to the band in subsequent conversations with other people).

² The first of Kripke’s two arguments against the descriptive theory of names is also known as the “epistemic” argument, and the second the “modal” argument against the descriptive theory.

If one accepts the causal theory of reference, or Strawson and Kripke's criticisms of the descriptive theory of naming, it seems like the only way of defending the notion that language is referential, at least in the case of proper names – the prime example of this phenomenon – is by taking the discussion into the realm of language use. For Strawson, as we saw above, reference happens not because of any inherent property of language, but through the way language is used. Similarly, for Kripke, reference is not an intrinsic property of language, but something that occurs through the use of language in initial baptismal acts, and then subsequent causally-related acts. Which means that the phenomenon of reference can only be accounted for by an *externalist* view of language, since it is only in this view that the (grammar-external) uses of language, rather than its internal structural properties, become points of focus.

However, as I have suggested before, developing a systematic theory of language use is challenging, given the vagaries of this phenomenon. And the same can be said for how language is used in the phenomenon of reference. So, by externalizing the basis for reference in language, i.e. by making it a product of certain actions performed by individuals when using language, we only complicate the origins of reference in language, rather than providing a better account of it than the descriptive theory of naming does. To understand this in another way, one could argue that Russell “confused meaning with mentioning”, to use Strawson's words, for an important reason, viz. because he wanted to assert that talking meaningfully about something requires knowing what that specific thing is that one is talking about. (The philosopher Gareth Evans has referred to this as “Russell's Principle” (Evans (1982): 89).) Russell subsumed this knowledge under his two famous categories of knowing, i.e. “knowledge by acquaintance” and “knowledge by description” (Russell (1912): 72-92). That is, one can meaningfully talk about an object if one knows what specific object that is, either through direct acquaintance (e.g. if one can directly point to it), or if one knows a discriminating description of that object that distinguishes it from all others.

None of this has any connection to the causal theory of reference though, because that theory only deals with how one uses certain words, e.g. names, to refer to things – the causal theory of reference is not concerned with how one goes about talking about those things meaningfully. (To put it simply, it is a

theory of reference and not of thought.) This is why it does not require a person to know the meaning of a name before that person can use it to refer to an object – as long as there is some causal connection between that person’s use of the name and the initial baptism that rigidly designates it. But what this means is that the causal theory of reference is not even a response, therefore, to the descriptive theory of naming proposed by Russell – it is a different beast altogether. It is only concerned with the use of names in referring to things, irrespective of their use in thinking about things. And one could argue that this still leaves the whole issue of how names refer a bit of a mystery – the causal theory just says *that* someone can use a name to refer to something even without having any thoughts about that thing, without saying *how* or *why* this is possible. This just leaves the whole phenomenon of reference up to the vagaries of society, and specifically the social use of language, without really describing how this phenomenon occurs, i.e. without accounting for the range of possibilities and conditions under which causal connections transpire between names and things. Gareth Evans illustrates the problem thusly:

“Suppose there were in fact eight emperors of China called ‘Chan’, and there is a modern practice of using the name ‘Chan’ as if for an ancient emperor of China: the modern practice in fact derives from the use of the name for one in particular of the eight emperors, though no participants in the modern practice can distinguish that one from any of the seven others. Here, making it clear which modern practice one is participating in hardly seems to be enough to make it clear that one is speaking of that emperor in particular. If we want to retain anything of the idea that the reference of a name is fixed by uses of it by producers, it seems that *we must understand the requirement that one manifest which practice one intends to be participating in in terms of an individuation of practices which does not merely concern the present, but involves an indication of who the relevant producers were* [my emphasis]. Or, to put it another way: if we want to use the idea that the reference of a name on our lips is fixed by the fact that we have taken over (a version of) a name by which someone was known, then we must impose the requirement that we be able to give some indication as to who knew the person in question by (a version of) that name.” (Evans (1982): 405)

This implies that our options are to either accept the descriptive approach to names, with its arguably richer account of thought and meaning in language – but with its complicated and sometimes contradictory views on naming and reference – or the causal approach, which takes on the task of explaining reference by moving it into the sphere of social language use, but which, in the end, does not seem to account for reference with any more certainty or completeness than the descriptive approach

does. With the end result being that the notion that language is capable of reference still remains unaccounted for.

One last point worth regarding in our discussion of reference has to do with what are called “logically proper names”, e.g. demonstratives such as “this” or “that” (as in “give me *that*”, as opposed to “give me the book”). Russell distinguished logically proper names from ordinary proper names such as “Led Zeppelin” for reasons related to his distinction between knowledge by acquaintance and knowledge by description. That is, we know the meaning of logically proper names by direct acquaintance – we know what “that” means in “give me *that*” when for example, someone points to the object “that” refers to – whereas we know the meaning of an ordinary proper name by description, since ordinary proper names are, after all, disguised definite descriptions in Russell’s view; their meaning is identical to that of one or more garden variety definite descriptions.

Now, one could respond to the above way of distinguishing logically and ordinary proper names as based on a confusion between meaning and reference. In other words, one could argue that the meaning of a logically proper name does not have much to do with the object that it is used to refer to, whose knowledge we then acquire through direct acquaintance with that object, because the meaning of any name is really the set of conventions governing how the name is used to refer to objects, rather than those objects themselves. In fact, this was Peter Strawson’s very criticism of Russell’s descriptive approach to naming, which was cited a few pages ago. But if we accept this criticism, we must then proceed to understanding reference in terms of social language use – which has its own inherent problems, as we saw above.

So, let us ignore Strawson’s argument against Russell’s theory of names for the moment, and assume that a genuine distinction can be made between logically and ordinary proper names. This is much better for a theory of reference, since logically proper names fare much better than ordinary ones when it comes to their referential abilities. Whereas the referential abilities of ordinary proper names can be brought into question, as we saw in the above criticisms of the descriptive theory of names, logically proper names appear to be the purest, least controversial examples of reference in language. This is

because when we use a logically proper name to refer to an object, that object must necessarily exist (i.e. it must exist for the statement containing the name to make any sense – to say “give me *that*” and point to a non-existent object would make no sense), and our direct perception of that object makes the referential link between the name and object clear and incontrovertible.³

But there is a significant problem with the above description of rigidity in logically proper names. What it implies is that the rigidity of the logically proper name has essentially a *grammatical* basis. In other words, the reason why ordinary names do not seem to designate rigidly is because they need another grammatical entity to allow them to do this, viz. the demonstrative “this”, which logically proper names already have. So, what gives a description rigidity – assuming, per Russell, that ordinary names are disguised definite descriptions – is a grammatical entity that is separate from the description itself. In fact, we could argue that this is really the only difference between ordinary and logically proper names – the grammatical entity which gives either of these rigidity has to be added on to ordinary names, whereas it is *implicit* in a logically proper one. This implies that the phenomenon of reference in language, best exemplified by logically proper names, is largely grammatical – and beyond that it just something that has to do with the complicated social uses of language, which are not susceptible to easy description or explanation.

This also means that proper names are never inherently referential, but only become so as the result of a grammatical process. And this has a big advantage too – depriving names of their rigidity allows us to use different names for the same object depending on context and need, without rigidly tying them to specific objects in a one-to-one relation. For example, Henry Jekyll and Edward Hyde ostensibly refer to the same individual, but using two different names for this individual allowed Robert Louis Stevenson to explore the complex psychology of this person, which provided English literature with one

³ This allows logically proper names to genuinely rigidly designate objects too, as per Saul Kripke’s above criterion for reference. In this light, one could argue that the causal theory of reference was really an attempt to endow ordinary names with the same referential abilities as logically proper ones, by connecting them, causally, to demonstrative acts – i.e. the “initial baptism”, as exemplified by the act of pointing to someone and saying “I name *thee* Charles”, as happens at actual baptisms – and which is what seemingly gives logically proper names their rigidity.

of the great allegories about good and evil in the process. (See Uriagereka (1998): 110-113 for a further discussion of these matters, and also Neale (1990).)

The conclusion that I wish to draw attention to from the above discussion is that to the extent that one wishes to develop a systematic theory of reference in language, the best hope for this lies in a *study of grammar* – i.e. in a study of the conditions that govern how grammatical entities like demonstratives merge with certain other grammatical entities.

What this means is that a discussion of meaning in language can to a large extent be subsumed under a theory of linguistic grammar. And to the extent that linguistic and *musical* grammar are identical, this implies that Lerdahl and Jackendoff's assertion that musical meaning "is in no sense comparable to linguistic meaning" is essentially without teeth. That is, even if musical meaning and linguistic meaning are divergent in some respect, e.g. in the case of reference, the best hope for explaining even these divergent aspects of music and language is by subsuming them under a *shared* grammatical theory of music *and* language – which implies that musical and linguistic meaning are indeed comparable, as long as one understands them as aspects of the shared grammar of music and language.

Another example of this can be found in the case of the phenomenon of *synonymy* in language – e.g. in the fact that we understand the passive and active forms of a sentence to mean the same thing. In their statement cited earlier in this chapter, Lerdahl and Jackendoff take this phenomenon to be another example of the difference between musical and linguistic meaning, given the absence of this phenomenon in music. But we also saw in chapter 1.2 that the similarity between the passive and active forms of a sentence has to do with how they are *grammatically* connected, as the products of a transformational operation. (That is, we can understand the passive form of a sentence as arising from a noun phrase movement transformation, which transforms the same D-structure from which the S-structure of the active form of a sentence is generated into the S-structure of the passive form of the sentence.) And to the

extent that music reveals a transformational grammar too, we could argue that synonymy in language is just a product of the same transformational grammar that identifies music with language.⁴

Now, we have seen in the preceding discussions that grammar is just one part of the explanation for phenomena like reference, and possibly synonymy, in language. The social use of language seems to play a part too, as we saw in Strawson and Kripke's proposals about how reference manifests itself in language. This is illustrated by the following sentences:

(3a) They saw him trying to escape from jail.

(3b) They saw him trying to escape from prison.

(3c) They saw him trying to escape from school.

Grammatically, these sentences are identical. But it is quite obvious that (3a) and (3b) are synonymous, and they both also differ in meaning from (3c). So, we have to have a way of relating (3a) and (3b) on the one hand, and distinguishing them both from (3c) on the other. And if we want to claim that meaning is contained within grammar – or that the synonymy of (3a) and (3b) can be explained in grammatical terms, as suggested on the previous page – we will have to find some grammatical way of achieving this. But unlike sentences (1a-b) and (2a-b) discussed earlier in this chapter, where the meaning of some words seemed to be contained within a grammatical description of how they form those sentences, the above examples suggest that sometimes grammatical properties cannot help determine the meaning of certain words. The issue at stake here is the same, or at least related, meanings of “jail” and “prison”, and the different meaning of “school”, which one could argue has at least something to do with the different ways in which these words are *used* in language.

⁴ Phenomena like synonymy led certain linguists to conclude that D-structures must really be semantic structures, which is what allows S-structures generated from the same D-structure to have the same meaning. These linguists, such as Jerrold Katz, Paul Postal, John Ross and George Lakoff – who came to represent the “generative semantics” movement in linguistic theory – disagreed with the generative grammatical notion that the only truly semantic structure within grammar is Logical Form, which is a *surface* and not deep level of structure. This led to much debate between the generative grammar and generative semantics camps in the mid 1960s to 1970s about the status of meaning within generative linguistic theory, debates that are known collectively as the “Linguistic Wars”.

This point is illustrated even more dramatically in the context of a phenomenon that is normally considered part of language's semantic makeup, but whose apparent absence in music is taken by Lerdahl and Jackendoff to be yet another example of the difference between musical and linguistic meaning. This is the phenomenon of *analyticity*, in which the truth of a proposition is supposed to be evident from the meaning of its constituent terms. The stereotypical example of this is "All bachelors are unmarried males", whose truth is evident from the fact that the term "bachelor", by definition, is supposed to mean "unmarried male".

But given the complicated social uses of language, this is not always the case. As Ray Jackendoff argues, one can find instances where it seems odd to call someone a bachelor, just because that person happens to be a man who is unmarried – two cases in point being the Pope, who cannot marry on religious grounds, and an unmarried man who is in a long-term relationship with a woman, but neither of whom seem deserving of the epithet "bachelor". This leads Jackendoff to assert that the "the meaning of *bachelor* is inseparable from the understanding of a complicated social framework in which it is embedded" (Jackendoff (2002): 375). He also concludes that the linguist's job, therefore, is to take on the task of understanding this framework in order to account for meaning in language.

However, it is not clear that such a task is even feasible. One of the themes that has run throughout this dissertation is that the social uses of systems like language and music are such that they do not yield themselves to systematic explication, meaning that the attempt to understand these uses systematically will quite possibly be futile for developing unified theses about music and language, even when they are considered individually, let alone jointly. At least this was the conclusion chapter 1.1 arrives at in its discussion of John Blacking and Harold Powers' proposals about music and language.

This seems to be how generative linguistic theory understands this issue too, at least within Minimalist approaches to these matters. For example, think about the complications surrounding whether to call a tall geological entity a "hill" or a "mountain",⁵ or what the word "field" means. Focusing just on

⁵ Which is an issue that underscores the humorous title of a movie starring Hugh Grant, "The Englishman Who Went Up a Hill But Came Down a Mountain".

geological or geographical matters again, and leaving aside its meaning in the context of specific *fields* of study, we might define “field” as an open area of land without trees or buildings, which is often used for a specific purpose (as in “sports field” or “battle field”), and/or is marked by the presence of particular objects or features (as in “oil field” or “mine field”). In fact, this is exactly how the Merriam-Webster dictionary defines the term. And even if one thinks of a field as an area of land covered in *grass*, which is how people often conceive of fields, the above meaning of the term still persists even if such an area of land is covered, for example, in snow (as in a snowfield on a mountain), or sand (as in a dune field in a desert) – or even a relatively abstract substance like gravity (as in a gravitational or other force field, as discussed in physics). But what if such an area of land is covered in water – would we still consider this a field, or would we perhaps consider it a lake? Moreover, if the particular objects marking such an open area of land are graves and gravestones – would we still consider this a field, or would we consider it a cemetery or graveyard? When does a *battlefield* become a *graveyard*?

In other words, even when an object is physically identical (i.e. in spatio-temporal terms) to what it was before, its meaning often does not remain the same. All of which suggests that the social uses of language are complicated and quite possibly intangible, which is what makes it so hard to develop a systematic theory of meaning in language. What this implies is that the aspects of language that depend on its social use, especially those that pertain to its meaning, should be excluded from the attempt to develop such a systematic theory. As Noam Chomsky says:

“The concept “human being,” with its curious properties of psychic continuity, does not enter the natural sciences. The theory of evolution and other parts of biology do try to understand John Smith and his place in nature; not, however, under the description “human being” or “person” as construed in ordinary language and thought. These notions are interesting for natural-language semantics and ethnoscience, but not for the branches of human biology that seek to understand the nature of John Smith and his conspecifics or what distinguishes them from apes and plants. ... The special sciences also go their own ways. To borrow Jerry Fodor’s example of a meandering river eroding its banks, the earth sciences do not care under what circumstances people take it to be the same river if the flow is reversed or it is redirected on a different course, or when they regard something projecting from the sea as an island or a mountain with a watery base. The same should be expected in the case of such notions as *language* and *belief*, and terms of related semantic fields in various languages and cultural settings.” (Chomsky (2000): 139)

The Minimalist conclusion, therefore, to the issue of meaning in language is that pursuing a theory of meaning beyond that which can be subsumed under grammatical theory is likely a futile and unnecessary endeavor. In other words, and contra Michael Dummett's assertions discussed at the beginning of this chapter, a theory of grammar *does* have much to say about meaning, but *not* when meaning is understood in terms of the grammar-external social use of language in communicative acts – because talking about meaning under this conception of language is ultimately a futile enterprise.

In other words, the Minimalist conception of meaning in language is one that does not involve a “semantics”, in the conventional grammar-external sense of the term. To quote Chomsky again:

“As for semantics, insofar as we understand language use, the argument for a reference-based semantics seems to me weak. It is possible that natural language has only syntax and pragmatics; it has a “semantics” only in the sense of “the study of how this instrument, whose formal structure and potentialities of expression are the subject of syntactic investigation, is actually put to use in a speech community,” to quote from the earliest formulation in generative grammar 40 years ago, influenced by Wittgenstein, Austin and others. In this view, natural language consists of internalist computations and performance systems that access them along with much other information and belief, carrying out their instructions in particular ways to enable us to talk and communicate among other things. There will be no provision for what Scott Soames calls “the central semantic fact about language,... that it is used to represent the world,” because it is not assumed that language is used to represent the world, in the intended sense.” (Chomsky (2000): 132)

This, then, is the ultimate internalist approach to meaning – i.e. one that seeks to understand meaning only as an internal component of grammar, in light of the conclusion that grammar-external aspects of meaning are either inexplicable or overblown in their relevance to a systematic theory of language. And we have already discussed the advantages of such an internalist approach to meaning for a theory of music. By declaring that phenomena like reference, synonymy, analyticity – those very aspects of linguistic meaning that have been taken as the basis for the difference between music and language – are comprehensible only in terms of grammar (that too a grammar that is *ex hypothesi* identical in music and language), the internalist approach renders the supposed ‘semantic’ differences between music and language null and void. And in the process, it provides a suggestive alternative to the various externalist approaches to meaning that dominate both music and linguistic scholarship.

This position has led some to assert that there really is no theory of meaning in Minimalist linguistics. For example, the linguist Jerrold Katz, who was affiliated with the generative grammar tradition earlier in his career, but broke with it later, says that:

“Given the philosophical sources of skepticism about meaning, one would expect C[homsky] – a strong opponent of empiricism and behaviorism in linguistics, philosophy, and psychology ... – to support meaning against its critics. But he has not done so. To the contrary, he has never taken anything stronger than an agnostic position on the prospects for a theory of meaning; from the beginning of his career to the present, his position has been a skepticism comparable to that of many of the empiricists or behaviorists.” (Katz (1980): 2-3)

But there *is* a theory of meaning inherent in Minimalist linguistics, except that it is a rather unusual, grammar-internal, one. In pre-Minimalist approaches to language within generative linguistics, this involved, for example, the Theta Criterion and Binding Theory, as constraints on the generation of D-structures, which we explored in section 1.2.3 of chapter 1.2. In Minimalism’s own approach it involves the notions of “Logical Form” and “Full Interpretation”, which we have also encountered before.

The notion of Logical Form or LF has antecedents in philosophical discussions about meaning – including, in fact, some of Bertrand Russell’s proposals about meaning that we explored earlier in this chapter. Russell developed his descriptive approach toward meaning and reference partly to deal with a logical problem that arises when we examine certain kinds of sentences, such as the famous one discussed by Russell himself, viz., “the present King of France is bald”. If this sentence expresses a true proposition, its negation “the present King of France is not bald” must express a false proposition, because the present King of France must either be bald or not bald – he cannot be both, according to the Law of the Excluded Middle, a principle of logic that Russell accepted, and whose origins can be traced back to at least Aristotelian logic. But there is no way of deciding this matter if there is no present King of France. That is, we do not know which of these sentences expresses a true proposition if there is no King of France who can directly verify or falsify these propositions by means of our acquaintance with him. (This reveals, again, Russell’s belief that the purest form of knowledge and reference arises through direct acquaintance with things.)

If the truth or falsehood of “the present King of France is bald” cannot be decided, does this imply that the sentence is meaningless? This does not seem to be the case since as a combination of words that individually make sense the sentence as a whole seems to be meaningful too. Therefore, as a solution to this conundrum, Russell suggested that the grammatical form of a sentence like “the present King of France is bald” obscures its *logical form* (not to be confused with the separate, albeit related, Logical Form of generative linguistic theory). The logical form of a sentence is its inherent *quantificational* structure, which can be expressed by means of a formal language of logical quantifiers, and which reveals that even sentences like “the present King of France is bald” have a (logical) meaning. Put another way, a sentence like “the present King of France is bald” can be analyzed logically into: (1) a proposition about the existence of its logical subject, viz. the present King of France, using the existential quantifier \exists , (2) a proposition about this subject’s uniqueness, using the universal quantifier \forall , and (3) a claim about this subject’s attributes, like its baldness. This gives us:

$$(4a) \exists x[K(x)]^6$$

$$(4b) \forall x \forall y [[K(x) \& K(y)] \rightarrow y = x]^7$$

$$(4c) \forall x [K(x) \rightarrow B(x)]^8$$

In simpler language this means that the logical form of the sentence, underlying its grammatical form, can be understood as saying that there is one and only one x , such that x is the present King of France, and x is also bald. This shows that the sentence has meaning, by virtue of the attributes ascribed to its logical subject by its logical predicate. This is separate from the issue of whether the sentence expresses true or false propositions, which depends on the existence of a unique thing such as the present King of France. If

⁶ This reads “there is an x , such that x is K ” (i.e. K being the present King of France).

⁷ This reads “for all x and all y , if x is K and y is K , then x and y are the same”, meaning that there is only one such thing that is K

⁸ This reads “for all x , if x is K , then x is also B ” (i.e. B symbolizing “bald”).

such a thing uniquely exists – i.e. if propositions (4a) and (4b) have a *reference* – then (4a) and (4b) will be true, which implies that the sentence “the present King of France is bald” expresses a true proposition. Contrariwise, if such a thing does not exist or is not unique, (4a) and (4b) will be false, which implies that the sentence expresses a false proposition.

Two important conclusions can be drawn from this description of logical form. First, it suggests that the meaning of a sentence depends on how its individual parts are related – i.e. it depends at least on the relation between its logical subject and predicate. Secondly, it suggests that the meaning of the sentence does not depend on whether the proposition it expresses is true or false – *that* depends on whether the proposition has a reference or not.

These conclusions matter because it shows that meaning and reference can be separated.⁹ Given the problems inherent in the phenomenon of reference in language, the idea that sentences have a logical form allows us to think about their meaning irrespective of their referential qualities, and also independently of any notions of truth or falsehood. In other words, the idea of logical form allows us to think about linguistic meaning as something internal to the structure of sentences, which arises from the way a sentence’s constituents are combined to generate the sentence, and which has no connection to objects external to the sentence’s structure or to issues of truth and falsehood. (This is particularly important when comparing linguistic meaning with musical meaning, because the latter does not have any connection to issues of truth or falsehood either – musical sentences do not express propositions – and musical meaning might be thought of as something internal to the structure of musical sentences too, which is the internalist approach to meaning that I will discuss momentarily.)

⁹ Although as we saw earlier, Peter Strawson objected that Russell’s theory presupposes that a sentence’s logical subject always has a reference, because of the propositions about the subject’s existence and uniqueness inherent in its logical structure, implying that a sentence will always express a true or false proposition. This has the advantage of showing how sentences like “the present King of France is bald” do not violate the Law of the Excluded Middle, which is something Russell was interested in demonstrating. But Strawson suggested that whether a sentence’s subject has a reference or not depends on how the sentence is used, implying that a sentence can actually fail to express a proposition, and therefore be neither true nor false, if it is used incorrectly. In other words, if “the present King of France is bald” is used incorrectly, e.g. during a time when France does not have a king, it fails to express a proposition – and is therefore neither true nor false, and also violates the Law of the Excluded Middle.

For these reasons, the notion of logical form is ideally suited for an internalist view of meaning – and it is this that led generative linguists to develop *their* notion of Logical Form or LF, as a basis for a grammar-internal theory of meaning in language (Hornstein (1995): 5, Chomsky (1991): 38). The generative (and ultimately Minimalist) description of LF is similar to the more traditional philosophical notion of logical form as discussed above, but it differs from the latter in that it dispenses with the grammar-external issues of reference and truth that Russell associated with it. For this reason, in linguistic theory, Logical Form is a grammatical structure – an S-structure generated by Merge – but one that also reveals the meaning of that structure, which is often obscured by the actual, articulated (i.e. Phonetic) Form of that S-structure, as we saw happens in structures like “the present King of France is bald”.

So, during the process of generating phrases, the computational system takes words as inputs that, as we have seen before, are marked by features – that determine whether a word is a noun or a verb, masculine or feminine, singular or plural etc. – and the computational system uses these features to generate phrases. Crucially, some of these features pertain to meaning as well, which means that when words are merged they ultimately give rise to an S-structure that is meaningful too. When all information not pertaining to meaning is abstracted away from the S-structure, this resulting ‘surface’ level of structure – or level of representation – is a purely semantic one, but one that is nonetheless internal to grammar, as an S-structure generated by grammar. And so this is what LF is, i.e. a *semantic level of representation*. Therefore, in the generation of an LF representation, meaning is coupled with grammar in the generative process – i.e. the *substance* of language is coupled with the internal, organic *form* of language, as was discussed in chapter 1.1, and which reveals the Aristotelian way in which the Minimalist Program explains language.

As we will see in the next chapter, the same phenomenon occurs in the case of an S-structure’s Phonetic Form or PF level of representation, with regard to information about how that structure should be articulated. So, LF and PF are representations of the same (grammatical) S-structure when considered in its different, semantic or phonological, aspects. This also suggests that there is a deeper connection between LF and PF – which is how generative theory understands the age-old idea that language is a

system of *sound-meaning pairings* – an idea that goes back to at least the celebrated work of the Swiss linguist Ferdinand de Saussure. So, from a Minimalist standpoint, we can argue, following tradition, that sound and meaning are paired in language, because of the fact that LF and PF are representations of the same S-structure – but this only occurs in a grammar-internal way, as a result of the workings of Merge in generating such an S-structure. This, then, is the unique internalist perspective that generative theory, and particularly the Minimalist Program, provides to the matter of language being a system of sound-meaning pairings.¹⁰

So, the postulation of an LF level of structure illustrates how we understand what a sentence *means* once it has been generated. In other words, we can understand what the sentence means because we have formed, *ex hypothesi*, an LF representation of the sentence in our minds. But what does it mean to “understand the meaning” of a sentence, as given by its LF representation?

The answer to this brings us to the second crucial concept in the Minimalist approach towards linguistic meaning, which is that of “interpretation”, and specifically, “Full Interpretation”. In order to understand this concept, remember that according to Minimalism the internal properties of the computational system of grammar exist precisely in order to meet conditions imposed on it by the external conceptual-intentional (CI) and sensorimotor (SM) systems. The CI system is the system that deals with our emotions, beliefs, desires etc. – i.e. the substance of thought. The SM system basically acts as our interface with the outside world, i.e. it is the system by which we sense information from and articulate information to the outside world. And importantly, as discussed in chapter 1.1, the SM and CI systems are shared between different cognitive systems – so, all cognitive systems that receive

¹⁰ It is also worth remembering that generative theory only says *that* LF and PF are paired. It is silent about *how* they are paired. This is because this pairing is arbitrary – language is not just a system of sound-meaning pairings, but *arbitrary* sound-meaning pairings – an idea that can also be traced back to Saussure. Which is why the sound [cat] means a certain kind of animal only arbitrarily – it does not have to mean what it means to us, as English speakers, and can mean something quite different in another language. What linguists like Saussure often understood to be *non-arbitrary* though, is the way in which certain sounds (like those associated with logically proper names like *that*) relate to objects in the world. In other words, the phenomenon of reference is often taken to exemplify the non-arbitrariness of language. Minimalism does not accept this position given its skepticism about the relevance of reference in a generative theory of language.

information from the outside world, like language, music, vision etc., share the SM system – and these systems are not even unique to humans in the way our computational system seems to be.

So, the computational system, though separate from the grammar-external CI and SM systems of the mind, has a very intricate connection with these systems. If grammar is the form through which we express the meaningful content of language, to put it in Kantian terms, then without that content form would be empty. To the extent that we use language (and *ex hypothesi* music) to communicate, these external systems therefore give the computational system something *to* communicate – or more accurately, they act as constraints on how the computational system is used in specific contexts, without being involved in the actual generation of linguistic structures themselves – the latter being the sole task of Merge.

And LF and PF serve as interfaces between the computational system and these external systems. In other words, the computational system can map to these external systems successfully, only when the LF and PF representations of an S-structure can be *interpreted* by these external systems – which then allows the S-structure to be articulated by the SM system, and understood by the CI system. This is all there is to the notion of Full Interpretation. That is, if the LF representation of an S-structure can receive a *semantic interpretation* (which allows for comprehension by the CI system) and if the PF representation of that S-structure can receive a *phonetic interpretation* (which allows for articulation by the SM system), they are said to receive Full Interpretation – i.e. the LF and PF representations *converge* (which is just another way of saying that meaning and sound are paired, albeit arbitrarily as discussed above), and the generation of the S-structure by Merge is successful. If such Full Interpretation does not occur, i.e. if an S-structure cannot be phonetically and semantically interpreted, a sound-meaning pairing does not occur, the SM and CI systems will have nothing to articulate or understand – and the generation of the S-structure will be unsuccessful. In this manner, and as discussed in chapter 1.2, Full Interpretation acts as a general constraint on the workings of Merge – if Merge’s attempt to generate an S-structure cannot satisfy Full Interpretation, i.e. if it does not result in convergent LF and PF representations, the generative process will crash.

In chapter 1.2 we also looked at some constraints placed by the external SM and CI systems on S-structure generation by Merge, which can be understood as specific manifestations of Full Interpretation. One such constraint was the Case Filter, which requires that all noun phrases be adequately Case-marked for the generation of an S-structure to be successful. The Case Filter is a constraint associated with the SM system, since noun phrases that are inadequately Case marked result in PF representations of S-structures that cannot be *phonetically* interpreted, which prevents the SM system from articulating them. So, an S-structure that contains an inadequately Case-marked noun phrase will fail Full Interpretation, meaning that its LF and PF representations will not converge (mainly due to its ill-formed PF representation) – which means further that the generation of the S-structure will ultimately be unsuccessful.

The above discussion shows us why the concept of interpretation is so crucial for generative theory, and specifically the Minimalist Program. The phenomenon of interpretation shows us how the Merge-based architecture of grammar meets the external conditions imposed on it by extra-grammatical systems, which is all that is necessary for Minimalism to show in order for it to sustain its particular account of language, and which it does by postulating LF and PF levels of representation of S-structures generated by Merge.

Importantly though, the relevance of interpretation to Minimalism’s picture of language does not mean that a successfully generated S-structure *will* necessarily be interpreted. All that Minimalism says is that a successfully generated S-structure *can* receive an interpretation. Whether the structure actually *is* interpreted, i.e. whether the SM system actually articulates it, or the CI system understands it, depends on the extra-grammatical vagaries of how these systems work – which fall under the broad heading of “performance” (as opposed to “competence”) factors in language use. So, even if an S-structure is successfully generated, and therefore *pronounceable*, the SM system might not be able to articulate it – possibly due to an injury to the vocal cords, or the speaker’s having suffered a stroke, or a myriad of other practical and pragmatic problems. A similar situation obtains in the case of a sentence like “the present King of France is bald”. This sentence is clearly grammatical, and also meaningful as we saw earlier.

Which implies that it *can* receive an interpretation. However, it quite possibly *will not* receive an interpretation if one were to exclaim, “the present King of France is bald” today – and one might be met with incredulous stares instead, from people who cannot *interpret* what is being said (despite the sentence being perfectly meaningful), perhaps because “present King of France” seems to lack a reference (or at least has since 1870, when the Second French Empire collapsed). Alternatively, on hearing a sentence like this, one might try to interpret it figuratively, taking “present King of France” to refer to some contemporary Frenchman, like Nicolas Sarkozy, who might seem to act like a king without actually being one.

The point is that the picture of language being painted here is a grammar-internal one, in which the job of the grammatical system is to generate structures that are *interpretable*, irrespective of whether they are *interpreted*. This is what an internalist view of meaning amounts to – a structure is said to be meaningful in this view if it *can* be interpreted, irrespective of whether it *is* interpreted, i.e. irrespective of whether it is taken to be meaningful in the more common sense of the term, because that more common sense of “meaning” requires a study of the various external, performance factors that govern the social use of language – which are beyond the scope of a grammatical theory, and perhaps any systematic theory of language, as argued before.

This point can be illustrated even more clearly with another famous example, viz. the sentence “Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo” (Pinker (1994): 208). This sentence seems downright nonsensical, and is certainly tremendously hard to interpret. (Plus, if heard in ordinary (i.e. oral) conversation it would not reveal the different uppercase and lowercase Bs that might help one in interpreting it, as is apparent in the written form of the sentence shown here.) But the sentence *can* receive an interpretation, i.e. it is *interpretable*, because it is the result of a successful grammatical generation. In other words, the sentence has a legitimate grammatical, and specifically LF, structure. This can be seen in the fact that the word “buffalo” can participate in grammatical generation in three different ways, viz. as:

- (a) A noun – which means a certain kind of animal, also known as a bison.
- (b) A verb – which means “to bully”, or “to intimidate”.

(c) An adjective – which means that something belongs to the city of Buffalo in the U.S. state of New York.¹¹

In light of these three uses of “Buffalo”, we are able to understand the hierarchical grammatical structure of the sentence, represented thus (which has the grammatical tree structure shown in Example 2.1-1):

Buffalo^{adjective} buffalo^{noun} Buffalo^{adjective} buffalo^{noun} buffalo^{verb} buffalo^{verb} Buffalo^{adjective} buffalo^{noun}

Which then allows us to interpret it thus (this is Steven Pinker’s reading):

(The) Buffalo buffalo (that) Buffalo buffalo (often) buffalo (in turn) buffalo (other) Buffalo buffalo.

(A more colloquial translation of this could be “Bison from the city of Buffalo, New York, who are intimidated by other bison from that city, also happen to intimidate other bison from that city themselves”.)

A striking feature of this construction is how the recursive use of just one word (or one trisyllabic speech sound) can generate such a complex structure. But it is precisely because of its complex, nested structure that it becomes so hard to interpret. This is typical of nested structures – consider the following grammatical (and therefore meaningful in the internalist sense) sentences that are extremely hard to interpret without a knowledge of their grammatical structure:

The cat the dog the cow the milkman milked kicked chased left. (Uriagereka (1998): 65)

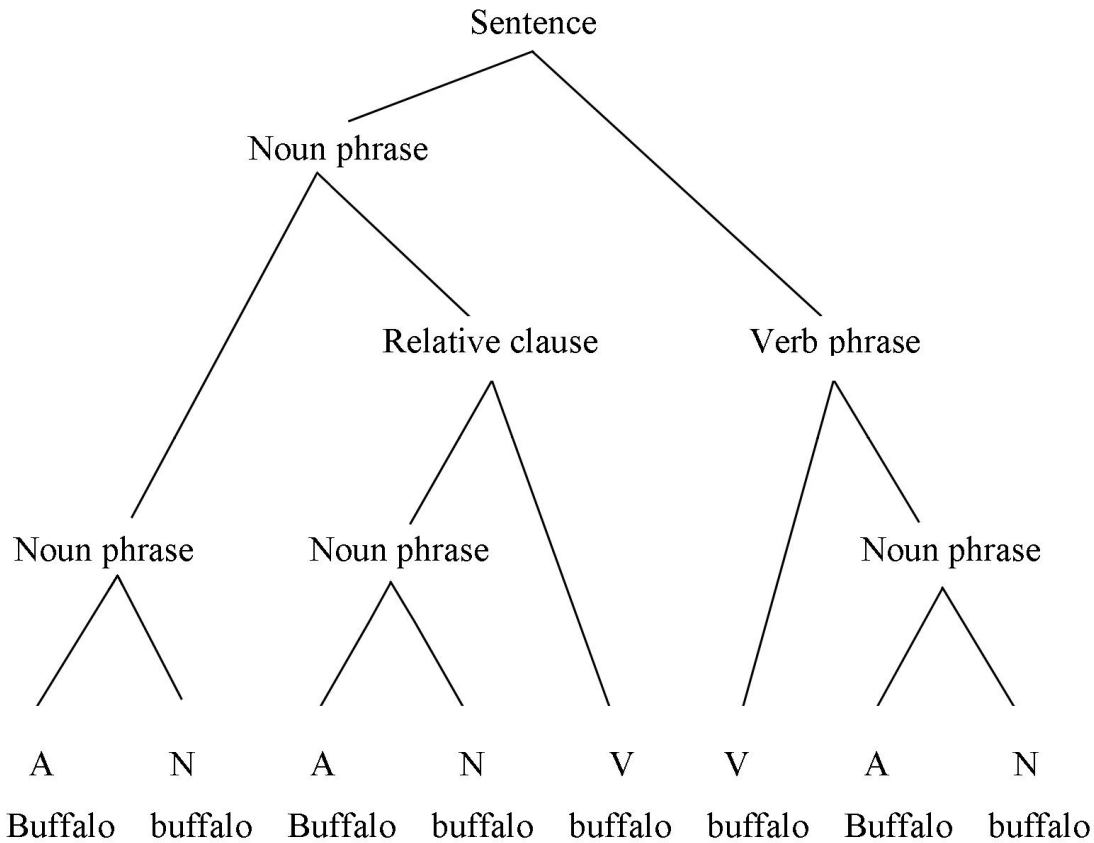
The umbrella the man sold despite his wife is in the room. (Bever and Langendoen (1971): 437)

The man who fishes goes into work seven days a week, but the man who hunts ducks out on weekends.

(Pinker (1994): 211)

¹¹ In X-bar theory, this would be an adjunct to the noun “Buffalo” (i.e. as described in (a)) that heads the noun phrase “Buffalo buffalo”. This is how (c) is often described in the literature on the above sentence.

Example 2.1-1. Tree for “Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo”



The point here is that many sentences appear to be meaningless (in the externalist, use, sense of the term) because they are hard to interpret, for a variety of performance factors – such as a stress on the information-processing abilities of working memory, of the kind imposed, for example, by heavily nested structures. But such sentences can still be meaningful (in the internalist, grammatical, sense of the term) if they *can* receive an interpretation, due to their grammaticality, i.e. due to their successful generation by Merge.¹²

¹² A final illustration of this phenomenon can be seen in the case of the sentence that has come to epitomize Chomskyan linguistics, viz. “colorless green ideas sleep furiously” (Chomsky (1957): 15). This sentence appears to be meaningless too, at least in the externalist, use sense of the term – e.g. ideas do not sleep and are not colorless, and something cannot be green and colorless at the same time. Which implies that such a sentence would be impossible to use in discourse, at least literally, because it defies all our pragmatic beliefs about how the world works. However, the sentence is grammatical, which means that it had to satisfy certain constraints on structure generation, some of which are *semantic* in nature. For example, it had to satisfy the Theta Criterion, according to

In this light, it does not make much sense to think of grammar as being something that divides the set of possible sentences into those that are correct and those that are incorrect – rather it divides them into those that are interpretable and those that are not.¹³

2.1.3. Hanslick and Internalism in Musical Meaning

Let us now turn to an examination of what the preceding discussion implies for a theory of *musical* meaning. That is, in light of the above Minimalist ideas about linguistic meaning, can we conceive of a similar, internalist theory of musical meaning, especially one in which musical analogues of LF and interpretation play a part? And does such a theory strengthen the thesis for the identity of music and language?

The idea that musical meaning *can* be approached from an internalist perspective certainly has a precedent in music-theoretic thought, most notably in the ideas of the 19th century German theorist Eduard Hanslick. Hanslick is well known for his defense of the position that music is meaningful solely

which the theta roles assigned by a sentence's predicate must map one-to-one with its arguments (review the discussion on this in chapter 1.2 in this regard). Since the predicate "sleep", being an intransitive verb, assigns just one, external, theta role to its subject, viz. that of *experiencer*, and since "colorless green ideas..." contains just one noun, that too in the subject position, viz. "ideas", the sentence satisfies the Theta Criterion. Moreover the noun phrase "colorless green ideas" has no antecedent NP in the sentence, and therefore remains unbound – which satisfies the constraint placed on non-pronominal noun phrases by Binding Theory. But the Theta Criterion and Binding Theory are both semantic constraints – which can be seen especially in the Theta Criterion's characterization in terms of predicates and arguments, i.e. terms that are associated with the logical, and therefore semantic, aspects of a sentence. This implies that in satisfying these constraints, "colorless green ideas..." *can* receive a semantic interpretation, i.e. it seems to have a legitimate LF representation. Which means that it is meaningful – but only in the grammar-internal sense described by generative theory. As a result of this, we *can* try to find all kinds of interpretations for it, even though the sentence appears to be meaningless. This can lead to figurative or poetic interpretations of the sentence, as was the case with "the present King of France is bald". (In fact, the Wikipedia article on this sentence mentions a competition in 1985 at Stanford University, in which competitors were asked to provide interpretations of the sentence within certain boundaries – see http://en.wikipedia.org/wiki/Colorless_green_ideas_sleep_furiously.)

¹³ This is relevant for an internalist description of music too, since one often hears the objection that musical sentences cannot be divided into those that are correct and those that are not, since musical structures apparently do not show the kind of grammaticality that linguistic structures do – implying that they should be understood in non-grammatical terms, e.g. in terms of something like Lerdahl and Jackendoff's idea of "preference". (Witness Lerdahl and Jackendoff's statement in this regard, "In music, on the other hand, grammaticality per se plays a far less important role, since almost any passage of music is potentially vastly ambiguous ... The interesting musical issues usually concern what is the most coherent or "preferred" way to hear a passage" (Lerdahl and Jackendoff (1983): 9).) But not only *can* we think about musical sentences in terms of grammaticality, as this dissertation has attempted to show, the above objection misses the real point about grammaticality in generative theory, which is that it is about what makes sentences interpretable, and not about what makes them correct or unambiguous.

by virtue of its internal structure, and for his criticism of the attempt to understand musical meaning in terms of extra-musical associations. This can also be seen in his position in the “Brahms versus Wagner” debate, where, as a friend of Johannes Brahms, and a partisan of the Brahms faction in the debate, Hanslick was vocal in his dislike for Wagner’s focus on word painting and extra-musical dramatics in his compositions, and equally vocal in his admiration for Brahms’ mastery of a more traditional and formal approach towards musical expression and innovation.

Hanslick’s main thesis was that musical meaning does not lie in the various functions that musical signs serve but rather lies within the various structures of music itself, within the various ‘plays of forms’ between these structures. Since plays of forms reveal the imagination of the composer, they thus reveal the meaningful, expressive content of the music. Unfortunately though, Hanslick himself took his internalist position on musical meaning to imply a separation between music and language, since he was not aware of the internalist nature of linguistic meaning:

“The essential difference [between music and language] is that in speech the sound is only a sign, that is, a means to an end, which is entirely distinct from the means, while in music the sound is an object, i.e., it appears to us as an end in itself.” (Hanslick (1986): 42)

From this Hanslick concludes that music is *autonomous* from language, and that musical meanings are not verbally recoverable. I will have reason to return to this view of musical meaning in a moment, especially its infelicitous separation of music and language. But first of all, it is worth noting that an internalist approach to musical meaning, Hanslickian or otherwise, has often met with resistance from music scholars – leading to the plethora of externalist approaches to this subject mentioned earlier in this chapter. I would like to show how these anti-internalist views on musical meaning can be rejected in favor of an internalist one, often for the same reasons discussed in the previous section, regarding the relative merits and demerits of an externalist approach to *linguistic* meaning. Which means that ultimately a Hanslickian perspective on musical meaning is worth defending – although it is in need of some revision, in order for us to see the connections between internal meaning in music *and* language, of the kind Hanslick failed to see himself.

The emphasis on finding meaning internally, within musical structure, implies a careful study of how such structures arise. Specifically, it involves a consideration of the computational system and grammatical processes from which musical S-structures are generated. This ultimately involves considerable technical facility with those aspects of music involved in structure generation, such as harmony, counterpoint, rhythm, etc. Many opponents of the internalist approach to music find this ‘technocentricity’ unappealing, especially if it involves a reactionary attitude towards the novel harmonic, contrapuntal or rhythmic features of new musical styles – as was partly the case with Hanslick (in his rejection of Wagner), and has often been a criticism levelled against Heinrich Schenker, e.g. regarding his attitude towards musical modernism, and his general rejection of any music of ‘non-Germanic’ origin.

I discussed in chapter 1.1 how the polemics of music theorists like Schenker can be ignored in favor of their considerable insights into musical structure, and for this reason I find the above criticism against musical internalism to be lacking in merit. But the rejection of the technical attitude towards musical structure inherent in musical internalism is often really the result of an interest in understanding music in its wider social, historical or political contexts – which is an interest that can be seen clearly in the more recent postmodern and critical attitude towards music that has come to be associated with the New Musicology (see Kerman (1980) for the classic statement of this interest). This interest has actually come to mark a rather patronizing attitude toward the technical aspects of music on the part of various music critics and New Musicologists too, e.g. see the Preface in Abbate and Parker (2012).

But accounting for the grammar-external social uses and functions of systems like music and language have their own inherent problems, so a rejection of musical internalism from this perspective seems rather unpersuasive. This point has been made in various forms earlier in this dissertation, but it is worth revisiting one last time in order for us to explore a problematic externalist theory of musical meaning that has received some attention in the music-theoretic community in recent years. This is the theory of musical *topoi*, associated with the work of Leonard Ratner and his students and colleagues. As mentioned earlier in this chapter, this theory attempts to reveal how certain musical surfaces signify certain grammar-external realities by means of various expressive gestures or “topics” present in them,

such as those that signify certain dance forms. It can be argued, however, that Ratner never conceived of these topics as any more than instruments for analyzing certain musical pieces, particularly in the Classical style, in order to show how eighteenth century listeners might have related to these pieces – meaning that for Ratner topics were quite possibly just a means to help teach a certain kind of music appreciation to modern audiences. Certainly his classic text on topics, i.e. Ratner (1985), never develops a *theory* of musical meaning based on the musical topics he discusses. But many later theorists who have researched musical topics, such as Robert Hatten and Raymond Monelle, have attempted to formulate explicit theories of musical meaning from them. These attempts are notable for seeking the meaning of certain musical structures, but beyond that which is contained within grammatical theories of these structures, especially since finding specific grammatical foundations for different topics has not proven to be a successful enterprise, as the music theorist William Caplin has pointed out (in Caplin (2005)). So, work in topic theory has focused on grammar-external ways of defining topics, in order to develop an externalist theory of topical musical meaning, and which has often involved pointing to ‘field exemplars’ of various topics found in different musical pieces.

But this endeavor has all the problems associated with the attempt to define the meaning of concepts extra-grammatically, as we saw in the case of linguistic examples like “bachelor”, and “field”. In other words, how does one define a term on the basis of its exemplars, when those exemplars vary considerably in their different social uses? This is demonstrated particularly well when one attempts to define topics that are supposed to signify certain dance forms. For example, consider the Polish dance form known as the “mazurka”, a dance form we will examine in greater detail in chapter 2.3, in the context of the great Polish composer Frédéric Chopin’s use of this form in many of his compositions. How does one define a “mazurka”? As just suggested, a possibility is to look at various examples of dances called “mazurka” in different social traditions, and then find “family resemblances” between them, as a Wittgensteinian philosopher might say. In consequence, one might define “mazurka” in terms of a certain short-long or short-long-long rhythmic pattern that seems to occur in several field examples of this dance. (Although this characterization is itself not without problems, as we shall see in chapter 2.3.)

The problem arises though when one considers passages in certain musical pieces that realize the above rhythmic pattern – but which do not seem to deserve the topical epithet “mazurka”, just as happens when one tries, problematically, to call the Pope a “bachelor”, or a lake a “water field”. Exactly such a problem can be found in the subordinate theme of the first movement of Beethoven’s last violin sonata, Op. 96. This theme is reproduced in Example 2.1-2. As the musicologist Maynard Solomon has remarked in his analysis of this piece (i.e. Solomon (2004)), this violin sonata is one of the few examples of a truly pastoral work in Beethoven’s oeuvre, and is thus replete with such pastoral themes as bird songs and folk dances, of the kind that can be found in the more famous pastoral composition of Beethoven, viz. his sixth symphony. So, the idea that one or more passages in this sonata would evoke a folk dance, possibly through their rhythmic structure, is quite reasonable. Solomon himself, therefore, understands the subordinate theme of the first movement as evoking some sort of dance, or more specifically a march, whose strong metrical accents provide a forceful contrast to the languid rhythms of the main theme – the latter purportedly evoking the sleepy rhythms of daybreak in the countryside.

But if you examine the specific rhythmic structure of the subordinate theme, it is the same as can be found in countless Chopin mazurkas (such as Op. 50 No. 1 and Op. 68 No. 3, to take just two examples we will examine in chapter 2.3 – the main theme of the latter is shown at the bottom of Example 2.1-2 for comparison with the Beethoven theme), and it accords with the above ‘statistical’ definition of “mazurka” too. So, is the subordinate theme of the first movement of Beethoven’s Op. 96 violin sonata a “mazurka topic” then? Does it even make sense to understand this theme as signifying a dance topic that is so closely associated with Polishness, a stylistic trait not commonly associated with Beethoven?

These are the kinds of problems that arise when one attempts to explain meaning in music (or language) without considering its grammatical bases, and in terms of its complicated social uses. So, for us to have a theory of topical meaning that is not subject to the above problems, it must describe topics in grammar-internal terms, i.e. it should show how topics are related to a generative grammatical procedure,

Example 2.1-2. Beethoven, Violin Sonata #10, Op. 96/i: Apparent mazurka 'topic' in subordinate theme

Violin

Piano

sfz

cresc.

rit.

sfz

cresc.

rit.

Chopin, Mazurka Op. 68 #3, mm. 1-8

Piano

p

and can therefore be given the kind of semantic interpretation available at an LF level of representation – of the kind we are not able to give the Beethoven theme (at least not in topical terms).

As mentioned earlier in the chapter, not all attempts to theorize musical meaning in externalist terms have involved discussions of the social use of music. Those less interested in the social uses of music, such as certain scholars working in the sciences (especially psychology or biology), have theorized musical meaning in externalist terms too, e.g. by relating it to extra-musical notions of affect. But notice what the music theorist Kofi Agawu, whose Schenkerian thoughts about musical meaning will influence much of the discussion later in this chapter, has to say about this attempt to connect musical meaning with affect:

“Interpretations that convey only overarching qualities of affect are held to be incomplete because they fail to lead us to the specifics of a musical text. Musical syntax, according to this view, is the true source of meaning. ...Musical particularities, according to this viewpoint, should be the determinants of musical meaning, not overarching attributions of affect or emotion that are not altered with the alteration of the specific notes that [certain composers] wrote.” (Agawu (forthcoming))

Now, compare this with the linguist Massimo Piatelli-Palmarini’s thoughts about *linguistic* meaning, in his foreword to Juan Uriagereka’s Minimalist text *Rhyme and Reason*:

“The unique role of grammatical meanings in Minimalist theory seem to be independent of any external, culturally-determined, domains of thinking and communicating. But even within syntax, it spans a range intermediate to that of individual words and full-blown sentences – and thus cannot be explained in terms of the meanings of individual word forms or the truth content of entire propositions. Moreover, they cannot be explained in terms of certain restrictions on word order, since the nuance with which such meanings affect grammatical generation operate across transformations of sentences, which violate surface word order. For these reasons and more, the level of analysis that involves such meanings cannot be considered a mere interface between, say, sound and meaning, but is rather an autonomous grammatical entity in itself.” (Uriagereka (1998): xxx)

So, the idea here is that even if music does have some connection to affect, this is a gross (or “overarching”, to use Agawu’s words) connection, one that does not relate to how nuances of meaning arise from combining individual pitch structures (such as chords), or words in language, into larger phrases. In other words, approaches to musical meaning that focus on affect, are more like the problematic linguistic theories of meaning that countenance meaning only in terms of the gross truth

content of propositions expressed by entire sentences. This means that grammar-internal approaches to meaning in music are better at accounting for the nuances of meaning expressed by musical texts. Moreover, they have greater explanatory power, since they can be related to grammar-internal approaches to linguistic meaning – both of which avoid any invocation of notions of truth and falsehood, in the way propositional approaches to linguistic meaning do.

What the above suggests is that one can at least conceive of the possibility of theorizing musical meaning from a linguistic perspective, because certain presuppositions about language – such as the belief that it is referential unlike music, and which have therefore obstructed a comparative study of musical and linguistic meaning – are not only based on faulty reasoning (as explored above), but also emanate from an externalist attitude to language that fails to account for the nuances of linguistic meaning.

For these reasons, I think a grammatical approach to musical meaning – founded in Hanslickian internalism – and a further comparative approach to musical *and* linguistic meaning, are therefore justified. Now Hanslick himself did not believe in the possibility of the latter project of course. But this seems to arise from his problematic, pre-theoretical assumptions about musical and linguistic meaning. For example, Hanslick's belief that music cannot be a system of signs, since it has no object other than itself, assumes that a sign can only be a sign when it signifies something other than itself. But as the philosopher Diana Raffmann has pointed out (regarding similar assumptions made by philosophers like Bertrand Russell), this just assumes what signs are, rather than defending this specific manner of understanding them – it rules out the very possibility of a non-linguistic symbol system (Raffman (1993): 40-41). Moreover, this assumption is partly based on the belief that signs have a reference, which is a belief that Russell held. But we have also seen that language is meaningful without necessarily being referential, as a result of which Peter Strawson criticized Russell for confusing meaning with reference.

The latter belief also underlies one way of interpreting the statement that language is a system of arbitrary sound-meaning pairings, in which meaning is taken to have a connection to reference. But from the preceding discussions we know that language can be considered a system of arbitrary sound-meaning

pairings without reference having anything to do with this. Unfortunately, the assumption that reference *does* have something to do with this prevents scholars from further considering the overlaps between linguistic and musical meaning. For example, in their recent proposals about music and language, the linguists Jonah Katz and David Pesetsky state that:

“All formal differences between language and music are a consequence of differences in their fundamental building blocks (arbitrary pairings of sound and meaning in the case of language; pitch-classes and pitch-class combinations in the case of music). In all other respects, language and music are identical.” (Katz and Pesetsky (2011): 3)

The point that the only formal differences between music and language are the result of their different building blocks is well taken, and one this dissertation accepts. However, this is not related to the *methodological* assumption that language is a system of arbitrary sound-meaning pairings, and that music is not. Such an assumption might make sense methodologically if one is not interested, or dismisses, a grammar-internal approach to explaining musical meaning, in favor of, say, a preference-rule approach, as Lerdahl and Jackendoff do. However, if a linguistic or musical theory does not even require such a definition of music or language, on methodological grounds, then there is obviously no need to assume it. Indeed, a Minimalist approach to music and language does not require this definition, as we have seen. So, rejecting an internalist approach to musical meaning – and thus a Minimalist research agenda such as ours – simply because the kind of meaning the system is dealing with does not fit a preconceived definition of meaning seems to be just the result of methodological dogma.¹⁴

This does not mean that an internalist approach is without problems though – especially problems that have to do with other problematic aspects of the general internalist approach to music being proposed in this dissertation. Two of these problems are the uncertain nature of the inputs and outputs to musical grammar. The inputs and outputs to grammar are, to put it in simple terms, lexical items and sentences respectively. But if we are uncertain about what a musical lexicon is, or what a musical sentence is, which

¹⁴ For this reason, Katz and Pesetsky’s above statement is particularly surprising, given its inconsistency with their generally Minimalist approach to music and language, and their belief that music and language are identical in important respects.

are issues we explored in the first two sections of chapter 1.2, we cannot be certain about lexical meaning or sentential meaning in music either. So, if it turns out that music and language have different lexical structures, it would make sense to assert that other parts of music and language that depend on these lexical ‘building blocks’ would be different too – which is the point Katz and Pesetsky make in their above statement. But since the matter of the musical lexicon is still up for grabs, as also discussed in chapter 1.2, I feel that asserting a dissimilarity between musical and linguistic meaning on lexical grounds still amounts to a methodological dogma in one’s choice of approach to these issues.

The uncertainty over what a musical sentence is, and the resulting uncertainty over what sentential meaning in music is, poses a bigger challenge for a Minimalist approach to music though. A musical sentence can be understood as a grammatical entity, when it is taken to be the output of a generative procedure, and is therefore governed by grammar-internal principles of structure generation. However, a musical sentence might be understood alternatively as a poetic entity when it is taken to be less a result of grammar and more a work of art, i.e. the result of a composer’s conscious artistic design of a musical piece, akin to the paragraphs, chapters and other discursive aspects of a literary artwork. Even though such extra-grammatical aspects of music usually fall under its social uses and functions, sometimes they can be the result of the same formal properties of structure generation that drive musical grammar, such as those connected with harmony and counterpoint. This is precisely the problem that arises when one understands the Schenkerian *Ursatz* as a grammatical structure in Schenker’s mature theories, when it can span entire symphonic movements, several hundreds of measures long – and therefore seems to be more poetic than grammatical in origin, even though it arises from the same principles of harmony and counterpoint that drive tonal grammar.¹⁵

¹⁵ Consider in this regard the parallel case in language, where a sentence like “Once upon a time there was an old man.”, can be understood in poetic or discursive terms, where it sets up the stage for a narrative to unfold, but it can also be understood as a grammatically complete structure, i.e. one that does not require further generation from a computational perspective, which is why we can end it with a period. So, grammatically the sentence is closed – if it were not, the sentence could not be interpreted. (Obviously, the sentence can be interpreted since it uses ordinary words from English that have well-known meanings associated with them.) So, this description of meaning will be a different one from one that engages inter-sentential discursive structures. And it is precisely this kind of meaning that I have been considering in this chapter.

But the poetic or discursive aspects of sentential meaning in music are somewhat secondary to its grammatical aspects, since without meaningful, grammatical sentences we cannot even have a discourse. So, discursive meaning is parasitic on the existence of grammatical meaning in the first place. In this light, semantic theories that illustrate how discursive meaning relates to grammatical meaning can therefore form an important supplement, rather than an alternative, to purely grammatical theories of meaning. For example, such a theory can show how musical sentences are related to, or vary from, the forms of other musical sentences. Such theories could then show how the grammar of a generated sentence, complete with its internal meaning, can affect the grammar of another meaningful, musical sentence, thus allowing us to describe, in turn, the intersentential meanings that arise as a result of this process. (After all, this process really amounts to no more than a description of the play of forms that determine how musical sentences are generated.)

Such intersentential theories are particularly helpful in understanding certain diachronic aspects of music – how music evolves, by composers of one generation altering the way sentences were generated by composers of a previous generation, through processes of variation, substitution etc. This is well-exemplified by composers who take themes written by earlier composers and write sets of variations on them in a new style (such as the variations on themes by Handel and Haydn written by Brahms), or composers who quote fragments of musical sentences from older music in their own compositions in a new style, often in humorous or ironic (and thus meaningful) ways (such as in the symphonies of Mahler or the smaller-scale works of various French composers like Saint-Saens, Debussy and Satie). In such cases, theories of intertextuality (such as the one proposed by Michael Klein (2005)) can provide suggestive ways of understanding such intersentential meaning.

We can see from the above that theories of intersentential meaning can also throw light on the problematic issue of musical style because they can show how the grammatical aspects of style influence its semantic aspects. After a composer has mastered the grammatical norms and procedures of a stylistic period, s/he usually wants to ‘say something new’. One way of describing how this happens is to show how new things are said by varying the way music was made in a previous stylistic period. In

consequence, new grammatical structures are generated, which give rise to a new stylistic period – but this is precisely why these developments are meaningful too, since they show how a composer is trying to *express* him/herself in novel, original ways. Thus, musical meaning yet again gets subsumed under a grammatical description of how musical phrases and sentences are generated.

To conclude, Hanslick’s view of musical meaning and the view of linguistic meaning contained in Minimalist grammatical theory show an important similarity, in that they both propose internalist theories of meaning for their respective domains. What Hanslick did not do, though, was ground his thoughts about musical meaning in a theory of musical grammar. This is what possibly led to him to ignore the important connection between music and language in the realm of meaning, which can be seen when both systems are understood in grammatical, and particularly Minimalist, terms. For this reason, one could even say that Hanslick did not really propose a formal, internalist theory of musical meaning, because this depends on a grammatical theory that explains how musical structure arises. (This is in contrast to the conventional wisdom that regards Hanslick as being the arch-formalist among theorists of musical meaning.)

I suggested a little while ago that the ideas of Heinrich Schenker, in line with similar observations made throughout this dissertation, contain certain proposals for thinking about musical meaning in grammatical terms – suggesting that Schenkerian theory might be the basis for developing a truly internalist theory of musical meaning, of the kind even Eduard Hanslick’s proposals fall short of. Moreover, it also seems to me that implicit in Schenkerian theory are proposals for musical analogues to the Minimalist notions of LF and PF representations. The notion of a musical PF grounded in Schenkerian ideas is something I will discuss in the next chapter. So, now I turn to a discussion of how a musical analogue to linguistic LF might be grounded in Schenkerian ideas about musical meaning.¹⁶

¹⁶ In addition to Schenker, Ludwig Wittgenstein had some interesting thoughts on the relation between music and language too, to the extent that he might be seen as making an internalistic claim for musical meaning, a la Hanslick – but *also* for language, in a way Hanslick never did. For example, in *The Brown Book* he makes assertions that appear to be overtly Hanslickian such as, “The same strange illusion which we are under when we seem to seek the

2.1.4. Schenkerian Poetics and its Relevance to Musical LF

Heinrich Schenker's proposals about musical grammar, discussed extensively in the first part of this dissertation, describe how the details of a musical surface are generated from a basic harmonic 'lexicon' in accordance with certain principles of counterpoint. But, more importantly, it is probably the only theory of music in which grammar is also seen as the basis for discussions on a variety of other musical phenomena such as musical meaning, rhythm, melody, form etc., as much recent scholarship in Schenkerian theory has demonstrated. In other words, Schenkerian theory is the kind of internalist theory in which phenomena like meaning are treated as inherent to grammatical theory, just as it is in the Minimalist approach to language. This seems to demonstrate the strongly Minimalist intuitions behind Schenkerian theory too, of the kind I discussed extensively in chapter 1.1, which owes possibly to the Goethe-ian philosophical foundation it shares with generative linguistics, and more importantly to the very identity of music and language, as the two identity theses of chapter 1.1 assert.

something which a face expresses whereas, in reality, we are giving ourselves up to the features before us – that same illusion possesses us even more strongly if repeating a tune to ourselves and letting it make its full impression on us, we say, "This tune says *something*, and it is as though I had to find *what* it says. And yet I know that it doesn't say anything in which I might express in words or pictures what it says. And if, recognizing this, I resign myself to saying, "It just expresses a musical thought", this would mean no more than saying, "It expresses itself"" (Wittgenstein (1958): 166). This statement clearly shows Wittgenstein's commitment to the Hanslickian idea that music meaning is not verbally recoverable, partly because music just "expresses itself". But he immediately goes on to say on the next page that, "What we call "understanding a sentence" has, in many cases, a much greater similarity to understanding a musical theme than we might be inclined to think. But I don't mean that understanding a musical theme is more like the picture which one tends to make oneself of understanding a sentence; but rather that this picture is wrong, and that understanding a sentence is much more like what really happens when we understand a tune than at first sight appears. For understanding a sentence, we say, points to a reality outside the sentence. Whereas one might say, "Understanding a sentence means getting hold of its content; and the content of the sentence is *in* the sentence". This shows that Wittgenstein, unlike Hanslick, was willing to extend the latter's internalist view of music to language too, anticipating Minimalism in this regard. (For more on this, see Mukherji (2010): 201-209.)

Wittgenstein's thoughts on these matters are interesting for another reason, viz. because of his connection to the Schenkerian tradition. Not only did Schenker and he share an internalist attitude to music, they had a shared personal history too. Wittgenstein was the uncle of Felix Salzer, one of Schenker's most noted students in Vienna, since his sister Helene had married Salzer's father, Dr. Max Salzer. Moreover, he was brought up in an amazingly musical environment, and apparently had perfect pitch, which might explain the interest he shows in music in his many writings. Like both Schenker and Hanslick, he disliked modern music, and was also an ardent admirer of Brahms, e.g. see Wittgenstein (1998): 27 and 29. In fact, Brahms was a family friend, who used to frequent the Wittgenstein household in Vienna, along with other musical luminaries like Gustav Mahler, and Richard Strauss, with whom his older brother Paul, a well-known concert pianist, allegedly played duets. Their brother Johannes, who died at the young age of 25 from a probable suicide, was a musical prodigy too (Monk (1991): 3-27).

Ultimately, though, Wittgenstein's developing interest in issues of language use, marked by his later publications such as the *Philosophical Investigations*, prevented him from developing the kind of grammar-internal theory of either linguistic or musical meaning that his earlier work hinted at.

So, the question is how does a discussion of meaning figure in the Schenkerian approach to musical structure, if indeed it does? How do we even know that Schenkerian theory has any bearing on a discussion of musical meaning? Well, for starters, consider the following quote by Schenker:

“The content of the second and subsequent levels is determined by the content of the first level, but at the same time it is influenced by goals in the foreground, mysteriously sensed and pursued.” (Schenker (1979): 68)

In this passage, Schenker is obviously describing the generative process through which higher levels of musical structure, all the way to the foreground, are generated from a background – the latter being of course the *Ursatz*, what Schenker refers to as the “first level” in the above quote. But despite this passage being about grammar, one could read it as illustrating Schenker’s sensitivity to the fact that musical grammar ultimately serves a semantic function. That is, he seems to be sensitive to the fact that a generative procedure has to produce musical surfaces that are interpretable (by grammar-external systems) – i.e. the generative procedure has to respond to the semantic and phonological conditions on surface structure generation, “sensed and pursued mysteriously”.

In addition to statements like the above by Schenker, scholars have often noticed a direct influence of Hanslick on Schenker as well, when it comes to discussions of musical meaning (e.g. Cook (1989a): 419, Korsyn (1993): 108), and the evidence for this is statements like the following, which clearly share Hanslick’s internalism:

“The musical motive is only a *sign of itself* [my emphasis], or better, nothing more or less than itself.” (Schenker (1988): 257)

So, it seems that musical meaning had a role to play in Schenker’s thought, and in an internalist manner akin to Hanslick’s views on the subject. But the preceding quote, made by Schenker early in his music-theoretic development, should be read with caution. This is because of the way it reveals Schenker’s interest in musical meaning only by means of his focus on musical motives. Schenker’s thoughts about motives would change drastically over the years, and his early interest in them is one of the things that put him squarely in the camp of someone who denies any deep connection between music and language.

Allan Keiler describes this early phase in Schenker's thought regarding the role of motives in musical structure, and how this blinded Schenker to the grammatical aspects of music – and also to any deep connection between music and language as a result:

“In order for the new musical material to be understood without the help of the word, the motive was created and repetition was necessary to insure musical understanding, since there was no associated referent to make understanding possible, as in the case of language.” (Keiler (1989): 278)

So, it was really his later interest in the problems of harmony and counterpoint, beginning with the *Harmonielehre* treatise of 1906, that would reveal Schenker to be a true grammarian of music, in a way that even Hanslick never was. In turn, this is what led Schenker to arrive at a model of musical meaning that was internalist in the true sense of the word, and which created the foundation for a deep comparison of music and language, of the kind this dissertation has been exploring.

Importantly, Schenker never did notice any deep connections between music and language, even in his later work, just like Hanslick. But unlike Hanslick he developed the way for such connections to be made explicit by others, because of his emphasis on the internal, grammatical basis for music. And this extends to issues of musical meaning too – for in the grammatical generation of surface from background, Schenker could show how musical meaning was strongly coupled to grammatical structure. This is something that many music theorists with an interest in Schenker have noticed, and some have incorporated this idea into their own explorations of meaning in music too (e.g. Samarotto (2007) and Rothstein (2004)). The theorist who has arguably explored such connections between Schenkerian theory and issues of musical meaning the most extensively though, is Kofi Agawu, who describes the connection between musical grammar and meaning, when seen in the context of a Schenkerian derivation of a Bach chorale, thus:

“This way of accounting for musical meaning is predicated on the significance of imagined origins, prototypes, invisible structures, fictional texts and background (including middleground) sonorities. These collective structures are functional – one feels or experiences their function in the passages to which they are assigned as origins or generators. For some musicians, it is precisely this juxtaposition of constructed proto-structures with Bach's actual structures that offers a compelling narrative of internal meaning. Within such acts of speculation we can glean something of Bach's tonal imagination. We can differentiate the ‘ordinary language’ of [the musical background and middleground] from the ‘poetic language’ of Bach's music. We can thus distinguish the conventional from the exceptional, the routine from the

inspired. This approach brings us into direct contact with the music itself, with the notes that Bach wrote in his score.” (Agawu (forthcoming))

And given that Schenker’s main concern in the vast majority of his published work was to explain, in grammatical terms, why composers wrote the notes they did in their scores, Agawu concludes that in Schenker’s approach we see “an earnest and relentless search for musical meaning in a range of configurations from the tiniest motive to the most elaborate sonata-form movement”.

Based on observations such as these, Agawu has devoted many pages of his writing over the years to proposals about how one might develop an internalist theory of musical meaning based on Schenkerian ideas. One of the clearest proposals in this regard is his notion of a “Schenkerian poetics”, which he discusses as a way of accounting for internal musical meaning, particularly in the analysis of song (Agawu (1992a): 24-29). Agawu’s contextualizing of Schenkerian poetics within song analysis is suggestive, because the phenomenon of song is one where one finds not only musical meaning, but also textual meaning (based on the lyrics of a song) – and the interplay of word meaning and musical meaning this allows enables one to think about a song’s musical meaning in textual terms, rather than in its own terms. Indeed, this is often the path taken by scholars who find textual analysis to be more acceptable than a purely musical one, given the technical complexities often inherent in the latter. This results in externalist descriptions of song meaning, in which a musical event is shown to be significant because of its connection to a significant, extra-musical, textual event.

An example of this can be found in the last line of the Schubert song “Einsamkeit”, from his *Winterreise* cycle, shown in Example 2.1-3. Notice the word “elend” (i.e. the German for “wretched” or “miserable”) in measure 44, which is harmonized by a German augmented sixth chord that functioned as an applied dominant seventh chord to flat-II in the previous measure. (The F-natural of the applied chord is respelled as an E# in the augmented sixth chord to facilitate this reinterpretation.) Such reinterpretations are fairly common harmonic practices in Western tonal music; but here it destroys the possibility of the

Example 2.1-3. Schubert, "Einsamkeit" from *Winterreise*, Op. 89 No. 12, mm. 41-48

Voice

Als noch die Stür - me tob-ten, war ich so e-lend, so e-lend nicht.

Piano

f *p* *fp* *pp* *dimin.*

B-minor: V⁷ VI

V⁷/₄II V₄ Ger⁵ V₄ i

8 : 7 : 3

chord resolving to the flat-II major triad, which it normally would do as an applied chord. This is because its reinterpretation as an augmented sixth chord makes it resolve to V instead (via a cadential 6-4 sonority), as augmented sixth chords conventionally do. This V then proceeds to the minor tonic triad of the piece, to give the song an authentic cadential ending in the home key. From an internal perspective, this signifies a motion away from major to minor. But since this motion also accompanies the feeling of misery that the singer is singing about, in a textual approach to song analysis *this* is what can be seen as giving the harmonic event its significance.

Of course such a passage can be given an opposite analysis, in which the meaning of the word “elend” is taken to be of particular significance *because* it is harmonized by the significant, major-modality destroying chord progression of the piece. But this is the point – in song, the co-existence of music and words allows for one to be interpreted in terms of the other, which leads to the possibility of “musico-poetic insights”, as Agawu calls them. However, such musico-poetic insights are only possible when there is an actual collaboration between text and music in a song. What happens, then, when there is no text to facilitate a musico-poetic reading of the song’s meaning, as happens in songs when the singer is silent, or in all purely instrumental works? As Agawu says:

“An enduring problem for the song analyst is deciding what to do with those aspects of structure that appear not to participate directly in the signifying processes of the text. Are they merely gaps to be ignored? Text-based interpretations... often bypass this problem by simply discounting musical information that the analyst does not need. This, however, does not make the problem go away.” (Agawu (1992a): 24)

This point is particularly relevant when seen in the context of our discussion in the last section of LF and its relation to internal meaning. There we saw that internal meaning spans a range intermediate to that of individual words and entire sentences, which is why internalist theories of meaning are not, or should not be, concerned with, e.g. the truth value of propositions expressed by complete sentences. This is also why theories of affect in *music* are problematic, because they do not deal with the nuances of meaning at the level below entire musical phrases, or at least in ways that track the fine grammatical structure of musical phrases. But it is for this reason that a musico-poetic approach to musical meaning, such as the one just

discussed in the case of Schubert's "Einsamkeit" is problematic – not because it deals with entire phrases, but because it deals with the opposite extreme, i.e. of individual words and their connection to individual musical sonorities, rather than with the nuances of internal meaning intermediate between these extremes.

It is problems such as these that led Agawu to propose a method in which one describes the meaning of a song not in terms of text-music relationships, but purely from the perspective of the internal 'play of forms' within a song's musical structure. Since a Schenkerian approach to the study of this play of forms is particularly powerful in its explanatory scope, for the many reasons we have explored in this dissertation, Agawu sees the best realization of his proposed method in a Schenkerian approach to song analysis, in which the focus could be on just the study of a song's music, or on how this music further informs its text-music relations – hence "Schenkerian poetics".

Agawu demonstrates the possibility of such a Schenkerian poetics via an analysis of the first strophe of "Seit ich ihn gesehen", the first song from Schumann's Op. 42 song cycle *Frauenliebe und Leben*. Through this analysis, which is essentially a Schenkerian interpretation of the grammatical structure of the strophe, Agawu shows how much of the singer's melody in this passage is an initial ascent to the scale degree 3 *Kopftón* D5 (the song is in B-flat major). After the *Kopftón* is reached, the inevitable, and in this case hasty, descent back to tonic, i.e. D5 – C5 – B-flat5, ensues. But this occurs over a deceptive V – vi harmonic progression rather than the conventional V – I one, which prevents the passage from attaining its structural close. When the closing V – I progression does occur, over the next two measures, the singer remains silent, so the descent to tonic in the melody ends up being performed by the right hand of the accompanying piano, which plays the notes of this descending line an octave lower. This means that the obligatory D5 – C5 – B-flat5 descent is replaced by a D5 – C4 – B-flat4 one, and this is how the strophe ends.

So, there are at least two structural features in the above passage that are unusual, and which therefore seem to be significant – first, the drawn-out initial ascent on the way to the climactic *Kopftón*, which builds tension in the passage, and second, the octave-transposed descent in the *Urlinie*, which closes the passage in an unusual way. Agawu reads these unusual structural features as implying a certain

meaning, viz. that of imbalance and uncertainty, especially in the way the passage closes, which creates an expectation for the next strophe. The drawn-out initial ascent in particular seems to suggest a struggle against the structural certainty that the attainment of the *Kopfton* will bring, with its inevitable subsequent descent to tonic – which Agawu reads as representing:

“...the protagonist’s awakening from her dream, conveyed forcefully by means of an *Anstieg*, [which] contrasts with the attainment of [scale degree] 3, which represents the real world. The sweep up to 3 encounters resistances and takes detours. The music continually falls back on itself, delaying but not ultimately denying the overall ascent. Musical metaphor may therefore support an interpretation of this prodding (bs 1-8) as marked by doubts and uncertainties. The turning point is reached in b.8, where the pattern is broken. The contrast between patterns on either side of 3 then becomes especially striking. The consistency (not straightforwardness) in the approach to 3 partly disintegrates in the second half of the song. It is almost as if, once the ambivalent status of trance or dream is broached, the course of the music grows less predictable.” (Agawu (1992a): 28)

In this manner, a purely grammatical description of a song passage can be shown to contain an internal meaning. This may further enhance an interpretation of the text of this passage – but importantly, it does not have to, since the grammatical description of the passage does not depend on the passage’s text. All of which shows how Schenkerian analysis has implications for an internalist theory of musical meaning, i.e. a Schenkerian poetics.

Now, Agawu’s description of Schenkerian poetics in the specific context of song analysis makes sense, because it helps contrast this internalist approach to theorizing musical meaning with the more externalist, text-based ones inherent in traditional song analysis. But this description can of course be extended beyond song analysis, and can inform a grammar-based, internalist discussion of musical meaning in any kind of music, precisely because it does not depend on textual content. Agawu demonstrates this in a more recent essay of his, which discusses the internal semantic aspects of the Bach chorale, “Ach Gott, wie manches Herzeleid”, BWV 153/ix (also known as “Herr Jesu Christ, meins Lebens Licht”). The score for this chorale is given in Example 2.1-4.

Agawu discusses the musical meaning of this chorale within a discussion of how the chorale’s phrases are generated. So, this situates the discussion of musical meaning within a discussion of musical

Example 2.1-4. Bach, Chorale "Ach Gott, wie manches Herzeleid", BWV 153/ix: Score

The image displays a musical score for the chorale "Ach Gott, wie manches Herzeleid" by Johann Sebastian Bach, BWV 153/ix. The score is written for a four-part vocal choir (Soprano, Alto, Tenor, Bass) and a keyboard instrument (likely organ or harpsichord). The music is in the key of D major and 4/4 time. The vocal parts are arranged in four staves, with the Soprano at the top and Bass at the bottom. The keyboard part is shown in a grand staff with a treble and bass clef. The score consists of 16 measures. The vocal parts enter in the first measure with a half note, followed by the keyboard part. The piece concludes with a final cadence in the 16th measure.

grammar again – and importantly, within a generative approach to musical grammar, which is the context in which we have been discussing internalist approaches to meaning so far, at least within the Minimalist Program. (This contrasts with the *analytical* orientation of his above analysis of Schumann’s “Seit ich ihn gesehen”.) And despite the fact that the chorale Agawu discusses sets a hymn text, this text never figures in his discussion, which illustrates how meaning and grammar are accounted for here in purely musical terms as well.

Let us now explore Agawu’s description of how meaning arises in this chorale, focusing specifically on how this happens in the generation of its last phrase, i.e. mm. 13-16 in the score. This generation is depicted in Example 2.1-5. In this context, Agawu says that he wants to:

“...show how a handful of relatively simple harmonic progressions lie behind its more decorated surface. What I want to show here is one path to the construction of the meaning of this chorale, a path to establishing its conditions of possibility. For these purposes, the harmonic and melodic content will be privileged, with the belief that rhythm plays a subsidiary role on the deepest levels of structure. The specific method of constructing meaning is to hypothesize certain prototypes for what Bach actually writes. This is not a claim about the way “Ach Gott” was composed; rather, it is a rational reconstruction of the chorale’s structure.” (Agawu (forthcoming))

In this light, Agawu’s description of the generation of the last phrase of the chorale in Example 2.1-5 begins with the *Ursatz* in its very first level, level A, which Agawu describes as “among the elementary utterances of the tonal system. ...that anyone who speaks the language of eighteenth-century tonality has internalized”. This *Ursatz* itself is elaborated with an 8-7 motion over the bass in the tenor voice, which changes the V chord of the *Ursatz* to a V⁷. At level B, a ii^{6/5} predominant chord is added to the background I – V – I structure, the bass of this chord leading to the bass of the following dominant chord via an added chromatic passing tone, viz. F#. This leads to the final, surface, level of the derivation, level

C. Regarding this transition from levels B to C, Agawu says:

“From here to Bach’s concluding four-bar phrase (level C) is but a short step. In the chorale, the initial tonic sonority postulated at levels A and B is extended by means of a voice exchange between outer voices; also, instead of the direct succession F-F#-G in the bass (level B), the note F is expanded into F-G-A before reaching F#. The effect is an expansion of the predominant area in a manner that is not unlike the expansion of the initial tonic in bar 13. With these simple steps we have been able to suggest the structural origins of Bach’s four-bar phrase.” (Agawu (forthcoming))

Example 2.1-5. Bach, Chorale “Ach Gott, wie manches Herzeleid”: Generation of mm. 13-16

A

I $V^8 - 7$ I

B

I ii^6 V^6/V $V^8 - 7$ I

C

I^6 vii^o6 I ii^6 V vi V^6/V V — I

But it is not as if we have just accounted for the structural origins of this Bach phrase. For it is in the derivation of this phrase that we can see what Agawu calls the “juxtaposition of constructed proto-structures with Bach’s actual structures that offers a compelling narrative of internal meaning”. And this

is what allows us to see how the “poetic language” of Bach’s music arises from the “ordinary language” of the phrase’s musical background and middleground.

Agawu continues his description of the internal meaning of the Bach chorale, by considering how its first and second phrases (mm. 1-8, and 9-12, respectively) are derived. These phrases are both elaborations of a simpler I – V progression, which in mm. 1-8 is expanded to twice the length of the expansion that occurs in mm. 9-12. I will not go into the details of how Agawu derives the actual surface structures of these phrases from their I – V foundation, but what is of interest is the internal meaning Agawu finds in this very I – V progression, and which can therefore be seen as the internal meaning of the first two phrases of the Bach chorale as well. Regarding this I – V progression, Agawu says:

“Unlike the I-V-I progression, which is closed, I-V is open, demanding or implying continuation and eventual closure. The psychological reality of a musical phrase framed by a I-V progression can, however, be stable and complete as a sound term in certain musical contexts; it can serve as a strong or potent manifestation of the competing tonal premises of a work. So even though an initiating I-V progression will normally feature a resumption and eventual completion of the process, I-V can sometimes feel less like a truncation of a fuller sentence than as an utterance that is at once complete and incomplete – complete on a local level, incomplete on a global level. Further, while the diatonic V that concludes a I-V progression conveys a sense of incompleteness, functioning analogously to a comma in grammatical punctuation, the tonicized V at the end of a I-V progression conveys a feeling of completeness because of the enhanced status of V as a temporary tonic. This double meaning of the I-V progression makes it especially suitable for expressing larger trajectories of musical thought.” (Agawu (forthcoming))

In this manner, we see how the grammatical generation of a musical structure is inherently meaningful, irrespective of any textual, or other extra-musical, associations these structures might have. And through a Schenkerian investigation of these structures, we can therefore account for this meaning, i.e. we can account for this meaning by understanding Schenkerian descriptions of musical structure as containing an internalist theory of musical meaning too, viz. a Schenkerian poetics.

There are a couple of things worth noting in this account of a Schenkerian poetics. First of all, this account deals with musical meaning within the context of grammar-internal operations on musical pitch structures – i.e. it deals with musical meaning within the context of the ‘plays of forms’ between musical pitch structures. However, this does not mean that this account is one of how meaningful pitch structures are *articulated* (i.e. sounded) in actual performance. It is just a description of the *abstract*

grammatical relations between pitch structures, and how their combination is inherently meaningful. This is an important caveat, because to the extent that this description of how internal meaning arises in music can be related to the Minimalist Program, how meaningful, grammatical structures are articulated in actual performance is something that falls under a *phonological* account of those structures, not a semantic one. (Such an account of musical structure is the focus of the next chapter.)

This is why Agawu's account of the Bach phrase in Example 2.1-5 does not consider the actual rhythmic articulation of the phrase's surface in its derivation of that surface. And it is for this reason that the account of musical meaning inherent in Schenkerian poetics is akin to the account of linguistic meaning inherent in Minimalist descriptions of *LF*, and not *PF*. Therefore, how all of the above discussions – of meaning in Schumann's "Seit ich ihn gesehen" and in Bach's "Ach Gott, wie manches Herzeleid" – relate to how these pieces are actually performed is something that comes under a description of musical phonology. But with such a description, the relationship between musical meaning and musical sound can be established too, i.e. this will demonstrate how music consists of sound-meaning pairings, just like language. The last chapter of the dissertation, chapter 2.3, will attempt such a demonstration through analyses of a few mazurkas by Chopin.

Another thing worth noting about Agawu's proposal for a Schenkerian poetics is that it actually fails to account for internal meaning in music in some important respects. Though it provides by far the most nuanced way of understanding the significance of various plays of forms in music, in some ways it is not nuanced enough. For example, Agawu's description of the internal meaning of the first two phrases of the above Bach chorale, in terms of the significance of the I – V progression from which they are generated, is spot on. But I – V is the gross progression from which these phrases are generated, meaning that a truly nuanced description of internal meaning in these phrases should be able to account for more than just the significance of I – V. Since Schenkerian theory attempts to account for each and every grammatical nuance in its description of how tonal phrases are generated, a Schenkerian poetics should account for the significance of all of these grammatical nuances too, or else it will fall short of its own

prescription that a theory of meaning should account for “musical particularities”, as Agawu puts it, and not just “overarching attributions” that remain unaltered with the alteration of specific notes in a piece.

Similarly, the I – V – I progression in the background of Example 2.1-5 is no doubt immensely significant, but a comprehensive Schenkerian poetics should be able to account for more than this. For example, it should be able to account for the significance of the voice exchange in the first measure of level C, which expands the initial tonic harmony. Also, it should be able to account for the chromatic passing F# in the bass in level B, whose significance is unquestionable since it did not have to be introduced into the passage – the ii^{6/5} predominant sonority preceding it could easily progress to the V following it without infelicity. In section 1.2.4.ii of the second chapter, I discussed how applied predominant sonorities, such as V⁷/V and vii^{o7}/V, seem to behave differently from diatonic ones, such as ii, IV and their seventh chord forms like ii^{6/5} or IV⁷.¹⁷ So, the introduction of the F# in level B of Example 2.1-5, which turns the diatonic ii^{6/5} into an applied V^{6/5}/V, has grammatical implications – and a Schenkerian poetics should therefore be able to account for the significance of this grammatical nuance in the phrase. So, just asserting *that* these grammatical nuances are significant, because of the way they perhaps reveal the imagination of the composer, is not enough – this significance needs to be explained.

In this regard, it might be worth revisiting linguists Jonah Katz and David Pesetsky’s proposal about musical meaning (Katz and Pesetsky (2011): 57-64), discussed before in chapter 1.2. Katz and Pesetsky do not propose an explicit theory of musical meaning, let alone a semantic level of representation within musical structure, akin to LF in language, despite their allegiance to the Minimalist Program, and their interest in the identity of music and language. Part of this has to do with their belief that tonal grammar deals only with pitch-class information, so they do not explore the possibility of a musical lexicon with inherent lexical features, including semantic ones. As a result, no semantic level of representation can result from their proposals about musical structure generation, even though such

¹⁷ Specifically, applied dominants appear to behave like *complements*, in the phraseology of X-bar theory, whereas diatonic dominants appear to behave like *adjuncts*.

structure is generated by the workings of Merge according to them – the result being the absence of an explicit theory of internal meaning within their theory of musical grammar.

However, Katz and Pesetsky do make a very interesting point about functional harmony, which has semantic implications. They observe, correctly, that the functions accorded to certain harmonies, such as tonic or dominant, and represented conventionally by Roman numerals such as I or V, change in the course of musical structure generation. For example, consider the first inversion D^7 chord in level B of Example 2.1-5, i.e. the third sonority in the level, with the $F\#$ in the bass. This D^7 chord is a $V^{6/5}$ of the following G-major harmony, as a result of which this latter harmony can be interpreted as a local tonic, i.e. I, within level B. But the G-major harmony itself arises in the passage as a V of C-major, at the level of the *Ursatz*, as can be seen in level A of the example. This is why in the *global* scale of the entire phrase, the first inversion D^7 chord is a $V^{6/5}$ of V , which merges with a subsequent *dominant* G-major harmony – which explains why these harmonies have been labeled the way they have in level B. All of which goes to show how the function of a harmony, and its associated label, can change in the course of generating a larger phrase that contains the harmony.

In light of this, Katz and Pesetsky argue that when a D^7 and G-major chord merge, to generate a larger structure whose head is the G-major chord, information pertaining to the harmonic function of these chords does not play a role in this generation – or more specifically, information pertaining to harmonic function does not *project* up a grammatical tree when these chords are merged (Katz and Pesetsky (2011): 60), meaning that the (local) tonic function of the G-major chord does not project higher up in the tree where the G-major chord merges with C-major chords to form the *Ursatz*, since of course at this level of structure the G-major chord no longer has the function of tonic. For this reason, Katz and Pesetsky believe that harmonic function is something that is *interpreted* from grammatically generated musical structures, instead of playing a direct role in this grammatical generation. This introduces the notion of *interpretation* into Katz and Pesetsky's thesis, and this, as we have seen above, has semantic implications.

Specifically, Katz and Pesetsky say that ascribing the labels of “I” or “tonic”, and “V” or “dominant”, to the constituents of a musical structure involves interpreting them as belonging to a key, such as C major, in which these constituents have tonic, dominant, or indeed some other harmonic function. But since such interpretation is separate from grammar itself, Katz and Pesetsky argue that this is the job of an extra-grammatical interpretive component, which they call the “Tonal-Harmonic Component”, and which they also take to be analogous to semantics in language, given the interpretive nature of the latter.

The advantage of this proposal, when seen in the context of a Schenkerian poetics of music, is twofold. First, it adds weight to Kofi Agawu’s argument about the expressive function of “tonic” and “dominant” in Western tonal music, which leads to the internal meaning of musical passages that contain progressions like I – V and I – V – I, such as those in Bach’s “Ach Gott, wie manches Herzeleid” chorale. In other words, the internal meaning that we find in passages like those in the Bach chorale, as revealed through a Schenkerian poetic study of such passages, can now be seen as resulting from the fact that understanding a passage in terms of its constituent functional-harmonic progressions is actually an act of *semantic interpretation*.¹⁸ Secondly, this also allows for a more nuanced understanding of the internal meaning of musical passages because if understanding such meaning involves just ascribing labels like “I” and “V” to the sonorities in a passage, then a detailed Roman numeral analysis will by itself be a nuanced semantic interpretation of the passage. In other words, my complaint against Agawu, viz. that his description of internal meaning in pieces like the Bach chorale are not nuanced enough, can now be refuted, on the grounds that Agawu’s very discussion of meaning in the chorale in terms of the role of “I” and “V” is as nuanced a semantic interpretation of such pieces as one can expect. So, supplementing a Schenkerian poetics with Katz and Pesetsky’s notion of a Tonal-Harmonic Component might be one way

¹⁸ Importantly, Agawu does not understand harmonic functions as being semantically interpreted entities, since they actually participate in the generative process for him, as Example 2.1-5 shows. (Although Agawu does consider them to have (internal) meaning.) But the inclusion of harmonic functions in grammatical generation is a problematic move, as we saw in the case of the ambiguous function of the G-major harmony in the above example. So, Katz and Pesetsky’s view on harmonic function, contra Agawu, seems to be the correct one.

of developing a nuanced, internalist theory of musical meaning, of the kind Minimalism develops for language.

Notice that we have not yet tackled the issue of whether the above internalist explorations into musical meaning have any implications for a musical equivalent to Logical Form, the semantic level of representation accorded to language by Minimalism. I think this is the reason why there is something rather unsatisfying about the conclusion of the last paragraph, regarding how a nuanced internalist theory of musical meaning might be developed by supplementing a Schenkerian poetics with the notion of a Tonal-Harmonic Component.

To understand this, let us review what we have established so far. On the one hand, we have considered the very sensible proposal that interpreting musical structures as having tonic or dominant function amounts to a semantic account of those structures, and that a Roman numeral analysis of a musical passage therefore amounts to a semantic interpretation of that passage. But on the other hand, there seems to be more to the meaning inherent in musical structures than their mere interpretation as having tonic or dominant function. This is why Agawu, when describing a passage that contains a I – V chord progression, does not find the meaning of the passage to be inherent in the functional-harmonic (i.e. “tonic” – “dominant”) aspects of this chord progression itself, but rather in the possibilities of, say, openness or closure, that the presence of these harmonies imply. And this is also why I made the point about Agawu’s description of internal meaning in the “Ach Gott” chorale not being nuanced enough – because he does not provide such descriptions of openness or closure for each and every grammatically significant sonority in the chorale, even though all of these sonorities are accounted for in a Schenkerian description of the chorale’s grammatical structure, and they are all interpreted in functional-harmonic terms with Roman numeral labels too (which should amount to a sufficient description of the chorale’s meaning in Katz and Pesetsky’s view of the matter).

The issue here, in my opinion, is that both Agawu in his Schenkerian poetics proposal, and Katz and Pesetsky, in their Tonal-Harmonic Component proposal, recognize the internalist nature of musical

meaning, and describe it as such, Katz and Pesetsky even connecting this to the Minimalist notion of interpretation. But we still do not have an account yet of *what* this entity is that is being interpreted, and that constitutes the grammar-internal representation of a musical structure's meaning. To put it another way, in comparison to language, an internalist theory of meaning in language describes meaning in a grammar-internal way too, but also in connection to entities that are clearly associated with some notion of semantics. And *this* is what leads to the notion of a Logical Form, which is a grammar-internal level of representation, being the semantic representation of a grammatically generated S-structure, but is also specifically the *semantic* representation of that S-structure. The entities that make up the Logical Form of a sentence are *semantic* entities, such as arguments and predicates and antecedents. So, if we want to conceive of a Schenkerian poetics as proposing a similar internalist theory of meaning in *music*, we need to think of what the specifically semantic entities are in music, which are necessarily the focus of such a theory.

From the preceding discussion, harmonic functions seem to emerge as such entities – i.e. according to these discussions, a Roman numeral analysis of a musical passage seems to be the musical Logical Form of that surface. But this is not a particularly satisfying conclusion, which is why Agawu says the things he does about functional harmonies like I and V that we reviewed two paragraphs ago. And even Katz and Pesetsky appear to sense this, since their discussion of harmonic functions only seems to be a step towards what they seem to think are the *real* semantic entities of music, and which therefore give rise to a Logical Form level of representation for music – i.e. entities like *cadences*, but which are entities that require a prior discussion of harmonic functions too, since it is functional-harmonic progressions like V – I that define cadences for Katz and Pesetsky:

“The role of the cadence in establishing a key, and the formal details of this role, motivated a distinction between the syntax of music proper (PR) and a Tonal-Harmonic interpretive component (THC) that mirrors the interaction of syntax and semantics within language.” (Katz and Pesetsky (2011): 59-60)

So, the question is whether we can propose a genuine *semantic level of representation* for music, akin to a Logical Form level of representation in language – over and above the internalist description of musical

meaning that we have developed so far, with its associated notion of interpretation? A second, related, question is how we should account for harmonic functions then, if they do not really constitute the semantic entities of music?

The answer to these questions will take us to the final discussion of this chapter, which will explore another proposal made by Kofi Agawu regarding internal meaning in music in the context of Schenkerian theory. So, to this final discussion I now turn.

Before he proposed his Schenkerian poetics of music, Agawu had already used Schenkerian ideas to explore internal meaning in music in his text *Playing with Signs: A Semiotic Interpretation of Classic Music*. In this text, Agawu, building on the linguist Roman Jakobson's notion of "introversive semiosis", asserted that one might understand meaning in music in terms of the internal significance of musical signs (Agawu (1991): 23). This internal significance has much to do with the ways in which musical grammar 'plays' with musical forms to generate intricately-structured, novel musical surfaces, which is therefore something that can be understood using the ideas and techniques of Schenkerian theory. But there is more to this significance than just its internal structure as described by Schenkerian theory, and which has to do with the fact that in generating structures, grammar endows musical surfaces with beginnings, middles, and endings. Describing this in relation to a passage from Mozart's K. 593 string quintet, Agawu says:

"Descending thirds, arpeggiations, stepwise motion, unfoldings – all these show the remarkable extent to which the music's inner processes are beautifully integrated. But were we to attempt to infer Mozart's foreground from these reductions, we would have great difficulty. This is not to claim that Mozart's is merely a surface drama, but rather to suggest that if the analysis does not reach the level of the surface where a great deal happens (or, as some might say, where it all happens), if there is no attempt to make sense of what is hardly a value-free musical surface, then one is missing something that is not only significant but fundamental to the piece. One way of approaching this surface drama is to reinterpret the Schenkerian analysis in light of the beginning-middle-ending paradigm in order to show the ways in which specific attitudes to these three related aspects of the piece are embodied in specific syntactical procedures." (Agawu (1991): 74)

So, beginnings, middles, and endings in musical passages are entities that are deeply coupled to the internal grammatical structure of these passages, but which also seem to inhabit a world of their own, i.e. a *semantic* world, in which musical structure is related to the *sense* of music. Therefore, beginnings,

middles, and endings might be considered truly semantic entities in music, akin to predicates and arguments in language, and the beginning-middle-end structure of a musical passage might therefore be considered the semantic representation of a musical passage – i.e. the musical equivalent to the LF level of representation ascribed to language by Minimalism.

This highly suggestive proposal might therefore be the answer to our search for a theory of Logical Form for music. And given its connection to Schenkerian descriptions of musical structure, as Agawu asserts himself, this could be one way of developing the notion of a Schenkerian poetics into a comprehensive internalist theory of musical meaning, whose connection to such a theory inherent in generative-grammatical descriptions of language strengthens the claim that music and language are identical, and further justifies the Minimalist Program for music *and* language being defended by this dissertation.¹⁹

For this reason, I will explore Agawu's thoughts on beginnings, middles, and endings a bit more, but I would first like to address the second question posed above, about harmonic function. That is, if beginnings, middles, and endings are the true semantic entities of music, what are harmonic functions? They cannot be grammatical entities, for reasons discussed earlier, and if they are not true semantic entities either, what role do they play in musical structure? My belief is that they do not play *any* role in musical structure. This might seem a radical thing to say, but harmonic functions are, after all, artifacts of a theory, viz. the theory of functional harmony developed by figures like Hugo Riemann and others – which means that their relevance depends on how relevant functional-harmonic theory is for music. In chapter 1.2, I claimed that the theory of scale steps, or *Stufen*, that Schenker subscribed to is sufficient for generating musical structures, and if this theory can be broadened to include the notion of beginnings, middles, and endings, as might happen in an expanded Schenkerian poetics, then this theory might be sufficient for interpreting the meaning of musical structures too. In other words, one could conceive of a

¹⁹ Obviously the claim here is not that linguistic structures have beginnings, middles and endings too, but that music seems to have an internal semantic level of representation like language. And since this level of representation depends on a grammar that seems to be identical in music and language, *this* is where the identity of music and language lies.

generative theory of music based in Schenkerian ideas in which scale-step information is used by Merge to generate S-structures, whose beginning-middle-end aspects can then be directly interpreted by the conceptual-intentional system – leading to a complete mapping of grammar to meaning in music, and *without* harmonic functions having anything to do with this.

This applies to the issue of Roman numeral analysis too. That is, the Roman numeral analysis of a musical passage does not have to be seen as a functional-harmonic interpretation of the passage, just because it uses labels like “I” or “V”. It could be just a shorthand for representing the scale-step structure of a passage (after all Schenker, who was not a fan of functional-harmonic theory, used Roman numerals in his analyses too, to represent scale steps), or possibly even a way of representing beginnings, middles, and endings. So, “V – I” could just represent an ending, which is many ways what Katz and Pesetsky seemed to be getting at with their notion of cadence, which they interpreted in terms of the specific functional-harmonic progression V – I.

Returning to the subject of beginnings, middles, and endings, Agawu traces the attempt to understand musical structures in terms of these notions to earlier music-theoretic endeavors, such as can be found in the writings of Carl Dahlhaus, or in the early 20th century Austrian musicologist Wilhelm Fischer’s notion of *Vordersatz-Fortspinnung-Epilog*. But Agawu’s important contribution was to relate the beginning-middle-end paradigm specifically to Schenkerian theory, so that in his proposal beginnings, middles, and endings are a way of “reinterpreting” (to use his words) the grammatical entities described in Schenkerian analyses of a musical passage. And this reveals the internal meaning of the passage – which Agawu characterized later as a Schenkerian poetics of music, although that later characterization was done in the context of song analysis, as we saw, and without explicit reference to beginnings, middles, and endings.

But despite this situation within a Schenkerian description of musical structure, beginnings, middles, and endings have their own specific characteristics, which is what makes them separate from, or rather interpretations of, the purely grammatical aspects of musical structure described by Schenkerian theory, and which is also what makes them genuinely semantic entities. As a result of this, each

grammatical entity in a musical passage can receive a semantic interpretation as a beginning, middle, or ending – as part of a musical LF representation of that passage – and this allows for a nuanced description of the internal meaning of a musical passage. This, in turn, allows a theory of beginnings, middles, and endings to serve as a comprehensive internalist theory of musical meaning, akin to that found in Minimalist descriptions of language. As, Agawu says, specifically regarding beginnings:

“Beginnings, then, are beginnings because they possess certain invariant characteristics. The most important of these in Classic music is that of a beginning as a detachable and internally complete abstraction, with a life of its own. I have suggested that every beginning contains or comprises a period in which is enshrined an expression of tonality evident in certain voice-leading patterns.” (Agawu (1991): 62)

Notice how Agawu describes the specific, “invariant” characteristics of beginnings, middles, and endings in grammatical terms, e.g. in terms of specific voice-leading patterns, which ensures that, as semantic entities, beginnings, middles, and endings are still *grammar-internal* semantic entities.

But it should be said that Agawu does not *always* think this to be the case for beginnings, middles, and endings. That is, though he does often think of these entities in grammar-internal terms, he does not *exclusively* think of them in grammar-internal terms. For example, he thinks that rhetorical characteristics sometimes play a role in defining beginnings, middles, and endings too. So, despite his above characterization of beginnings, Agawu also says that:

“A beginning constitutes or includes a period, it exploits the same closural functions as an ending. The difference between the two is therefore not only structural, but *rhetorical* [my emphasis]. Thus, the ending proper makes a point of its ending precisely because it is the completion of the global utterance – whereas the beginning necessarily understates its ending, since there are no competing claims as yet.” (Agawu (1991): 71)

That Agawu should take this stance is not surprising, because his attitude toward Schenkerian poetics, though governed by Schenkerian internalism, is never one that sees musical meaning in *exclusively* internalist terms. This can be seen in the very name of Schenkerian “poetics”, and its initial description within song analysis, where it allows not only for descriptions of (internal) musical meaning, but also for (external) musico-poetic meaning, as we saw earlier. This can also be seen in the fact that even in his Schenkerian description of mm. 21-33 of the first movement of Mozart’s K. 593 string quintet – from

which all of the above citations on beginnings, middles, and endings are taken – Agawu illustrates the beginning-middle-end structure of these measures *separately* from his Schenkerian analysis, and also presents a separate topical analysis of these measures – but with the express purpose of showing how these separate analyses are correlated, and can jointly inform an understanding of the meaning of these measures (see Agawu (1991): 76-77 for these analyses).

These hints of externalism that make their presence felt from time to time in Agawu’s characterization of Schenkerian poetics, introversive semiosis, and internal meaning in general, suggest that Agawu is not interested in restricting himself to an exclusively internalist position on musical meaning²⁰ – and that is fine; his attitude is that of a pluralist, who finds value in multiple approaches to understanding music, and for whom an ideological commitment to a single position, such as an exclusively internalist one, is perhaps unnecessarily restrictive, and maybe even repressive or reactionary.

But I have argued earlier how externalist approaches to musical meaning, including the topical one that Agawu invokes even in internalist, Schenkerian discussions of musical meaning, are problematic, and best ignored when trying to develop a systematic *theory* of musical meaning.²¹ For this reason, I think it is worthwhile to continue developing Agawu’s proposal of a Schenkerian poetics, via a focus on beginnings, middles, and endings, but in an exclusively grammar-internal way that Agawu himself does not attempt. And that such an endeavor is possible is evident from Agawu’s own analyses of meaning in musical pieces. For example, in his analysis of the above Mozart quintet passage, the information about the beginning, middle, and ending of the passage is all contained in the Schenkerian graph of this passage (to confirm this, compare the top two stave lines, which contain the Schenkerian graph of the passage, with the bottom two, which depict its beginning-middle-end structure). This means that the beginning-

²⁰ To this extent, his attitude toward musical meaning is similar to that of music semiologist Jean-Jacques Nattiez (whom I cited at the beginning of this chapter), and other thinkers in the structuralist tradition, for whom the study of music and language is an internalist enterprise, but who often move to an externalist position when it comes to studying issues of meaning and signification.

²¹ A further justification for this comes from the fact that the rhetorical characteristics of musical meaning, if included within a theory of musical meaning, will quite possibly preempt it from ever reaching the generality that any theory should, given that “attitudes toward the [rhetorical component of a musical sign] vary from genre to genre” (Agawu (1991): 67).

middle-end structure of the passage is inherent in the Schenkerian, grammatical, description of this passage – as we would expect it to be anyway, given the preceding proposals about how beginnings, middles, and endings are grammar-internal semantic entities and how the beginning-middle-end structure of a passage is essentially a musical LF representation. So the issue now is to develop an exclusively grammar-internal theory of the beginnings, middles, and endings of musical structures that is comprehensive enough, so that nothing more need be invoked to account for the internal meaning of these structures.

In the interests of space, I will not attempt to develop such a theory right now – especially since much more needs to be done in order to establish the feasibility of such a project, which presents an exciting avenue for future research in an internalist Schenkerian poetics of music. I would like to briefly discuss another project in music theory though, which has implications for future work in such a Schenkerian poetics.

This is William Caplin's well-known theory of form (as in "sonata form", not "logical form") in Classical music (Caplin (1998)). In this theory, Caplin discusses formal structure in the works of Haydn, Mozart and Beethoven, beginning with the form of simple 8-bar themes such as the Classical period and sentence, and leading all the way up to the formal structure of full sonata form movements. In dealing with these formal structures, Caplin discusses a variety of structural parameters that have a role in shaping different forms, including melodic-motivic structure, rhythm, meter, grouping, but also harmony, which plays by far the most important role in these discussions.

Now, the focus on form means that Caplin's theory is not a theory of grammar. But the importance of harmony in his description of form strongly couples his theory to actual theories of (harmonic) grammar, particularly Schenker's, and Caplin does cite parallels between his reading of the formal structure of certain musical passages, and Schenkerian descriptions of the grammatical structure of these passages too (e.g. see example 3.6, Caplin (1998): 38-39). Moreover, even though his theory is not a theory of grammar, it is a theory of beginnings, middles, and endings, and to this extent can be considered

a theory of musical meaning. (Interestingly, this suggests an equivalence between the notion of form as discussed in theories of music like Caplin's, and form in the philosophy of language and logic, given the former's connection to *musical* meaning through the beginning-middle-end paradigm, and latter's connection to *linguistic* meaning in the context of logical form.)

In his theory, Caplin discusses beginnings, middles, and endings under his concept of a "formal function" (Caplin (1998): 9-21), which is the function played by the parts of a musical structure, such as the parts of a Classical period, in shaping the larger form of that structure, either as the beginning of that structure, something that occurs in the middle of the structure, or something that ends it – and all of this specifically in musical structures found in the works of Haydn, Mozart, and Beethoven, although they might be applied to the works of other Classical composers, and even musical works from other styles or eras upon careful study of those other styles and eras.²²

William Caplin adapted the concept of a formal function from the form-theoretic writings of Arnold Schoenberg, and especially his student Erwin Ratz, who understood formal functions in a particularly Goethe-ian, organicist way, as arising from an abstract *Urform* – an idea that has interesting implications for a music-theoretic research project based in organicist ideas, even such as this one (although I will not pursue this particular issue here in the interests of space). Schoenberg and Ratz's influence on Caplin is particularly evident in a musical passage all three authors are well known for using to illustrate their ideas on musical form. This is the main theme of the first movement of Beethoven's Op. 2 No. 1 piano sonata, shown in Example 2.1-6. Schoenberg used this 8-bar theme to illustrate his concept of the *Satz*, also known as the "sentence", which is one of the main theme types in Classical music.

²² Sometimes the parts of a musical structure can have the formal function of both a beginning and a middle, or a middle and an ending (although not a beginning and an ending for obvious reasons), which is something Caplin refers to as "form-functional fusion" (Caplin (1998): 45). Also, a structure that contains such formal functions is something Caplin refers to as a formal "type" (Caplin (1998): 9). So, Classical periods and sentences, small and large binary forms, and even sonata forms, are all formal types, because each of these phenomena can be understood as being made up of smaller parts that have a beginning, a middle, an ending, or a fused function, all of which shapes the larger beginning-middle-end forms of these formal types.

Example 2.1-6. Beethoven, Piano Sonata #1, Op. 2 No. 1/i: Sentence structure of main theme

The musical score is divided into two main sections: **Presentation** and **Continuation**.

- Presentation**
 - Basic Idea**: The first measure, marked *p* (piano).
 - Repeat of Basic Idea**: The second measure, marked *sf* (sforzando), featuring a triplet.
 - Cadential Idea**: The third measure, marked *ff* (fortissimo), featuring a triplet and a dynamic hairpin.
- Continuation**
 - fragment**: The fourth measure, marked *sf* (sforzando), featuring a triplet.
 - Cadential Idea**: The fifth measure, marked *p* (piano), featuring a triplet.

Chord symbols below the bass staff are: i , V^6 , V^5 , i , V^3 , i^6 , ii^{o6} , V , and **HC** (boxed).

Caplin, building on Schoenberg's ideas, describes the sentence, as exemplified by this Beethoven theme, in the manner shown by the horizontal brackets in Example 2.1-6. That is, the sentence is normatively 8-bars long, and is made up of two 4-bar constituents, which are themselves made up of two smaller, 2-bar constituents each. The two 4-bar constituents that make up the normative 8-bar sentence are known as its "presentation" and "continuation" phrases. The presentation phrase is itself made up of a 2-bar "basic idea" (a term Caplin adapted from Schoenberg's *Grundgestalt*), which is repeated. The first two bars of the continuation phrase are normatively made up of fragments of the basic idea and its repeat, which in the case of this Beethoven theme are each 1-bar long. The final two bars of the continuation phrase are made up of a 2-bar "cadential idea".

This is the basic formal structure of the prototypical 8-bar sentence. But as mentioned earlier, in Caplin's theory, such a formal structure is deeply coupled to its grammatical structure, in particular the harmonic aspects of this grammar. So, as Example 2.1-6 illustrates again, the first two bars of the sentence, which contain its basic idea, prolong tonic harmony. (In fact, in this Beethoven theme, the prolongation of tonic harmony in mm. 1-2 does not even involve a second, prolonging harmony, since tonic harmony is all that we have in these two measures – which is why this theme is such a good illustration of the tonic-prolongational nature of the basic idea.) The next two measures of the sentence, which repeat the basic idea, could also be harmonized by tonic harmony, but they often prolong a second harmony. In Example 2.1-6, this is a first inversion dominant seventh harmony (i.e. $V^{6/5}$). If the repeat of the basic idea prolongs tonic harmony, it is described as an "exact" repeat of the basic idea. But if the repeat of the basic idea prolongs some form of dominant harmony, as happens in this Beethoven theme, or even subdominant harmony, it is described as a "statement-response"-type repeat of the basic idea. The basic idea can also be repeated sequentially, up a step, when the repeat prolongs supertonic harmony, or down a third, when it prolongs submediant harmony.

But even when the repeat of the basic idea takes a statement-response form, or is sequential, the harmony prolonged in its two measures can be understood as *itself prolonging* the tonic harmony of the basic idea – meaning that the larger 4-bar presentation phrase that is made up of the basic idea and its

repeat is also tonic-prolongational. So, in Example 2.1-6, the $V^{6/5}$ harmony prolonged in mm. 3-4, itself prolongs the tonic harmony of mm. 1-2, as dominant seventh harmony usually does, so that the entire presentation phrase can be understood as prolonging tonic harmony.

Turning to the continuation phrase, the first two measures of this phrase are just fragments of the basic idea and its repeat, as mentioned earlier, meaning that it shares their harmonic foundation too (although measure 6 is harmonized by a different inversion of the dominant seventh harmony of mm. 3-4, viz. $V^{4/3}$). This means, further, that these two measures also prolong tonic harmony, just as the basic idea and its repeat do. In other words, the initial tonic harmony of the basic idea is prolonged in this Beethoven theme into the first two measures of the continuation phrase – a point worth remembering for the discussion that will follow. This finally brings us to the last two measures of the continuation phrase, which contain the cadential idea of the sentence. These measures are harmonized by the chord progression $i^6 - ii^{06} - V$, which ends in a half cadence (or “HC”) at the end of the sentence. Caplin calls such chord progressions, unsurprisingly, “cadential progressions” – and specifically, a “half-cadential progression” in the case of this Beethoven theme. Importantly, Caplin defines cadential chord progressions in very specific terms, in order to distinguish them from, say, tonic-prolongational chord progressions, such as the $i - V^{6/5}$ tonic-prolongational chord progression of mm. 1-4 (i.e. the presentation phrase). So, a half-cadential chord progression, as can be seen in Example 2.1-6, involves a tonic or predominant chord, usually both and in that order, progressing to a *root* position dominant *triad*. So, just a dominant chord, or a progression from ii^{06} to an inverted dominant chord (such as V^6) or a progression from ii^{06} to a dominant seventh chord, even in root position (i.e. V^7) – would not be examples of half-cadential chord progressions according to Caplin.

So, this is the harmonic structure of the Beethoven theme, and it is strongly coupled to its formal structure too, as we can see in how the 2-bar and 4-bar constituents of the theme have specific harmonic foundations. But this is *exactly* where we see the scope of Caplin’s form-functional theory as a *grammar-internal theory* of musical meaning. This is because the terms we have been using to describe the 2- and

4-bar constituents of the Beethoven sentence, such as “basic idea”, “cadential idea” etc., are examples of Caplin’s formal functions – meaning that they are actually beginnings, middles, and endings. And more importantly, these, essentially semantic, entities (being beginnings, middles, and ends as they are) are strongly coupled to the harmonic grammar of the sentence, as we just saw, which confirms that Caplin’s theory of formal functions is essentially an internalist theory of musical meaning.

To understand this better, consider the notion of the “cadential idea” in mm. 7-8 of the Beethoven theme in Example 2.1-6. A cadential idea is normatively a 2-bar entity, which makes up the second half of a continuation phrase, as we saw above. But it also has a specific harmonic foundation – i.e. a cadential idea is a cadential idea because it is harmonized specifically by a *cadential chord progression*, as we saw above as well. But – and this is crucial – a cadential idea is *not the same* as a cadential chord progression. Whereas a cadential chord progression is a grammatical phenomenon, i.e. something that results from chords being merged according to grammatical principles of musical structure generation, a cadential *idea* is a formal function, specifically a closing formal function (Caplin (1998): 43) – in other words, it is an *ending*, a semantic phenomenon. (This is why the word “cadence” can be understood in multiple ways, at least one of which has to do with grammar, i.e. when one speaks of cadential *chord progressions*, and at least another of which has to do with semantics, i.e. when one speaks of a cadential *idea*.)²³

The point becomes clearer when we consider the fact that a cadential chord progression does not *have* to occur at the end of a phrase or theme, because this is not what defines a chord progression as a cadential chord progression.²⁴ Instead, what defines a cadential chord progression is something purely grammatical, e.g. the relation between the constituent chords of the progression, what inversion they are

²³ This is something Jonah Katz and David Pesetsky fail to notice in their discussion of cadences, which is why they speak confusingly of cadences as both grammatical phenomena, which act as instances of internal Merge in music according to them, and also as phenomena that are interpreted, i.e. as semantic phenomena (Katz and Pesetsky (2011): 58-60). Sensing this confusion, they seem to forfeit the latter use of “cadence” for a description of musical interpretation in functional-harmonic terms instead. That is, they seem to forfeit the idea that cadences are interpreted for the idea that the functional-harmonic components of a cadence, i.e. the functional-harmonic progression V – I, is what is interpreted. But Katz and Pesetsky need not have made this move, because cadences *are* semantic phenomena too, but only when we understand them as *endings* or *closing formal functions*, i.e. only when we understand them as *cadential ideas* rather than *cadential chord progressions* (like V – I).

²⁴ In fact, some cadential chord progressions, such as the deceptive cadential progression I⁶ – ii⁶ – V – vi, almost never appear at the end of a phrase because they are followed by another, usually authentic, cadential progression, the latter being the progression that ends the phrase.

in etc., as we saw above in the definition of the half-cadential progression. In other words, a cadential chord progression, or any chord progression for that matter, does not have a meaning of its own – *this* only happens when a cadential chord progression is *interpreted* as an ending, i.e. when we interpret it, form-functionally, as a cadential *idea*.

On the other hand, a cadential idea *is* defined by its being harmonized by a cadential chord progression. In this manner, we see how the formal function known as a cadential idea is a semantic entity in music, but one whose meaning is internal to musical grammar. And this is true of all the other formal functions proposed by Caplin in his theory of Classical form. For example, we understood mm. 5-6 in Example 2.1-6 to be fragments – but something can be a fragment only when it is a fragment *of* something else, viz. the basic idea and its repeat in this case. So, the meaning of mm. 5-6 clearly depends on the preceding basic idea and its repeat in mm. 1-2 and 3-4, which implies that mm. 5-6 cannot be a beginning.²⁵ But it cannot be an ending either, since it is not harmonized by a cadential chord progression – so we must interpret the fragments of mm. 5-6 form-functionally as “being in the middle”.²⁶ Caplin associates middles with the formal function of “continuation”, a word that suggests the persistence of something that has begun, but not yet ended – and hence the name of the phrase in Example 2.1-6 that mm. 5-6 initiates.

Finally, let us examine the form-functional aspects of the basic idea of mm. 1-2, and the presentation phrase of mm. 1-4, in Example 2.1-6, the latter made up of the basic idea and its repeat. We

²⁵ Notice how this resembles the phenomenon of “binding” in language, in which the meaning of certain noun phrases known as anaphora (like “himself”) depends on the meaning of certain antecedent noun phrases, the latter therefore “binding” such anaphora, according to Binding Theory in generative linguistics. (Review section 1.2.3 in chapter 1.2 in this regard.)

²⁶ Note how these fragments have not been described as formal functions themselves. This is because they are not formal functions, but just structures that arise from a certain formal *process* (Caplin (1998): 9), viz. that of fragmentation, which just involves breaking down a larger structure into smaller parts – and which therefore could happen to any structure, any where within a phrase. But the process of fragmentation creates musical structures that *then* must be interpreted, form-functionally, as middles, for the reasons mentioned above. There are other formal processes, similar to fragmentation, that also create structures that must be interpreted as middles. In Example 2.1-6, the fragmentation of the basic idea and its repeat in mm. 5-6 leads the harmonies of the basic idea and its repeat (i.e. i and V^{6/5}) to appear at a rhythm of one per measure in mm. 5-6, which is an increase from their earlier rhythm of one per two measures in mm. 1-2 and 3-4. An increased harmonic rhythm suggests heightened activity and climax, and is usually associated with the increase in tension one finds in the middle of musical passages. Therefore, *increasing harmonic rhythm* is another process that can make one interpret the structure in which it occurs as a middle.

have seen that the local harmonic foundation of mm. 1-2, and the more global harmonic foundation of mm. 1-4, is tonic prolongational. In other words, these measures are not harmonized by the kind of cadential chord progression that would merit their interpretation as an ending. Moreover, the very notion of tonic prolongation suggests a beginning formal function. (In this regard, think of Allan Keiler's grammatical proposal, discussed in section 1.2.4.i of the second chapter, which suggests that the tree structure of the *Ursatz* of a tonal phrase is made up of an initial "Tonic Prolongation" branch, and a final "Tonic Completion" branch.) The initiating formal function of a tonic prolongation becomes especially clear if we examine Example 2.1-6 in Schenkerian terms, which shows us that the pitches on the downbeats of mm. 1, 3, 5, 6, and 7 – all of which is harmonized by the same tonic prolongation that grounds the basic idea – also form an ascending line F – G – A-flat – B-flat – C, i.e. a Schenkerian *initial* ascent or *Anstieg*, which leads to the C5 *Kopfton* on the downbeat of measure 7 – which, in turn, initiates the cadential idea that closes the theme. So, the tonic prolongational basis of these measures allows us to interpret them as having a beginning formal function, which is exactly the formal function of a basic idea at the 2-bar level, and the presentation phrase at the 4-bar level.

What all of the preceding discussion shows is that the notion of a formal function, grounded in musical grammar yet independent of it, is a semantic notion – a notion of that which is *interpreted* from the structures generated by musical grammar. Which means that a form-functional description of a musical passage amounts to a description of that passage's semantic level of representation, i.e. its musical LF level of structure. And this description is nuanced too, which is a requirement of a theory of meaning we discussed earlier, since the theory of formal functions can deal, hierarchically, with the meaning of structures at the 1-bar level at least, and also the 2- and 4-bar levels as Example 2.1-6 shows – and beyond this as well, since the 8-bar sentence of Example 2.1-6 is but the *beginning* of a larger sonata form movement, several dozens of measures long, implying that the sentence itself can be interpreted as

having a formal function (viz. that of a beginning), which is an interpretation that can easily be extended to even larger spans of the movement's structure.²⁷

All of this goes to show that William Caplin's theory of formal functions, as a theory of Classical beginnings, middles and endings, is a genuine theory of musical meaning. But more importantly, it is a truly *internalist* theory of musical meaning too, because it describes beginnings, middles, and endings in Classical music in *exclusively* grammar-internal terms, without ever invoking any extra-musical reasons for why a musical structure should be given a specific form-functional interpretation.

So, in sum, a Schenkerian approach to musical grammar, when supplemented with Caplin's theory of formal functions, might be one way of developing a truly internalist Schenkerian poetics of musical meaning, akin to a Minimalist description of meaning in language, and one which also presents an associated description for a musical LF level of structure. There have indeed been some attempts to synthesize Schenkerian ideas with those of Caplin in this regard, a notable example of which can be found in the work of Janet Schmalfeldt (1991, 2011), however none of these attempts have been situated within a cognitive approach to semantic internalism in music and language, let alone from a Minimalist perspective. Therefore, for those interested in musical and linguistic grammar, meaning and their mutual relationship, this will be an exciting opportunity for future research.

²⁷ A final illustration of the nuance with which form-functional theory describes musical meaning can be seen in its ability to provide multiple interpretations of a musical passage's meaning, which is necessary for a rigorous theory of musical meaning, given the ambiguous ways in which music expresses itself. This can be seen in the theory's account of musical passages that begin with cadential chord progressions, such as the Trio of the third movement of Mozart's K. 551 "Jupiter" symphony. Since form-functional theory distinguishes chord progressions from formal functions, the beginnings of such passages will not be interpreted as endings just because they are harmonized by cadential chord progressions. If, for example, these beginning reveal a Schenkerian *Anstieg*, they can be interpreted as beginnings too. But the presence of a cadential chord progression right at the outset of the passage does suggest something anomalous and ambiguous about the passage, and form-functional theory, being the nuanced theory it is, is able to account for this too – given that cadential ideas *are* defined in terms of the cadential chord progressions that harmonize them, the *cadential* nature of the *beginnings* of musical passages that are harmonized with cadential chord progressions is something that form-functional theory is able to recognize, *in addition* to understanding them as beginnings.

Chapter 2.2

Musical PF, Meter and Schachter's "Tonal vs. Durational Rhythm"¹

Our discussion of the various associations between music and language has focused primarily on the pitch aspects of music so far. In this chapter, I will finally turn toward the other primary parameter of music, viz. rhythm. For the kind of music-theoretic inquiry being pursued in this dissertation, the study of rhythm is indispensable because pitch and rhythmic organization make up the two halves of music's formal structure. Moreover, rhythm is one of the important links between music and language too, given the prominence of rhythm in both music (as discussed in metrical theory) and in language (as studied in prosodic theory).

Despite the importance of the study of rhythm for a generative investigation of the structures of language and music, actual comparisons of rhythm in the two domains have been few and far between. As neuroscientist Aniruddh Patel says:

"The comparative study of spoken and musical rhythm is surprisingly underdeveloped. Although hundreds of studies have explored rhythm within each domain, empirical comparisons of linguistic and musical rhythms are rare. This does not reflect a lack of interest, because researchers have long noted connections between theories of rhythm in the two domains." (Patel (2007): 96)

Part of the reason for the lack of progress in this area is the fact that our understanding of rhythm *within* these two domains is far from being complete or coherent. Regarding rhythm specifically in music, Carl Schachter says:

"Meter [i.e. the primary heading under which rhythm is studied in Western music theory] is a problem. Theorists investigating rhythm have reached no consensus about such fundamental issues as the nature of the metrical accent, the possibility of meter over large spans of time, and the status of discrepancies between the meter as notated and some of the obvious rhythmic emphases in a piece. When different theorists attempt a metrical analysis of the same piece or passage, the divergent results can be striking indeed." (Schachter (1999c): 79)

¹ The ideas presented here arise from my Schenkerian studies with William Rothstein and Poundie Burstein, whom I would like to thank for helping me develop my ideas on Schenkerian rhythmic theory. Parts of this chapter will also be presented at the 13th Biennial meeting of the Society for Music Perception and Cognition, at Ryerson University, Toronto, Canada, under the title "Schachter's "Tonal Rhythm", the Grammar-to-Rhythm Mapping in Music, and its relation to Linguistic PF."

I believe that one of the reasons for the lack of consensus over different issues in rhythmic theory is that rhythm can be studied in at least two different ways, both of which are often conflated in rhythmic theory. On the one hand, rhythm can be understood in the context of other musical parameters, especially pitch. Therefore music theories that place an emphasis on the primacy of pitch grammar in understanding musical structure (such as Schenkerian theory) often take this approach towards studying rhythm too. This is the approach I will take in my discussion of musical rhythm in this chapter, partly because of the Schenkerian influence on the ideas to be discussed here and because the emphasis on pitch grammar has been the focus of this dissertation as well.

However, in contrast to the above approach, rhythm can also be studied as an independent entity that works according to its own rules, and is therefore often in conflict with the other aspects of musical structure. Since rhythm is one of the prime sources of variety and multiplicity in musical creation, it is this aspect of rhythm that creates problems for pitch-based theories of music. The different ways in which composers use rhythm to create imaginative new structures often stand in the way of theories that treat rhythm uniformly, as a result of pitch-based processes in music. In his discussion of the rhythmic theories of Edward Cone, and of Grosvenor Cooper and Leonard Meyer, Carl Schachter clearly asserts this problem:

“But neither approach makes headway against the central problem posed by this assumption [that rhythm is connected to pitch structure]: to distinguish between *rhythm as an active force* and *rhythm as resulting* from the activity of tonal events.” (Schachter (1999a): 33)

So, it is the distinction between these two ways of thinking about rhythm that has prevented theorists from arriving at a consensus about how to treat rhythm in music. As a result, even theories that are interested in delineating certain general principles of rhythm (which includes Schachter’s own rhythmic theory) often get sidetracked in this attempt by their entanglement with instances where rhythmic practice departs from a norm, or where rhythmic structure conflicts with other structural aspects of a musical passage, especially pitch structure.

The classic case of this phenomenon is the problem surrounding explanations of the metrically-unaccented cadence in tonal music, known previously by the politically-incorrect term “feminine cadence”. A metrically-unaccented cadence occurs when the (tonal-) hierarchically superior chord that ends a cadence falls on a weak beat relative to the hierarchically inferior chord that precedes it. (The best example is an authentic cadence in which the dominant chord falls on the downbeat of a measure, with the cadential tonic falling on the subsequent afterbeat.) Such cadences therefore pose a direct problem for pitch-based theories of rhythm, since they give clear examples of how a rhythmic phenomenon such as a metrical accent can conflict with a pitch phenomenon, such as the hierarchically-superior chord that ends a cadence. The theoretical challenge then becomes one of explaining how metrically unaccented cadences can occur in music that is *ex hypothesi* determined by pitch structure, and it is precisely the inability to explain such phenomena that leads to Schachter’s criticism of Edward Cone’s rhythmic theory (since Cone’s theory revolves around the notion of the ‘structural downbeat’, in which tonally-important events are equated with metrically-strong beats – the opposite of what happens in a metrically unaccented cadence).

In this chapter, I will first discuss another important distinction made by Carl Schachter, viz. the distinction between “tonal rhythm” and “durational rhythm”, which I believe will help us diffuse the conflict that arises from trying to arrive at a consensus between descriptions of rhythm as an active force and as resulting from pitch structure. Next, I will attempt to show how Schachter’s distinction actually reflects the very nature of generative *pitch* grammar, implying that rhythm is influenced by pitch grammar in important ways, which ties in with the centrality of grammar being discussed in this dissertation. I will build on this idea to explain the primacy of quadratic meter in music – a phenomenon that has not been discussed much in music theory despite its unparalleled prevalence in the world’s musical cultures. Since Schachter’s rhythmic theories are well known as being one of the most influential Schenkerian approaches to the problem of rhythm in tonal music, his thoughts will help place our discussion of musical rhythm in the broader cognitive and grammatical context of this dissertation, especially one that unites Schenkerian music theory with Chomskyan linguistic theory. This will allow us to end our

discussion by exploring an interesting parallel between rhythm in music and language, particularly with regards to the notion of Phonetic Form in linguistic Minimalism.

2.2.1. Schachter's "Tonal" vs. "Durational" Rhythm

Carl Schachter makes a distinction between *tonal rhythm*, i.e. the rhythm inherent in the hierarchical relations of pitches that make up tonal structure, and *durational rhythm*, i.e. the actual rhythm perceived at different levels of structure in a piece due to pitches being assigned real durational values (Schachter (1999a): 36-38).² As is implicit in these definitions, tonal rhythm is non-durational – the rhythmic properties it assigns to chords do not occupy time. So, tonal rhythm cannot determine meter, since meter

² Schachter's notion of tonal rhythm has some connections to the broader idea (expressed by some Schenkerians, e.g. Rothstein (1990a)) that musical rhythm results from the inherent temporal aspects of the relationship between consonant and dissonant pitch events in species counterpoint. When dissonant pitches are introduced in species counterpoint, specifically in the second species exercise, they are introduced as the second of two melodic pitches against a stationary cantus firmus, the first melodic pitch being consonant with the c.f. So, the entire second species model is, as is well known, a model of 'two-against-one' counterpoint, in which for every note in the cantus firmus, there are two notes in the counterpoint. This pattern has an inherent rhythm since the rhythm of the pitches of the cantus firmus is durationally twice as long as that of the pitches in the counterpoint. (The rhythmic relation is therefore a relative one – again, two-against-one – because no absolute temporal values are assigned to pitches in species counterpoint.)

This "rhythm as a result of pitch counterpoint" model of tonal rhythm is problematic for two reasons. First, there is a big difference between the consonance-dissonance relationships discussed in Schenkerian theory and those discussed in species counterpoint, because the former is an abstract theory of pitch grammar whereas the latter is a model for actually composing music. For this reason, the latter necessarily has a durational component (since one cannot compose real music without the pitches of that music occupying time), whilst pitch and duration are conceptually independent entities in Schenkerian theory – though they obviously get merged in actual music when a musical surface is created by prolonging a harmonic deep structure. It is this inability to distinguish the abstract properties of tonality, which composers might not even be conscious of, from the concrete and conscious attempt to compose music (inherent in pedagogical approaches to music) that ultimately made Schenker a critic of Johann Joseph Fux (the popularizer of the species model), and counterpoint pedagogy in general, even though the species model was a significant influence on Schenker's theorizing. For this reason, the Schenkerian *Ursatz* cannot be considered an example of a first species exercise, since the latter has a definite rhythm (normally notated with whole notes on a musical staff)

The second problem with the species approach to tonal rhythm is that the inherent relationships of consonance and dissonance that guide tonal rhythm are really the result of *harmonic* structure – counterpoint (and melody in purely monophonic textures) being but the means for realizing this harmonic structure in a musical surface through the process of prolongation. (This idea was explored earlier in chapter 1.3.) As Channan Willner says, "Many theorists have noted the simultaneous operation of various durational strata in tonal music. The durations and rhythms of these strata animate their tonal counterparts: Like the tonal strata, they are realized either explicitly or implicitly ... The all-important proliferation of implicit tones and durations, foreign to the species exercise, is made possible by the harmonic framework of the music, which pedagogical species exercises don't possess either" (Willner (1999): 210). So, Fux's species model is best seen as a pedagogical system that *instantiates* the relationship between deep tonal structure and surface rhythm – it is not a system from which this relationship is derived. In any case, Carl Schachter never explicitly connects his rhythmic theory with species counterpoint, so I will ignore that approach here.

is, according to most definitions of the term, a periodic division of time. It is only when pitches are assigned temporal, durational values (i.e. when they acquire durational rhythm) that we can talk about grouping them into metrical divisions, i.e. measures and hypermeasures. But even though tonal rhythm and meter are separate they are not equal, since the former definitely guides or constrains the latter in important ways. As Carl Schachter says:

“The designs of tonal rhythm often maintain a high degree of independence from any metrical scheme. Very frequently, however, aspects of tonal rhythm underscore the meter. The avoidance, in many styles, of bass repetitions from a weak to a strong beat, the related tendency to “change chord over the bar line,” the normal practice of beginning the measure with the lowest tone in Alberti or dance accompaniments – all of these show how tonal movement can help to express the meter. From a broader perspective, the appearance of an important goal of tonal motion at a metrically accented place can be an important compositional resource.” (Schachter (1999a): 40)

So, pitch can influence meter or, to put it in terms of the two new concepts that have just been introduced, tonal rhythm can influence durational rhythm.³ However, we have seen earlier that this influence is sometimes weakened, such as in phenomena like the metrically-unaccented cadence. This is why Schachter also insists on the “high degree of independence” of any durational phenomenon, such as meter, from tonal rhythm.

Lerdahl and Jackendoff share this view too, but phrase the discussion in terms of their theory of accents. For them, “structural accents” (i.e. accents resulting from tonal rhythmic activity) are different from “metrical accents” (Lerdahl and Jackendoff (1983): 17 and 30-35). Lerdahl and Jackendoff’s distinction between the two types of accents also helps explain why Lerdahl and Jackendoff famously separate “grouping” from “meter”. Since the former is determined by structural accents and the latter by metrical ones, the distinctness of the two kinds of accent results in the separation of grouping and meter too. Put in a different way, pitch groups are distinct from metrical units (such as measures) because structural accents do not correspond to metrical accents – if that were the case all structurally important pitches would fall on downbeats, which patently does not happen in metrically unaccented cadences.

³ Schachter does not equate durational rhythm with meter, but does say, “meter is much more closely bound to durational than to tonal rhythm ... durational rhythm can give rise to [pulse and meter].” In a sense, meter is a special case of durational rhythm – i.e. durational rhythm that has been organized into a system of hierarchical, periodic time spans.

But if tonal and durational rhythm – or structural and metrical accents, or grouping and meter – are separate, then how do we account for the asymmetry between structural events and metrical events – the fact that structural events (or at least their inherent tonal-rhythmic properties) seem to ‘underscore the meter’, as Carl Schachter says, but not vice versa? Or more specifically, how do we explain the effect tonal rhythm has on meter without collapsing the distinction between them, or that between tonal rhythm and durational rhythm? This is just a different way of stating the problem that we encountered earlier in the chapter – the fact that rhythm, though often an active force in musical passages, is also the ‘result’ of tonal events. It is one thing to assert, correctly in my opinion, that rhythm is not determined fully by pitch, especially if we do not want to make the same mistakes Schachter accuses Cone’s or Cooper and Meyer’s theories of making – and if we want to maintain the distinction between grouping and meter in the way that Lerdahl and Jackendoff do. However, it is an entirely different thing to say that rhythm is not influenced by pitch at all – and neither Schachter nor Lerdahl and Jackendoff make such an extreme statement in their rhythmic theories. In any case, Schachter ultimately subscribes to the Schenkerian belief in the primacy of pitch grammar in determining musical structure, so his assertions regarding the active role played by rhythm have to be subsumed under a pitch-based theory of musical structure in the end.

But even though he admits the asymmetry between pitch and rhythm in music, Schachter fails to provide an explanation for this asymmetry. Even Lerdahl and Jackendoff seem to accept this asymmetry (seen especially in their Metrical Preference Rules 5f, 6 and 7, henceforth referred to as MPRs), but ultimately stress the separation of pitch and rhythm via their insistence on the separation of grouping and meter. So how does pitch structure influence rhythmic structure, while still allowing it to be “an active force” in music?

2.2.2. Rhythm, Binary Pitch Grammar and Quadratic Hypermeter

The answer to this question takes us to the very heart of the grammatical story being told in this dissertation. One of the fundamental properties of generative grammar, as accepted by generative

linguists for over fifty years now, is its binarity – the fact that grammatical derivation proceeds by starting with *two* lexical items that have an inherent headed, hierarchical relationship with each other. The Minimalist Program in contemporary generative linguistics states that these binary pairs of items are then combined by Merge into (binary-branching) grammatical trees – which are, in turn, subjected to semantic or phonetic interpretation at the appropriate levels of structure. In the story I have told so far, this is *all* that is needed to describe the cognitive faculty of language (and, *ex hypothesi*, music), and thus a substantial part of human linguistic/musical behavior.

As I discussed in the first half of this dissertation, musical grammar is binary branching too, i.e. musical phrases are generated by merging two chords that have some hierarchical relationship into a binary tree. So, a I and a V chord, or a I chord with scale degrees 8, 5 or 3 in the soprano and another with scale degree 1 in the soprano, can be merged in a headed, binary, grammatical tree – the hierarchically-superior item forming the head of the tree – and this hierarchical tree forms the basis for the rest of the (pitch) generations in a musical phrase. Level A of Example 2.2-1 shows just such a binary tree, made up of two C major chords (of the kind that would realize a Schenkerian 3-line *Ursatz*, since the first chord has scale degree 3 in the soprano and the second scale degree 1).⁴

Now, the chords in Example 2.2-1A do not have any temporal duration, since their relationship is an abstract, formal one. So, a composer can choose to give each chord whatever duration s/he chooses in the actual surface of a piece. And this is exactly what composers do too – if a chord (or the phrase that arises from prolonging the chord) lasts a measure at the musical surface of a piece, that measure itself can have any duration at a given tempo – it can last two, three or four beats (the common time signatures in Western music), but also five or seven beats. However, this is only rhythm at the bar level.

It is a curious fact of music though, that whatever duration a composer assigns to the first chord/chord-phrase of a binary-branching pair is often assigned to the second one as well. (Obviously,

⁴ The second C major chord has been depicted as the head of this tree, following the conventional belief that tonal endings are more stable than beginnings. However, bear in mind our discussion in chapter 1.2 about how this position is open to debate – a debate I will ignore for our present purposes.

Example 2.2-1. Tonal rhythm and durational interpretation

A

Diagram A shows a binary chord pair. The upper staff (treble clef) contains two chords: a triad on the left and a dyad on the right. The lower staff (bass clef) contains two single notes: a whole note on the left and a whole note on the right. A single line connects the two chords in the upper staff, indicating a single durational unit for the entire pair. Below the staff, the letter 'I' is written under each chord.

B

Diagram B shows a binary chord pair. The upper staff (treble clef) contains two chords: a triad on the left and a dyad on the right. The lower staff (bass clef) contains two single notes: a whole note on the left and a whole note on the right. Two lines connect the two chords in the upper staff, indicating two durational units. Below the staff, the letter 'I' is written under each chord. Two arrows labeled 'prolongation' point from the first 'I' to the second 'I', one above and one below the staff.

Durational Interpretation: (♩ = 2 x n measures in any time signature)

C

Diagram C shows a binary chord pair. The upper staff (treble clef) contains two chords: a triad on the left and a dyad on the right. The lower staff (bass clef) contains two single notes: a whole note on the left and a whole note on the right. Two lines connect the two chords in the upper staff, indicating two durational units. Below the staff, the letter 'I' is written under each chord.

this does not always happen, since phrase groups can be of unequal lengths, and durationally-equal phrase groups are more typical of certain styles of music, such as the Classical style, compared to others, e.g. in the Baroque. This is an important point, to which I will return at the end of this section. For now, I will focus on cases where such symmetrical durational assignment does happen.) If a composer does assign equal durations to both halves of a binary chord pair (or the pair of phrases that arises from prolonging

these two chords), the duration of the entire binary chord/chord-phrase pair will be twice the duration of each individual chord/chord-phrase in the binary pair. So, if each chord/chord-phrase lasts one measure, the entire binary structure will last two measures. In other words, even though meter at the level of the notated measure can be of any length, meter at levels larger than the notated measure, i.e. at the level of *hyper-meter*, occurs in lengths that are often multiples of two – such as in the paired antecedent and consequent phrases of the Classical period or the paired presentation and continuation phrases of the Classical sentence. Moreover, the phrases that make up each half of a period or sentence are themselves often ‘quadratic’ in their hypermetrical structure as well because they commonly arise from a binary chord/chord-phrase pair. This explains the normative 4-bar length of antecedent phrases, which are

Example 2.2-2. Tchaikovsky, Symphony #6, Op. 74/ii: Quadratic hypermeter in main theme

The musical score illustrates a 4-measure phrase with quadratic hypermeter. The score is divided into four measures (1-4) and includes parts for cellos, flutes, oboes/clarinets, and a woodwind section. Dynamics range from *mf* to *ff*.

Measure 1: Cello part starts with *mf* and a triplet of eighth notes. Flute part starts with *mf* and a triplet of eighth notes. Oboe/Clarinet part starts with *mf* and a triplet of eighth notes.

Measure 2: Cello part continues with a half note. Flute part continues with a half note. Oboe/Clarinet part continues with a half note.

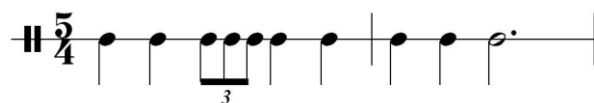
Measure 3: Cello part continues with a half note. Flute part continues with a half note. Oboe/Clarinet part continues with a half note.

Measure 4: Cello part continues with a half note. Flute part continues with a half note. Oboe/Clarinet part continues with a half note. The woodwind section (flutes, oboes, clarinets) enters in measure 5 with *ff* and a triplet of eighth notes.

themselves made up of an initial 2-bar basic idea, followed by a 2-bar contrasting idea, leading to the half cadence that usually ends the antecedent phrase (Caplin (2005): 12-13).

The above phenomenon, which we might call “quadratic hypermeter”, is not merely restricted to the Classical style, or to measures in common time signatures. Take for example the famous second movement of Tchaikovsky’s 6th symphony, shown in Example 2.2-2, which is in 5/4 time. Despite this unusual time signature, and despite this piece being written well after the end of the Classical era in Western art music, the main theme of this movement has the form of an 8-measure sentence, shown in the first two staves of this example, where it is played by the cellos, after which it is repeated in the last two staves of the example by the woodwinds, with a slight variation at the end. The theme itself is divided into a normative 4-bar presentation phrase, followed by a normative 4-bar continuation phrase, the first two bars of the presentation phrase, i.e. the “basic idea” of the sentence (Caplin (2005): 9), sounding the familiar ‘limping waltz’ rhythmic motive of the movement:

Example 2.2-3. Tchaikovsky, Symphony #6, Op. 74/ii: ‘Limping’ waltz motive



So, the total metrical structure of this theme is incredibly regular and four-square – even though each measure is in complex 5/4 time.

This symmetry occurs even in musical genres outside of Western Classical music. Example 2.2-4 depicts two such cases, from the first verse melodies of the songs “The Number of the Beast” by the heavy metal band Iron Maiden and “Them Bones” by the grunge band Alice in Chains. The verses of both songs are in complex time signatures, 10/4 and 7/8 respectively, though the choruses (not shown here) retain a common-time metrical structure. But despite the irregular bar-level meter, the hypermetrical

Example 2.2-4. Quadratic hypermeter in Rock: (a) Iron Maiden, “The Number of the Beast”, (b) Alice in Chains, “Them Bones”

1 2 3 4

I left a - lone. My mind was blank. I need - ed time to think to get the memo - ries from my mind. What did I see? Can I be - lieve that what I saw that night was real and not just fan - ta - sy? etc.

1 2 3 4

I be - lieve them bo - nes a - re me. Some we're bo - m in - to the grave.

structure of the verses in both songs is very regular. Each verse contains *two* lines, shown in successive staves in the example, each line being realized by a *four*-bar phrase in the music. Notice that this regularity does not owe to the rhyme structure of the songs' lyrics either, since the two lines that make up each verse do not share an end rhyme. In other words, the quadratic hypermetrical regularity here seems to be the result of purely musical factors. Another example of such hypermetrical regularity in the Western popular music tradition is 12-bar Blues, which is made up of three 4-bar phrases, each of which is comprised of two, paired 2-bar phrases.

Finally, this quadratic symmetry occurs outside of Western music too. Example 2.2-5 provides a transcription of the first minute of a duet Indian music performance by the *tablā* maestros Ustad Alla Rakha and Ustad Zakir Hussain – but the transcription is not of the *tablā* parts; rather it is of the *sārangī* part performed by Ustad Sultan Khan. In *tablā* performances, the *sārangī*, being the only melodic instrument, has the important function of repeatedly playing a melody, called the *lehrā*, that lasts a fixed number of metrical cycles (and often just one), and which therefore acts as a sort of metronome that cues the *tablā* soloists to where they are in a cycle. (In fact, the word “*lehrā*” derives from the Hindi/Urdu word for “wave”, so in effect the *lehrā* is like a melodic ‘wave’ that has the same periodicity as the metrical cycle to which the *tablā* players are soloing.) Example 2.2-5 transcribes the first minute of Sultan Khan's *lehrā* from the above *tablā* performance, which is set to the complex 7/8 meter *tāla* (i.e. metrical cycle) called *Pashto*, a folk meter of the Pashtun people who live in the trans-border region of Pakistan and Afghanistan.

As the example shows, this *lehrā* is actually four cycles long (which I have labeled as “Theme” here), and is made up of two, paired, two-bar phrases that I call “phrase I” and “phrase I variation”, the latter phrase acting as a response to the former. As the performance progresses, the first and then the second *tablā* enters and begins its improvisation, but the *lehrā* theme repeats over and over again, as it is supposed to – so its four-bar structure is heard over and over again too, sometimes with a new two-bar phrase inserted in place of phrase I (to create a variation of the Theme), and sometimes with minor

Example 2.2-5. Quadratic hypermeter in North Indian Classical music: *Pasho tāla*, theme and variations
 (performed by Sultan Khan, Alla Rakha and Zakir Hussain)

Theme
 Phrase I

Theme (Repeat)
 (Tabla 1 enters) Phrase I

Theme (Repeat)
 (Tabla 2 enters) Phrase I

Theme (Variation)
 Phrase II

Theme (Variation)
 Phrase II Variation

Theme (Repeat)
 Phrase I

x 3
 etc.

variations to one of these phrases as well. The end result being that even this 7/8, complex, metered piece has a consistently quadratic hypermetrical structure, made up of the four-bar *lehrā* theme and its two-bar components.

Generally in Indian music, the preference for quadratic durational lengths exists not only at the level of hypermeter but at the level of the ‘notated’ measure too. The most common *tāla* in the North Indian tradition of Indian Classical music is the supremely quadratic 16-beat *Teen tāla* cycle which is subdivided into two halves of 8 beats (the two halves being divided at beat 9, called the *khāli*), with each half being further subdivided into two halves of four beats each (Naimpalli (2008): 72). Similarly, the most common metrical cycle in South Indian Classical music is the different, yet equally quadratic, 8-bar *Ādi tāla*.

Finally, the *sthāyi* and *antarā* phrases that constitute most *rāga* compositions in the North Indian idiom normally last an entire *tāla* cycle each. This cycle is often the common *Teen tāla* cycle, as just mentioned, but it can sometimes be a much more complicated cycle of 13 beats, 11 ½ beats etc. But however complex the individual metrical cycle, the *sthāyi* and *antarā* phrases are normally *paired* together, just like an antecedent-consequent pair in Western tonal music, so that the length of the whole *sthāyi-antarā* group is equal to the length of *two* cycles – again, a simple quadratic number.

So hypermeter often tends to be quadratic. But the question is why composers choose to assign equal durational lengths to binary chord/chord-phrase pairs in the first place – which is how quadratic hypermeter arises, at least in Western Classical music. Remember, the quadratic nature of hypermeter does not result directly from binary chord grammar, since chords in a binary tree structure do not have duration. They have to be assigned a certain, equal duration for quadratic hypermeter to arise – indeed for any metrical structure to arise, since there cannot be any meter without duration.

One could take an externalist perspective here and argue that the preferred status of quadratic meter in the world’s music is a result of something external to the pitch grammar of music, something that has to do with the very nature of meter seen independently from pitch structure. Along these lines one can

cite arguments that base the preference for quadratic meter on the ‘dupleness’ of various biological rhythms, such as the heartbeat, or on the bipedal gait of humans. However, chords/chord-phrases in a binary tree do have a certain grammatical relationship, even if they do not have duration, which is why they are *paired* together in a binary tree structure in the first place. In this light, one might propose the *internalist* hypothesis that when composers assign equal durations to chords/prolonged chord phrases in a binary tree, they do so to reflect this paired status that pitch structures have within a headed, hierarchical relationship – and this is where the preference for quadratic meters in the world’s music might seem to arise as well.⁵ Given the internalist orientation of this dissertation, this is the perspective I shall explore in the following pages.

Now, one of the main features of generative grammar, as discussed at length in this dissertation, is the fact that it generates surface structures that are *interpretable*. That is, it generates structures that can be semantically interpreted (i.e. understood) by those parts of our minds that deal with meanings, beliefs, intentions etc. (also known as the conceptual-intentional system), and that can be phonetically interpreted (i.e. parsed and pronounced) by those parts of our minds that deal with the perception and performance of sounds (also known as the sensorimotor system). In fact, the Minimalist Program insists that this is the *only* reason why grammar generates surface structures, or else postulating a grammatical system would be conceptually unnecessary. (See the discussion on this in section 1.2.3 of the second chapter, to review this argument.)

Moreover, for grammar to generate a surface structure that is interpretable by the conceptual-intentional and sensorimotor systems, it has to fulfill certain constraints imposed on it by these systems. The Minimalist Program refers to this phenomenon as Full Interpretation. In the case of music, one could

⁵ This equal durational assignation could also involve a semantic element. A musical structure that signifies a beginning (e.g. the presentation phrase of a Classical sentence) is usually paired with a phrase of *equal* length that signifies an ending (e.g. a continuation phrase) – but this semantic relationship is itself one contained within tonal grammar. The very ideas of ‘beginning’ and ‘ending’ are semantic ideas, but ones determined by tonal *grammar* – which is seen most clearly in Allan Keiler’s division of a binary musical tree into an initial “prolongational” branch and a final “completion” branch (Keiler (1977, 1983-84)), as was discussed in chapter 1.2. So, the durationally-equal pairing of phrases seems to be ultimately a grammatical phenomenon.

understand the above proposal as stating that musical grammar needs to fulfill certain constraints imposed on it by extra-grammatical systems too, in order for it to generate musical surfaces successfully, i.e. for it to generate surfaces that are interpretable, by those parts of our minds that deal with *musical* conceptions and intentions, and with musical perception and performance.

If we consider (durational) rhythm and meter to be an extra-grammatical phenomenon, which we must in order to maintain the separation of tonal and durational rhythm, or structural and metrical accent, then we might consider rhythm to be governed by the sensorimotor apparatus of music, in the way speech rhythm (i.e. prosody) is said to be governed by the sensorimotor apparatus of language. This means that for musical grammar to successfully generate a musical surface, it has to be *rhythmically interpretable* by the musical sensorimotor system. This is an issue I will deal with in greater detail later in the chapter, when exploring the Minimalist notion of Phonetic Form in the context of music.

But all that the above thesis implies for our current discussion about pitch-rhythm relationships is that for a musical surface to be rhythmically interpretable, the pitch structure of that surface, as determined by musical grammar, should align with relevant aspects of rhythmic structure, as decided by the system that governs rhythm in our minds – i.e. the musical sensorimotor system, as per the above hypothesis. In other words, pitch structure should *map* to rhythmic structure successfully for rhythmic interpretation to occur, and for the generation of the surface by grammar to be successful.

Now, one crucial aspect of rhythmic structure, and specifically metrical structure, is that a rhythm is metrical only when it has evenly spaced beats. That is, any musical passage can be characterized as having a metrical structure only when the rhythmic aspects of that passage arrange themselves into evenly spaced beats. This is what is often referred to as a principle of metrical *well formedness* (Lerdahl and Jackendoff (1983): 69-74). But what this suggests is that the generation of a surface by pitch grammar will be *metrically interpretable* (which is just a special case of rhythmic interpretation), and therefore successful, only when the (pitch) structure of that surface aligns with the evenly spaced beats of a meter, as required by the rhythmic system in its principle of metrical well formedness. And since pitch grammar

is binary branching, the binary structure of the grammatical surfaces composers write will *automatically align with the evenly spaced meter of a musical passage if these binary structures have equal duration*.⁶

It is in this manner that rhythm is asymmetrically ‘influenced’ by, or is the result of, pitch structure (or the “activity of tonal events”, as Carl Schachter puts it). In other words, composers end up assigning metrical accents to phrases in a way that aligns those accents with the pitch structure – the result being that surface durational rhythm ends up corresponding with deep tonal rhythm too. This also results in the the alignment of metrical accents with the *structural beginnings* of a pitch surface, which is why the beginning of a new phrase, section or any other pitch group is often seen as a metrically strong event – resulting in the phenomena that Carl Schachter listed in his previously cited remark on the relation between tonal rhythm and meter, i.e. the fact that chords (and the phrases resulting from them) tend to change over the barline, or that bass repetitions (one of the stronger determiners of phrase beginnings, and also metrical accents, see Lerdahl and Jackendoff (1983): 88) tend to occur from *strong* to *weak* beat.⁷

All of this happens despite the fact that metrical and structural accents are separate, which implies, importantly, that structural accents can occur at places other than phrase beginnings too, i.e. in places where pitch groups do not align with meter. This must be so, otherwise left-branching pitch groups would not be possible, i.e. structures in which the head of the group (the locus of a structural accent) is *preceded* by pitch events that prolong this structurally-superior event. This idea is illustrated in level B of Example

⁶ Of course, they do not have to align; but the even spacing of phrase groups would then give rise to a conflicting meter – a “shadow meter” as Frank Samarotto (1999) calls it – that would conflict with the regular meter of the piece. Though shadow meters frequently occur in music, they are usually more of a surface phenomenon and cannot explain the rather deep connection between phrase groupings and meter at the level of tonal rhythm.

⁷ Note that the tonic event that represents the structural beginning of a passage was *not* treated as hierarchically superior to the tonic ending of the passage in the tree diagram of Example 2.2-1A. This might suggest that composers should align the meter of a passage with the pitch events that end the passage, if pitch really does map to meter in the way being discussed here. However, the labeling of the ending tonic as the head of the tree in Example 2.2-1A is itself controversial, as I argued in chapter 1.2. Moreover, metrical accents are not identical to structural accents, so metrical accents do not have to occur wherever structural accents do. All that I am saying here is that metrical accents tend to align with the *left branch of the highest tree* in a musical tree diagram – in other words, they tend to align with the *structural* beginning of a passage. Any subsequent structural and metrical accents within that passage may or may not line up – but they will line up again at the beginning of a new passage. Incidentally, if this new beginning occurs on a beat that would have been weak in the previous metrical cycle, this previously weak beat will now become a strong beat in the new cycle because of the mapping of pitch to meter – leading to the much-discussed phenomenon of *metrical reinterpretation* in tonal music.

2.2-1, where the second half of the binary chord pair from which the example is generated in level A, is prolonged by a preceding dominant harmony, resulting in the metrical accent (shown by the barline) aligning with the onset of the dominant harmony, not the following tonic chord where the structural accent occurs. This aspect of the relation between pitch grammar and rhythm not only reinforces the difference between structural and metrical accents but also shows how metrically unaccented cadences can occur, as actually happens in level B of the example.⁸

We can formalize the intuitions expressed in the above paragraph in the following way: since structural accents are likely to occur in multiples of two, given the binary manner in which tonal grammar generates them, it logically follows that it is likely for durational spans to occur in multiples of two as well. This leads to pitch groups and metrical accents lining up with each other, because of the requirement that pitch structure align with metrical structure for rhythmic interpretation. This idea is formalized in level C of Example 2.2-1 with the notation “whole note = $2 \times n$ measures in any time signature”, which implies that whatever n measure length a composer assigns to the chord/chord-phrase that makes up half of a binary tree, s/he must assign to the other half as well, to reflect the mapping of pitch to metrical structure, the result being that the entire passage will be $2 \times n$ measures in length.

As I noted earlier, the situation in actual music is frequently more complicated than the idealized situation described in the previous paragraphs, because composers often do not write durationally-equal phrases for chords/chord-phrases in a binary pair. However, unequal phrase pairs only occur in special circumstances and there are usually good grammatical reasons for their occurrence too. So, the relation between pitch and rhythm displayed in Example 2.2-1 still remains the norm – which just hammers in the point about rhythm being a ‘result’ of tonal events, and not just an active force in tonal music. Given the mapping of pitch to rhythmic structure, composers often follow the tonal rhythm of a piece very closely at the metrical surface, which leads to the kind of alignment between grouping and meter we have been

⁸ The above also helps us resolve a problem mentioned above – how to describe the influence of pitch on rhythm, in the manner that Edward Cone’s rhythmic theory attempted to, without preempting the possibility of metrically unaccented cadences, as his theory ended up implying.

discussing so far. However, a composer may create a conflict between meter and tonality in a given musical surface, for the sake of musical variety, in a way that resists interpretation. You might recall from discussions in previous chapters that just because grammar must generate surface structures that are interpretable, for the reasons discussed a few pages ago, this does not mean that successfully generated, and therefore interpretable, surface structures will be *interpreted*. The interpretation of an interpretable surface is not something that the grammatical system does itself, since this does not fall within the scope of musical “competence”. Instead, the relevant extra-grammatical systems are responsible for interpreting such structures, and this is susceptible to factors that fall under the notion of “performance” – which can include a composer’s idiosyncratic treatment of a musical surface, among the other social manifestations of music. (In this regard, review the discussion on competence and performance in chapter 1.1.)

I will discuss the above issue within a more detailed discussion of rhythmic, and more generally, durational, interpretation in the next section. For now, let us explore one of the more important ways in which a non-alignment between pitch and rhythmic structure can lead to difficulty in interpreting a generated surface. This can be seen in the phenomenon of *phrase expansion*, in which a phrase of a certain normative length is expanded to a phrase of greater duration by prolonging one or more of the component measures of the phrase. Example 2.2-6 from the second movement of Beethoven’s fifth symphony (taken from Rothstein (1989): 81) gives an example of such a phrase expansion. The measures marked “2” and “4” show how those single measures have been expanded to two measures each, thus turning the normative 2-bar groups in which they occur (i.e. mm. 1-2 and 3-4, which also mark the celebrated short-short-short-long rhythmic motive of this symphony) into expanded 3-bar groups. Importantly though, the underlying tonal structure of the piece is what makes the expansion of these two individual measures possible; the measures being expanded *prolong* the *tonic* C-major harmony in those measures for two bars rather than one, emphasizing the crucial structural role of that harmony, and of tonal grammar, in phrase expansion.⁹

⁹ Also, phrase expansions do not always expand phrases ‘rightwards’; they can also expand them prior to the structural beginning of the phrase, so that the phrase begins on the upbeat of a previous measure in the surface of the

Example 2.2-6. Beethoven, Symphony #5, Op. 67/ii: Phrase expansion from Rothstein (1989)

The musical score shows a piano part in 3/8 time, marked *tutti* and *ff*. The key signature has two flats. The score is divided into seven measures, with numbers 1 through 7 above the measures. Measure 1 contains a triad. Measure 2 contains a triad. Measure 3 contains a triad. Measure 4 is a prolonged chord. Measure 5 contains a triad. Measure 6 contains a triad. Measure 7 contains a triad and is marked *dim.* The piano part has a melodic line in the right hand and a bass line in the left hand.

Two aspects of pitch structure are made clear by this: (1) pitch does not determine duration, or else the durational length of the triad in mm. 150 and 153 would be fixed to one bar (as they are when first played at the beginning of the piece), and could not be expanded to two measures. So, a deterministic view of rhythmic structure is certainly *not* what is implied by the assertion that rhythm is the result of, or more accurately, aligned with, tonal events. All that pitch does is to participate in grammatical procedures that generate musical surfaces – which are then open to interpretation, in this case durational (rhythmic or metrical) interpretation, which is itself subject to a variety of performance factors. However, (2) even though pitch structure does not generate rhythm/meter, it is strongly aligned with it, since the task of the grammatical procedure that generates musical surfaces is to make these surfaces interpretable. As a result of this, grammatical procedures such as phrase expansion that lead to a breakdown of this alignment affect how rhythmic/metrical interpretation occurs. For example, in Example 2.2-6, the harmony being prolonged, and the phrase thus being expanded, helps clarify the important structural role of the C-major

piece. This is why phrase groups and meter can appear misaligned at the surface even when the beginning of the phrase group is actually aligned with the downbeat of the metrical unit in which that group is contained, in the deeper levels of musical structure. Hence the emphasis on the alignment between the *structural*, as opposed to surface, beginning of a pitch group and the beginning of a metrical cycle in the above description of how pitch maps to meter.

harmony in the piece – but it also *conflicts* with the regular 2-bar hypermetrical structure at the surface level of the piece, which is what makes interpreting the meter of this passage difficult.

In order to make the relation between pitch grammar and rhythmic/metrical interpretation clearer, we can ‘collapse’ the expanded phrase back into the quadratic, 2-bar prototype from which it arose, simply by taking away the expanded durational values assigned to the chords at the surface. This reverses the expansion procedure that made interpretation difficult, and therefore shows how pitch and rhythm are actually aligned at the deeper levels of structure from which the piece was derived, which is what makes interpretation possible. This process of reversal, called rhythmic normalization (or durational reduction when it involves hypermeasures), can show how apparent conflicts between meter and tonal structure at the surface get resolved at deeper levels of structure, where duration is stripped away and tonal structure (and tonal rhythm) govern the organization of a piece. (See Schachter (1999b) and Rothstein (1989, 1990, 1995) for wonderful demonstrations of this technique.)

So, composers can play with pitch grammar to create groups that conflict with the prevalent meter, often in the interests of variety. This is what justifies the separation of structural and metrical accents as well – e.g. in tonally weak structures that are placed on strong beats, such as in metrically unaccented cadences and also suspensions, or in cases of phrase expansion. Importantly, this also means that a durational reduction of such a surface can show that the greater duration assigned to an event in the musical surface through prolongation, and in the interest of variety, does not mean that it is structurally important in the tonal grammar as well. In fact, if the composer is intentionally emphasizing a tonally weak event for musical variety, a durational reduction will show the opposite to be true, i.e. it will reveal the event for the tonally weak structure that it is.

Example 2.2-7, the opening of the last movement of Beethoven’s Op. 31 No. 2 “Tempest” piano sonata, illustrates this manner of creating conflict between tonal and metrical structure at the surface of a piece. First of all notice how the upbeat figure right at the beginning of the passage expands the structural beginning of the main theme in measure 1 leftwards, so that the theme actually starts prior to the

Example 2.2-7. Beethoven, Piano Sonata #17 “Tempest”, Op. 31 No. 2/iii: Harmony in mm. 1-16

Allegretto

Norm: i _____ V _____

Actual piece: i _____ V _____

6

_____ iv _____

_____ i _____ iv _____

12

_____ V _____ i _____

_____ 6 - 5 _____

_____ v 4 - 3 _____ i _____

beginning of the meter on the downbeat of measure 1 – thus making the theme appear to be misaligned with the meter in this measure. Now, the main theme appears to have the structure of a 16-measure sentence, i.e. it can be divided into an 8-measure presentation phrase that prolongs a basic motivic idea in the tonic harmony, and an 8-measure continuation phrase that continues this idea in the next eight measures, leading to a cadence at the end of the structure. In music that has this formal design, the presentation phrase can usually be further subdivided into two halves, viz. a 4-measure phrase that presents the basic idea in tonic harmony and another 4-measure phrase that repeats this idea either in tonic

or another harmony, usually the dominant. (This formal design is exemplified, famously, by the main theme of the first movement of another Beethoven piano sonata, the one in F minor, Op. 2 No. 1, which we discussed in the last chapter, in Example 2.1-6.)

Example 2.2-7 reveals that this Beethoven theme has a similar design in its first 8 measures – a basic motivic idea consisting of 3 sixteenth notes and an eighth note is stated in the first half of the presentation phrase (mm. 1-4) in the tonic D-minor harmony and then repeated in the dominant A-major harmony for the second half of the presentation (mm. 5-8). However, the harmony that supports this motivic progression does not divide into the normative 4 measures each (of i and V). Instead, it follows an unusual harmonic rhythm – as can be seen in the lines drawn in score in the example, the initial tonic lasts just 3 measures, followed by 4 measures of V, and then one more measure of tonic.

Beethoven's non-normative treatment of harmonic rhythm here can be explained by showing how it works to hold this presentation phrase together. When the i and V chords occupy their normative four measures, the V chord gets a lot of emphasis, since it is given as much hypermetrical weight as the tonic – but this can also lead to a feeling of disjunction between the i and the V, since the V 'stands its ground' apart from the i. In a presentation phrase, this may even be desirable because the detachment of the V from the i gives the phrase a sense of forward propulsion that keeps in line with the 'presentation' nature of this phrase – the phrase is supposed to present, not sum up or close. But this progressive drive normally harmonizes an initial ascent (Schenker's *Anstieg*) in the melody, which ends with the attainment of a structurally-important melodic tone (Schenker's *Kopftön*) normally harmonized by a tonic triad – all of which serves to hold the presentation phrase together, as a tonic-prolongational event. Indeed, this is what happens in the Beethoven Op. 2 No. 1 theme just mentioned (review Example 2.1-6 in this regard). But as that example also shows, such a structure normally has tonic harmony at the *beginning* of the subsequent continuation phrase to confirm the tonic-prolongational nature of the preceding presentation phrase. However, Beethoven's presentation phrase in Example 2.2-7 is structured differently, since its continuation phrase begins with subdominant harmony. So, in order to maintain the tonic-prolongational nature of the presentation phrase in this theme, the tonic harmony that supports the *Kopftön* ends up

appearing at the *end* of the presentation phrase itself, rather than pushing forward into the continuation phrase.

The point is made clearer in Example 2.2-8, which presents Heinrich Schenker's own graph of this passage (Schenker (1979): figure 104). In this graph, Schenker clearly shows how the rising top voice in the presentation phrase is supported by a prolonged tonic harmony. This realizes an initial ascent to the F5 *Kopfton* of the passage. The generative nature of this sonority is demonstrated by the way Schenker beams this pitch to the final D5 over the tonic chord at the cadence in measure 15 since, after all, the entire passage is derived from the binary tree generated by merging these two chords. So, we see that the motion from measure 1 to measure 8 signals a completion of the initial tonic prolongation, not just a progression to the continuation phrase of mm. 9-16. By displacing the V chord from its normative position in mm. 4-8, Beethoven prevents a sense of disjunction from occurring and thus integrates all eight measures of the initial tonic prolongation into a whole. This means that the *three* part i – V – i prolongation of tonic that realizes the harmony of the presentation phrase has to be accommodated within the *eight* measures of the presentation phrase itself. This is impossible to do in an even way – and this is what seems to conflict with the 4 + 4 normative division of the presentation phrase.

Example 2.2-8. Beethoven, Piano Sonata #17 “Tempest”, Op. 31 No. 2/iii: Schenker’s graph, mm. 1-16

But this is actually true only of the musical surface of the presentation phrase. The i – V – i harmonic structure of the phrase, is really just its surface harmonic structure. The background harmony of this

phrase is really four measures of *i* + four measures of *i* *again*, since the phrase is generated from two tonic chords in a hierarchical relationship with each other (the one in measure 1 that starts the ascent, and the one in measure 8 that ends it). Given this tonal rhythm, the metrical structure of the presentation makes sense. It is divided into two symmetrical sub-phrases, each made up of four measures. The first four measures expand the first tonic, the next four measures the next tonic. The first tonic is expanded by means of the V triad in measure 4, and the next by means of the V⁷ harmony in mm. 5-7. Crucially, the change from a V to V⁷, by means of an 8-7 motion over the bass, happens at the halfway point in the presentation phrase (m. 5), which shows how the binary-branching structure of tonal rhythm affects this phrase's metrical interpretability in binary terms. So, the pitch and rhythmic structure of the presentation phrase really are symmetrical and therefore aligned, when its grammatical origins are examined. Which goes to show how Beethoven creates conflict between tonal and metrical structure at the surface of a piece, even though no such conflict exists in the deep structure of the piece – and which is why such a piece is interpretable.

Of course, the presentation phrase in the above Beethoven theme is not without asymmetries; 3 measures of *i* + 1 measure of V makes mm. 1-4 asymmetrical, as do the 3 measures of V⁷ + 1 measure of *i* of mm. 5-8. However, as mentioned earlier, a durational reduction of the passage can show how these asymmetries are resolved at a deeper level of structure. Example 2.2-9 presents just such a durational reduction. In level A of the example, each quarter note is equivalent to a dotted quarter in the actual piece, i.e. each quarter note equals a measure's worth of music, since the time signature for the piece is 3/8. The graph clearly shows the asymmetry in the harmonic rhythm discussed above, but it also shows how the presentation phrase divides neatly into two 4-bar halves. When we reduce the prolonging dominant chords out of the figure, we get the tonal deep structure of the passage, represented in level B of the example – which is nothing but the two tonic chords that generate the passage. Importantly, at level B the durational structure of the passage corresponds exactly with the tonal rhythm of the passage, most

Example 2.2-9. Beethoven, Piano Sonata #17, Op. 31 No. 2/iii: Durational reduction, mm. 1-8

The image displays two levels of durational reduction for the first eight measures of the third movement of Beethoven's Piano Sonata #17, Op. 31 No. 2.
Level A: Shows the original musical notation in 4/4 time. The treble clef part has a dotted quarter note followed by a quarter note in each measure. The bass clef part has a half note in each measure. A durational reduction line below the staff is marked with 'i' at the beginning and end, and 'v⁸ - 7' in the middle, indicating a binary division of the passage.
Level B: Shows a simplified harmonic background. The treble clef part has a whole note chord in each measure. The bass clef part has a whole note chord in each measure. A durational reduction line below the staff is marked with 'i' at the beginning and end, and '8' in the middle, indicating a binary division of the passage.
 A bracket connects the two levels, showing their alignment.

significantly in its binary division of the passage at the deepest level of tonal structure into its two hierarchically-related components. This reflects the grammatical origins of this structure, and its rhythmic/metrical interpretability – which owes to the alignment of grammar and rhythm at such deep levels of structure.¹⁰

Interestingly, the *final* tonic chord of the above 16-bar theme, in measure 15, does appear on a weak measure in the surface, which conflicts with the tonal rhythm of the passage. (Although, this conflict is unproblematic for the rhythmic theory I am suggesting here, as it happens at an ending – not a beginning – and involves the common, aforementioned phenomenon of the metrically unaccented cadence.) In order to work around this conflict, composers often ‘push’ the tonic chord onto the downbeat of the next measure to emphasize its important structural status. Such “end-accented phrases” (Temperley

¹⁰ In this light, the common-time signature of level B is somewhat arbitrary, since that level does not merely represent the durational values of the harmonic background of the passage – it also represents the non-durational tonal rhythm of the background. I chose the common-time signature to help relate level B to level A, but this level could have been represented equally well in cut time, or in 2/4 with the tonic chords notated by half notes.

(2003)) often occur at the end of a piece so that tonal and metrical resolution coincide, leading to a strong close for the piece. They can be created by expanding the tonic part of the phrase so that it extends to the next downbeat, or by metrically reinterpreting the entire phrase, so that the phrase itself starts not on count 1 of a hypermetrical cycle, but on a different hyperbeat. Eric McKee gives a wonderful example of this process in the finale of Mozart's G-major piano concerto, K. 453, in which the 4-bar main theme of the movement is gradually pushed backwards within its metrical frame during the course of the movement, so that at the end of the movement it actually starts on the second beat of a hypermeasure, and ends on the downbeat of the following hypermeasure (McKee (2004b)). Since the theme closes with an authentic cadence, this cadence therefore occurs on the hypermetrical downbeat at the end of the movement, bringing the movement (and the concerto) to a strong close.

Let us consider one last aspect of the Beethoven theme we have been examining. In Example 2.2-8, notice how Schenker treats the iv chord in measure 9 as structurally more important than the i^6 chord in measure 12, even though the latter chord supports the *Kopfton* as well. In contrast, Schenker treats the i^6 as an unfolding of the cadential 6-4 chord, which in turns prolongs *dominant* harmony. This clearly weakens the hierarchical status of this i^6 chord in the grammatical structure of the theme. We explored this passage within our discussion of musical ambiguity in section 1.2.4.vii of the second chapter, and as mentioned there, it is in this aspect of the Beethoven theme that we see the connection between grammar and interpretation most vividly. This is because treating the iv as a hierarchically-superior pitch structure aligns the pitch structure of the theme with its metrical structure – which makes it metrically interpretable, according to the preceding discussion. But this is what makes it *semantically* interpretable too. This is because once the tonic of measure 8 has been reached, at the end of the initial ascent, the presentation aspect of the theme, i.e. its *beginning*, is complete – it is time to move on. Beethoven stresses this by placing the iv chord, which initiates the continuation phrase of this sentence, on a strong beat. Therefore, tonally speaking, this is a true continuation phrase, it *departs* from the tonic and builds toward the final cadence of the passage. And as you might recall from our discussion of the theme from Beethoven's Op. 2 No. 1 piano sonata in the last chapter, presentations and continuations are *formal functions*, as William

Caplin calls them, that represent the meaning of a musical structure. So, by taking the iv in measure 9 of the Beethoven Tempest finale theme to be structurally stronger than the i⁶ in measure 12, Schenker maps the grammatical structure of the theme to its semantic and rhythmic aspects – i.e. he shows us how pitch grammar makes musical structures interpretable.

In the past few pages, I have tried to show how the binary-branching nature of tonal grammar maps to metrical structure in tonal music, specifically through the phenomenon of durational interpretation. In other words, the rhythmic properties inherent in the tonal system, in tonal grammar (i.e. tonal rhythm) show a strong mapping to durational rhythm (and subsequently meter), without being durational themselves. I will now turn to a more detailed exploration of some of the connections between pitch and rhythm in my Schenkerian approach to this topic, especially in the concept of durational interpretation. This will help us contrast it with Lerdahl and Jackendoff's proposals about musical rhythm, which constitute arguably the most comprehensive generative theory of rhythm in Western music. This will also allow us subsequently to compare our above conclusions about musical rhythm with descriptions of linguistic rhythm in Chomskyan linguistic theory.

2.2.3. Durational Interpretation and Linguistic Phonetic Form

We have now explored the idea that rhythmic structure in music is not determined by pitch structure. Binary-branching pitch grammar does not 'generate' rhythmic events in the true sense of the term – if it did we would not be able to maintain the distinction between structural and metrical accent, grouping and meter etc. So, the specific rhythmic structure of a musical surface is the result of interpretation, i.e. it is the result of how pitch structure is interpreted, possibly by the mind's sensorimotor system, as suggested earlier. In other words, it amounts to a composer or performer interpreting what durational values specific pitches should get, based on the structure of those pitches as decided by musical grammar – i.e. the process is one of *interpreting pitches durationally*. At the very least, a durational interpretation is a rhythmic one, since assigning pitches specific durations automatically endows them with a rhythm. (After

all, durational rhythm is nothing but the realization of musical pitch events in actual spans of time.) If the composer also interprets the duration of pitches in a musical passage in a way that either aligns or conflicts with a certain metrical scheme, we could call such an interpretation a “metrical interpretation” too. So, the phenomena of rhythmic and metrical interpretation we discussed in the previous section, are all just instances of durational interpretation.

There is one aspect of durational interpretation that we have already touched upon earlier, and which is particularly important for a generative theory of music. This is the fact that a composer or performer’s durational interpretation of a musical surface depends on a variety of performance factors. But most importantly, it depends on that musical surface’s being interpretable. This is, musical grammar must generate a surface that is *interpretable* to begin with, before it can be *interpreted* by a composer or performer. And durationally interpretable surfaces are generated by musical grammar by being correctly mapped to musical rhythm, which, as we have seen, has much to do with their shared binary structure.

Let me elaborate on this point a bit more by revisiting Example 2.2-1. I noted previously that when structural accents occur at the end of a (left-branching) pitch group, they often do not align with metrical accents (allowing for metrically unaccented cadences etc.). But when structural accents occur at the beginning of a pitch group, as occurs in the first half of level B in Example 2.2-1 (and in any right-branching pitch group), they *do* tend to align with metrical accents. This reveals an important cognitive aspect of meter that we have examined before – the fact that strong metrical accents align with pitch groups, usually the beginnings of such groups. Lerdahl and Jackendoff formalize this in their MPR 9 (Lerdahl and Jackendoff (1983): 89-90), as a preference for a musical structures in which conflict between metrical accents and structural accents are minimized.¹¹ William Rothstein describes this preference as a canonical regulation in several metrical traditions (Rothstein (1989): 75-76), to which he gives the title “the rule of congruence” (Rothstein (1995): 173). I believe that this preference is one of the

¹¹ This might also be a factor in another aspect of meter, which is that strong beats also tend to appear relatively early in a musical passage, which is what one would expect if metrical accents are to align with the *beginnings* of pitch groups. Lerdahl and Jackendoff formalize this aspect of meter as a preference for “strong beat early” in their MPR 2 (Lerdahl and Jackendoff (1983): 76).

strongest pieces of evidence for the asymmetry of pitch and meter – i.e. for the fact that meter can only be interpreted when pitch correctly maps to it. The congruence of structural and metrical accents is really nothing but an instance of the fact that when a pitch group begins, the meter tends to align with it – thus placing a strong downbeat at the point a new pitch group commences.

So, we tend to have a preference for aligned metrical and pitch structures because of the requirement that musical grammar generate pitch structures that are durationally interpretable. Lerdahl and Jackendoff say that, “Acutely out-of-phase passages [i.e. passages in which pitch groups and meter are not aligned] are more complicated for the listener to process because the recurrent patterns in the two components conflict rather than reinforce each other” (Lerdahl and Jackendoff (1983): 30). (This is similar to the point I made about shadow meters in footnote 6.) So, if a composer composes a structure that actively conflicts with meter, i.e. does not align with it correctly, it will be hard for a listener to process the structure – in other words, it will be hard for a listener to *interpret* the structure. It is in this manner that a composer is constrained by tonal grammar regarding just how s/he can compose certain structures – and this is why pitch and rhythm are asymmetrical; composers just happen to realize this by composing pitch structures and rhythmic structures that align with each other. This also seems to be the reason why composers write passages that involve pitch-meter conflicts only under special conditions, e.g. by composing a phrase that is actually aligned with the meter at a deeper level of structure but is made to appear out-of phase at the surface through, say, expansion. (So, the conflict is resolved at a deeper level of structure, after the expanded phrase has been normalized or reduced in some way.)

This brings me to an important point of comparison between the Schenkerian approach to rhythm being explored here, and that of Lerdahl and Jackendoff. In general, I view the Schenkerian approach to rhythm as being one that emphasizes the role of pitch grammar even in rhythmic structure (as exemplified by Carl Schachter’s very notion of “tonal rhythm”). This falls in line with Schenker’s broad belief that the *Ursatz*, the fundamental pitch-grammatical structure of tonal music, is the basis for *all* other kinds of musical structure. This implies that whenever an important rhythmic event occurs, there is often a pitch-based explanation for it. This grammar-centricity even in rhythmic events does not imply an *equivalence*

between pitch and rhythm, and it certainly does not imply the identity of pitch-based structural accents and metrical accents, as we explored earlier in Schachter's critique of Edward Cone and Cooper and Meyer's views on rhythm. However, it does imply a rather asymmetrical relation between pitch and rhythm. If we rephrase this in terms of Lerdahl and Jackendoff's theory, we could describe the Schenkerian view on rhythm as being one that suggests an asymmetric relation between grouping and meter – a view that suggests that grouping and meter are separate, but not equal. And as we just saw, there are good reasons for this asymmetry, which have to do with the role of pitch grammar in generating durationally interpretable structures.

Lerdahl and Jackendoff describe such a view as an instance of either Hypothesis A or Hypothesis B in their schema of the different ways in which grouping relates to meter (Lerdahl and Jackendoff (1983): 31-32). According to Hypothesis A, strong beats coincide with the beginnings of (right-branching) groups, but not with the ends of groups, whereas strong beats coincide with the (left-branching) ends, but not beginnings, of groups, in Hypothesis B. My above arguments about the relation between grouping and meter clearly suggest a preference for Hypothesis A, which is a view Lerdahl and Jackendoff ascribe to both Schenker and Donald Francis Tovey (Lerdahl and Jackendoff (1983): 335) – unsurprisingly perhaps, given the Schenkerian tenor of this essay. On the other hand, Lerdahl and Jackendoff view Hypothesis B as illustrating Hugo Riemann's, and Grosvenor Cooper and Leonard Meyer's, approach to rhythm. Finally, Lerdahl and Jackendoff also discuss a third alternative, Hypothesis C, in which strong beats occur at both the beginning and end of a group – a view that clearly tries to equate metrical accent with structural accent, since in this view both the initial tonic and the final, cadential tonic of a phrase will receive strong beats, preempting the possibility of metrically unaccented cadences. This is the problematic view that Lerdahl and Jackendoff (and as we have seen before, Carl Schachter) ascribe to Edward Cone, and also to Peter Westergaard.

Although they clearly disfavor Hypothesis C, Lerdahl and Jackendoff remain neutral between Hypothesis A and B, claiming that, "A dogmatic preference for either hypothesis would distort the flexible nature of the situation; one or the other – or perhaps something more complicated – pertains in a

given instance. In a broad sense, in-phase passages usually yield Hypothesis A and out-of-phase passages usually yield Hypothesis B” (Lerdahl and Jackendoff (1983): 32-33).

Lerdahl and Jackendoff are certainly correct in asserting that the relation between pitch and rhythm can be complicated in actual music. (After all, the perfect alignment of grouping and meter is an idealization, which is routinely violated in cases of phrase expansion etc.) However, it is one thing to (correctly) insist on the flexible relation of grouping and meter and quite another to (incorrectly) insist on the symmetry of this relationship. And it is not as if Lerdahl and Jackendoff actually insist on a symmetrical grouping/meter relationship anyway – they seem to be fully aware of the asymmetry of this relationship. At the beginning of this chapter I mentioned how Lerdahl and Jackendoff’s MPRs 5f, 6 and 7 betray the strong influence of pitch on rhythmic interpretation. Similarly, earlier in this section I mentioned how MPR 9 suggests an influence of pitch structure on rhythmic interpretation that actually aligns Lerdahl and Jackendoff’s metrical theory with Schenkerian theory. (A similar argument can be made for MPR 3, according to which a metrical accent is preferred at places where pitch events occur.) Last but not least, Lerdahl and Jackendoff generally seem to favor the Schenkerian approach to rhythm over the other approaches they discuss, specifically the same rhythmic theory proposed by Carl Schachter that has been the basis for this chapter (Lerdahl and Jackendoff (1983): 335-336).

Then why is it that they do not commit to either a Hypothesis A- or B-type understanding of the relation of grouping and meter, especially the former, as this dissertation does, and as they believe Schenker did? Related to this, why do they view rhythm important enough to discuss in as much detail as they do pitch structure, given the general underemphasis on rhythm in Schenkerian theory? (Note that Lerdahl and Jackendoff do not have to commit specifically to a Schenkerian, Hypothesis A-type understanding of the relation of grouping and meter to reveal a further commitment to the asymmetry of this relationship, since a commitment to a (Riemannian) Hypothesis B-type understanding of the relation of grouping and meter would reveal this further commitment to grouping/meter asymmetry as well. The point is that one has to commit to *either* Hypothesis A or B to reveal this further commitment to grouping/meter asymmetry – but Lerdahl and Jackendoff commit to *neither*.)

One of the answers to these questions is that Lerdahl and Jackendoff consider rhythm to be as critical as pitch in discussions of musical structure, so much so that they believe that musical structure cannot be described adequately without taking rhythmic factors into consideration (Lerdahl (2009): 188-190). This is particularly evident in their notion of “time span reduction” (Lerdahl and Jackendoff (1983): 124-178) in which both pitch and rhythmic factors interact to determine the grammatical structure of a musical passage. This influence of rhythm in grammatical thinking, or rather this ‘network’ approach to grammar – in which multiple components interact in the determination of musical structure – has been central not only to Lerdahl and Jackendoff’s musical theory, but also to Jackendoff’s own work in linguistic theory (Jackendoff (1999, 2002)), to the extent that Jackendoff has parted ways with the more grammar-centric approach of Chomskyan linguistics, to embrace a theory of linguistic structure that greatly emphasizes the role of semantics and phonology as well (including the rhythmic aspects of phonology, viz. prosody). So, if a (Schenkerian) emphasis on grouping *over* meter, on pitch *over* rhythm, betrays such a grammar-centric attitude (as I have tried to argue is the case throughout this dissertation), then it is understandable why Lerdahl and Jackendoff would oppose such a strategy in describing musical structure.

Now, as should hopefully be clear from the above discussion, the Schenkerian attitude toward rhythm certainly does not deny the importance of rhythm in musical structure. However, it does make a distinction, following Schachter, between rhythm “as the result of tonal events” and “as an active force”. Lerdahl and Jackendoff only seem to be interested in rhythm in its latter manifestation. So they tend, in my opinion, to overemphasize the importance of rhythm in understanding musical structure vis-à-vis pitch, while ignoring the strong influence of pitch on rhythm, i.e. on durational interpretation, even though this influence is clearly implicit in the formulation of their MPRs.

One of the reasons Lerdahl and Jackendoff seem to be interested only in the active role of rhythm is because their approach is geared toward listening. It aims to describe how the informed listener assigns structure to a perceived musical surface – an issue that we have touched upon several times in previous chapters. And rhythm is certainly an essential tool in assigning structure, especially if the surface has a

complicated structure. As we saw in the last chapter, even grammatically-correct *linguistic* surfaces can be hard to interpret without additional knowledge – of the sentence’s grammatical structure in particular, but of other things like the meaning of its words too. In the case of music, therefore, non-grammatical phenomena (i.e. non-pitch phenomena) like rhythm can really aid the informed listener’s attempts to parse a complex musical surface.

But this is a very different enterprise from the actual generation of musical surfaces, the process of *making* music – which is something that happens before the problem of interpreting a generated musical surface even arises. And the Schenkerian tradition seems to be more interested in this sort of generative question than Lerdahl and Jackendoff are – leading to the more ‘bottom-up’ generative approach of Schenkerian theory, compared to the ‘top-down’ parsing approach of Lerdahl and Jackendoff – which is point that several thinkers have noticed, as we saw in chapter 1.1. In this generative context, rhythm just results from the “activity of tonal events”, as Carl Schachter put it, which affects the kinds of durational interpretations one can make of musical surfaces. And Lerdahl and Jackendoff do not necessarily disagree with this, but their awareness of this is only implicit in the formulation of their MPRs, not in the explicit focus of their theorizing. So, it is the perceptual emphasis of their work that makes them over-emphasize the distinction between grouping and meter, and that gives their theory its anti-grammatical attitude. Moreover, the fact that they concern themselves with assigning structure to musical surfaces that have undergone phrase expansion (see their examples 2.22 and 2.23 (Lerdahl and Jackendoff (1983): 26)), makes Lerdahl and Jackendoff focus on a level of musical structure where phrase groups and meter do not align anymore. This is in spite of the fact that grouping and meter are often aligned at deeper levels of structure, which can be seen after the rhythmic surface has been normalized.

To sum up, Lerdahl and Jackendoff’s unwillingness to commit to a Hypothesis A- or B-type understanding of the relation between pitch and rhythm seems to arise from their focus on durational *interpretation* (i.e. how listeners interpret the rhythmic/metrical structure of a perceived musical surface) over durational *interpretability* (i.e. how the rhythmic/metrical structure of a surface can be interpreted to begin with, based on the ways in which it is generated by pitch grammar). But a focus on durational

interpretability is crucial because this tells us why composers write certain kinds of musical structures and not others. This itself is a critical component of a generative theory of music. Since Lerdahl and Jackendoff do not focus on these aspects of music, their theory does not deal with the kind of generative concerns that Schenkerian theory does.¹²

We have explored the notions of Logical Form (LF) and Phonetic Form (PF) in Minimalist linguistics earlier. For the rest of this section I will illustrate some connections between musical rhythm and the Minimalist notion of PF, since the links between pitch grammar and durational interpretation we have discussed so far have some fascinating similarities to the analogous relationship between grammar and phonetic interpretation in language. I believe these similarities provide further evidence against Lerdahl and Jackendoff's critique of Schenkerian approaches to rhythm, and also suggest a kinship between Schenkerian music theory and Chomskyan linguistics on rhythmic grounds – both of which I will examine in due course.

I have suggested that to be interpretable pitch grammar has to map to rhythm, i.e. composers have to write musical structures that align with rhythm, and especially meter. A different way of saying this is that for a phrase to be interpretable, grammar must produce a structure that meets well formedness conditions imposed by rhythmic structure. We have seen how the binary-branching nature of pitch grammar manages this, by generating structures that align with the evenly spaced beats of meter, required by principles of metrical well formedness.

Now, we have seen that similar ideas about interpretability exist in the Minimalist Program's views on language. That is, linguistic surface structures can only be interpreted if they meet certain well

¹² Lerdahl and Jackendoff also argue that phrase structure is not restricted to any normative length. Specifically, "There is no prior restriction that the group extend between beats at the same metrical level. A group can have any arbitrary length" (Lerdahl and Jackendoff (1983): 29). Of course, groups can be of any length, but they *tend to be* of specific, and especially quadratic, lengths because of the requirement that (binary) pitch grammar map to rhythm according to principles of rhythmic/metrical well formedness – which is also why groups tend to align with the meter (i.e. they tend to extend between beats at the same metrical level). Therefore, phrase groups normally do *not* tend to be of any arbitrary length, unless one is dealing with very specific cases such as those of phrase expansion. Ultimately, this is also what decides which kinds of groups are interpretable.

formedness conditions too, imposed on linguistic grammar by semantics and phonology. The inability of grammar to generate structures that meet such conditions results in grammar not generating legitimate Logical and Phonetic Form representations of these structures, which are, respectively, the semantic and phonetic representations of these structures that can be interpreted by the semantic (i.e. conceptual-intentional) and phonological (i.e. sensorimotor) systems. And this leads to the semantic and phonetic uninterpretability of those structures, and a crash in the generative process.

We explored how the above issue pertains to a musical equivalent of Logical Form (or LF) in the last chapter. So, let us now see what the above implies for a musical equivalent of Phonetic Form (or PF), given our preceding discussions about the relation of pitch and rhythm in music. A PF representation, as we saw in chapters 1.1 and 1.2, is a surface (or more accurately, S-) structure whose non-phonological features have been ‘siphoned out’. In other words, a PF representation is a purely phonetic representation of an S-structure, which can therefore receive a phonetic interpretation by the sensorimotor system, i.e. it can be pronounced by a speaker – although this is subject, as discussed earlier, to a variety of performance factors (Chomsky (1995a): 2, Uriagereka (1998): 91).

But a legitimate PF structure will arise only if grammar maps to phonology correctly. As we saw for example in chapter 1.2, a noun phrase that is not checked for Case will block an otherwise successful S-structure from mapping to phonology (Chomsky (1981): 49). The linguists Jonah Katz and David Pesetsky illustrate this with the following sentences:

(1a) The Romans destroyed **the city**.

(2a) The Romans’ destruction **of the city** was a tragedy.

(2b) *The Romans’ destruction **the city** was a tragedy.

For the sentence in (2b) to map to phonology, the noun phrase “the city” inside it would have to be generated in a Case marked position, i.e. its position in the sentence would have to be preceded by the Case marker “of”. Since “of” is missing in (2b), the sentence cannot map to phonology and is therefore phonetically uninterpretable. As Katz and Pesetsky put it, the inability of (2b) to map grammar to

phonology creates the percept of deviance in the sentence, as the “unavailability of an otherwise expected pronunciation” (Katz and Pesetsky (2011): 11).

Now, for grammar to map to phonology successfully, their respective hierarchical tree structures have to align, so that information from one can be ‘read off’ to the other. This very fact immediately suggests a striking parallel with music, since we have seen that for a musical passage to be durationally interpretable, pitch grammar has to map to metrical structure – and for this to happen both structures need to align as well. (By alignment here, I do not mean a literal *temporal* alignment between pitch groups and meter, though this undoubtedly happens too in the idealized situations I have described above. Rather, I am thinking about the abstract alignment of the formal structures of pitch and meter in music, which is exactly what happens at the level of tonal rhythm, and which shows how groups that are misaligned or out-of-phase with the temporal metrical surface (perhaps due to phrase expansion) are perfectly aligned at a deeper, abstract level of structure, where the rhythm of expansions etc. has been normalized.)

It is important to note at this point that the above parallel between linguistic and musical structure does not imply that *meter* is the musical equivalent of phonology in language. Phonology deals with many things, only one of which is linguistic rhythm. Only the subset of phonology that concerns itself with rhythm, viz. prosody, is the appropriate locus for the analogy between linguistic and musical rhythm – so, a musical equivalent of a PF structure will only be analogous to a linguistic S-structure’s *prosodic* representation. Moreover, *musical* rhythm is not always metrical, i.e. it is not always organized in periodic cycles. We have been studying musical rhythm in its metrical guise here only because this is the primary way in which musical rhythm has been studied in Western music theory, especially in Schenkerian contexts. However, treating meter as the driving force behind the music-language analogy is problematic, since meter in speech rhythm is a highly problematic issue (Patel (2007): 96 and 159). The ideal analogy between music and language would therefore be between linguistic prosody and musical rhythm in general (whether metered or not). This point will be relevant to the subsequent discussion.

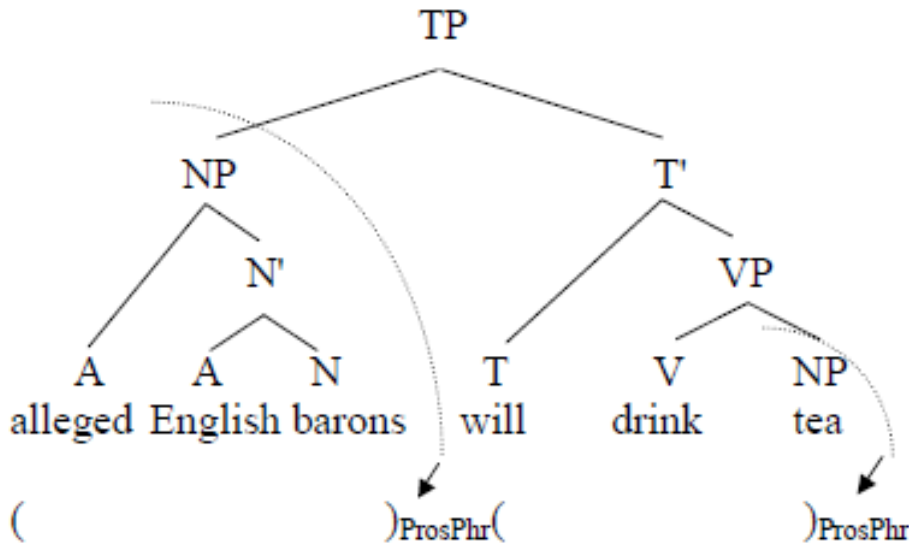
A current view in linguistic theory is that both (linguistic) syntactic structures and prosodic structures have the form of headed, hierarchical trees, just as musical pitch structures do, as we saw in

Example 2.2-1. Now, hierarchical prosodic structures are not identical to syntactic structures but, importantly, prosodic structures are usually *aligned* with their corresponding syntactic structures because the edges of certain prosodic categories match up with the edges of certain syntactic categories (Katz and Pesetsky (2011): 29). Specifically, the stress-bearing linearly rightmost edge (which is an example of a “phase marker”) of certain syntactic structures, such as most noun phrases, lines up with the edges of prosodic phrases. This gives rise to the phenomenon of “prominence”, in which a certain syntactic item is stressed prosodically. Katz and Pesetsky give the example of the sentence “alleged English barons will drink tea”, shown in Example 2.2-10, in which the rightmost edge of the noun phrase “alleged English barons” lines up with a prosodic boundary, since the last item (the noun “baron”) is the most prominent prosodic item in the phrase too.¹³

Katz and Pesetsky give an interesting argument for why grammatical and prosodic structures are aligned in language, and they use their argument to suggest that the same happens in music. According to them, the manner in which a grammatical structure is generated, i.e. the manner in which grammatical items branch off of each other in a tree structure, is constrained in such a way that it reduces conflicts between grammatical structure and prosodic structure – thus allowing the two systems to map on to each other. The primary way in which grammar is constrained is through what they call Region Conditions for language and music (Katz and Pesetsky (2011): 30-35). When a syntactic structure is generated in language or music, the Region Conditions restrict the ways in which words or musical

¹³ An interesting point to note here is that grammatical phrases and prosodic phrases line up on the *rightmost* edge of the grammatical phrase, which seems to be the opposite of the case in music, where Hypothesis A discussed earlier seems to be true, according to the views expressed in this chapter. That is, contrary to language, grammatical structures in music (i.e. pitch groups) seem to align with meter on the left edge of the structures, i.e. at the beginnings of groups. If musical and linguistic rhythm are indeed analogous, this could either mean that this difference is a parametric difference between linguistic and musical structure, or that some version of Hypothesis B (in which groups and metrical structure align on the right edge of those groups, as in language – a view defended by the music theorist Hugo Riemann and more recently by Cooper and Meyer) has greater validity than Hypothesis A. This is an interesting question for future research. (Although, Lerdahl and Jackendoff (1983): 321, report that in some languages prominence does align with the left edge of the structure, and also that some musical idioms display a Hypothesis-B type alignment of pitch and metrical accent, such as Javanese gamelan music. They cite the phonologist Elizabeth Selkirk’s assertion that the directionality of stress at least in English might be a matter of preference rather than necessity (Lerdahl and Jackendoff (1983): 343).)

Example 2.2-10. Grammar-to-prosody mapping in English from Katz & Pesetsky (2011)



pitches can be combined to generate that structure. So, to understand how the Region Conditions work, we have to understand how syntactic structures are generated first.

As we have seen several times in this dissertation, words and musical pitches combine to generate a syntactic structure by being merged into headed, binary trees. So, the words “the” and “cake” are merged to form a tree-like syntactic structure, specifically the noun phrase, “the cake” with both “the” and “cake” as its branches, and with “cake” being designated as the head of the tree – which is why the phrase is designated a noun phrase, not a determiner phrase.¹⁴ The noun phrase “the cake” can be merged further with other words to form larger tree structures with different heads. For example, it could be merged with the verb “ate” to generate the phrase “ate the cake”. “Ate the cake” is designated a verb phrase now because the verb “ate” is the head of the new phrase, instead of “cake”.

¹⁴ Chapter 1.2 explored the idea proposed by some linguists that noun phrases *are* in fact determiner phrases, with the determiner, and not the noun, as the head of the phrase. This proposal was made to fix an inconsistency within X-bar theory. However, the proposal has problems of its own, so I will ignore it in the present discussion on rhythm, and I will continue to treat structures like “the cake” as noun phrases rather than determiner phrases.

To quickly recap, words and the phrases generated from them can be combined with each other to generate full sentences, the totality of which can be represented as a large tree made up of different branches. Importantly though, all the branches of the tree occur in pairs to reflect the fact that they represent binary sets of syntactic items that are merged to generate part or all of the sentence that is represented by the tree. Each such pair of branches also has a head, such as “cake” or “ate” that determines how syntactic items are merged – so that a verb like “ate” can merge with “the cake” because the head of “the cake” is “cake” and not “the”. For this reason, the generation of a sentence follows a very deterministic path, based on which syntactic items can be combined with which other items. Finally, the ways in which syntactic items can be combined is based on the *hierarchy* that exists among these items. “Cake” is ‘higher’ in the hierarchy than “the”, which is why it is designated the head of the noun phrase “the cake”, and which is why it is possible for the verb “ate” to combine with this noun phrase – or even directly with “cake” (e.g. in “They ate cake for dessert”) in a manner that it cannot do directly with “the”.

This hierarchy is crucial for constraining how syntactic structures are generated. Specifically, if a syntactic unit Y falls in between two other syntactic units X and W in the hierarchy, X and W cannot legitimately merge in the grammar. “Cake” lies in between “ate” and “the” in the hierarchy of these words, with “ate” being the head of all three when they combine, or the head of “ate cake” when just those two words combine. Also, “cake” is the head of “the cake” but not of “ate the cake”, and finally “the” is not the head of anything at all. Therefore, “the” cannot directly combine with “ate” prior to combining with “cake” first, in the manner described above. So, this is a factor in constraining syntactic structure. To put this in Katz and Pesetsky’s terminology, “the” cannot “linearly intervene” between “ate” and “cake” before combining with “cake” first, i.e. the only way it can be added to the linear string “ate the cake”, is by merging with “cake” to form the noun phrase “the cake”. Which implies that if three words or phrases X, Y and W (ordered higher to lower in a hierarchy) are to merge to form a larger phrase or complete sentence, the hierarchically-lowest item W must combine with the middle entry Y before it can combine with X. (Alternatively, X can only merge with W after W has merged with Y first.)

Let us look at this another way. There are three ways of combining the items X, Y and W in a binary-branching tree approach. We can:

- (1) Merge (YW) first, assign either Y or W as the head and then merge this entire unit with X.
- (2) Merge (XY) first, assign either X or Y as the head and then merge this entire unit with W.
- (3) Merge (XW) first, assign either X or W as the head and then merge this entire unit with Y.

(3) is impossible because W must combine with Y first before it combines with X. However, (1) and (2) are both acceptable for reasons discussed above. (1) is just a formal way of describing how we generated “ate the cake” earlier, i.e. by first generating “the cake” and then merging this whole unit with “ate”. (2) is acceptable because it is an example of linear intervention. But (2) will only be a legitimate linear intervention if W merges with the Y branch of the (XY) pair, i.e. if it ‘rewrites’ (XY) as (X(YW)) – it will not be acceptable if W tries to merge with X prior to combining with Y.

(2) shows us the way in which one of the items among X, Y and W can linearly intervene between the other two. But this is also the *only* way in which a linear intervention can occur between the three items because (1), (2) and (3) exhaust the possibilities of how X, Y and W can be merged to form a syntactic structure, and only (2) gives an instance of a legitimate linear intervention. This brings us to an important fact about how items merge to generate syntactic structures – if a syntactic structure is generated from three syntactic items by one of the items linearly intervening between the other two, the intervening item *has to be the hierarchically lowest item in the group*. This is because the only legitimate intervention in (1), (2) and (3) involves W (which is hierarchically lower than both X and Y) intervening between the two other items. X and Y cannot linearly intervene in any legitimate way at all.

This fact about syntactic generation therefore constrains how a syntactic structure can be generated because it restricts the ways in which syntactic items can be merged in the generation of the structure. The constraint, as we have seen, works by determining which specific items in a group another item can legitimately combine with to generate the structure. So, in a way the constraint determines the

combinatorial *region* for that item. Specifically, the combinatorial region for W in the above example is restricted to Y, since W cannot directly merge with X.

Now imagine that the syntactic units X, Y and W involved in grammatical generation also encode prosodic information, e.g. if they happen to be *phase markers*, which are syntactic units that also contain prosody-related stress information. Prosodic theory in linguistics holds that when two phase markers are merged to form a prosodic tree structure, a third item can linearly intervene between them only if it is less prosodically prominent than the other two, i.e. if it is less stressed. But we have already seen that an item can linearly intervene between two other items in a *syntactic* tree only if it is lower than the two other items in the syntactic hierarchy of those items as well. In other words, there is a direct correspondence between how syntactic trees and prosodic trees are generated, when items that generate syntactic trees also encode prosodic information, because of the shared constraints on such generation in both syntax and phonology. To consider this in the light of Example 2.2-10, in the phrase “alleged English barons”, “English” is the lowest item in the syntactic hierarchy – so it cannot linearly intervene between “alleged” and “barons” prior to combining with “barons”. But as we have also seen, “English” is the least prosodically prominent item in the phrase too, since prosodic prominence is generally found in the linearly rightmost stressed item in the phrase. Thus, there is a wonderful mapping between grammar and prosody in language.

This mapping is nothing but Katz and Pesetsky’s Region Condition for language at work.

According to that condition:

“For every pair of distinct events (α, β) , such that α and β have property P, if e linearly intervenes between α and β , both α and β exceed e in prominence.” (Katz and Pesetsky (2011): 30)

One example of property P, as we have seen, is that of being a phase marker. So, if α and β are phase markers, e can only linearly intervene between them (syntactically) if it is hierarchically lower than both α and β , and (prosodically) if it is lower in prominence than α and β – thus ensuring a perfect mapping of grammar to prosody at the level of PF.

Now, we know from Example 2.2-1 that a basic region constraint on syntactic generation exists for music too. That is, a pitch event in musical grammar can only combine with other pitch events in its combinatorial region. So, the dominant chord in level B of Example 2.2-1 cannot combine with the initial tonic chord of that example before combining with the final, cadential tonic chord first – it cannot linearly intervene between the two tonic chords prior to merging with the final tonic. (As another example, a supertonic chord cannot normally linearly intervene between a tonic and a dominant chord in Western tonal music, without first merging with the dominant chord it prolongs – which is why supertonic chords are often referred to as “predominants” in tonal music.) Is there a region constraint on rhythmic structure in music too, analogous to prosodic structure in language, thus allowing for a mapping between grammar and rhythm in music just as in language – and therefore justifying a musical version of PF?

The answer to this question lies in whether there is a Region Condition for music too, and Katz and Pesetsky argue that there is such a condition, thus defending the identity of music and language in rhythmic contexts. I will now quickly review their argument, and then discuss what I believe to be a major problem with it, which can potentially be resolved under the more Schenkerian approach to rhythm being explored in this chapter.

Basically, Katz and Pesetsky support a Region Condition for music by equating linguistic prosody with Lerdahl and Jackendoff’s Time Span Reduction (TSR), and then arguing that TSR maps to musical syntax just as prosody maps to linguistic syntax, under region-based constraints of structure generation. (Katz and Pesetsky’s adoption of ideas from Lerdahl and Jackendoff’s theory here builds on the latter’s own recognition that there is a strong parallel between TSR and linguistic prosody (Lerdahl and Jackendoff (1983): 314-328).) TSR, as we have seen earlier, is a tree-diagrammatic representation of musical structure that considers rhythmic, in addition to pitch, information, to assign a structure to a given musical passage. So, TSR seems to encode *prima facie* for music the kind of information that a prosodic representation of structure does for language. Moreover, Katz and Pesetsky argue that the manner in

which TSR assigns structure to a passage is nearly identical to the way musical grammar generates a surface, thus allowing for a mapping of TSR to syntax.

However, the way TSR assigns structure to a passage is interesting. First of all, TSR does not really *generate* a musical passage because of Lerdahl and Jackendoff's listening-based approach to grammar. Rather, TSR assigns a structure to a *heard* passage that has already been generated. This in itself is not a problem because all that really matters in the time span reduction of a musical passage is the issue of which musical events in the passage are allowed to branch off of which other events in the tree representation of the passage – and these branching relationships remain the same whether one is generating the passage bottom-up, or assigning structure to the generated passage top-down while listening to it. Katz and Pesetsky formalize these relationships in terms of the notion of “root distance” (RD), viz. the distance of a musical event's highest projection in the tree structure (i.e. the highest projection at which it forms the head of a binary pair) from the top (i.e. “root”)¹⁵ of the tree. (For example, in the phrase “ate the cake”, the highest projection of “cake” occurs in the tree not when it is the branching sister of “the”, but when it is the head of the whole noun phrase “the cake”.) So, the RD value of an event in a TSR diagram negatively correlates with the event's hierarchical status in the TSR tree – in other words the lower the RD value, the higher the event's highest (or “maximal”) projection is to the top of the tree, the *closer* it is (in *Root Distance* terms) to the top of tree. Also, since TSR measures both rhythmic and pitch information, a lower RD number means that an event is higher up in the rhythmic hierarchy of the passage too, implying that it is more rhythmically stressed – or more prominent, to make an analogy with prosodic structure in language.

But given Lerdahl and Jackendoff's insistence on the importance of rhythm in determining the grammatical structure of a musical passage, they proceed to derive the (pitch-based) grammatical structure of a passage (known as the “prolongational reduction” (PR) of a passage) *from* TSR. The determination of a passage's grammatical structure depends, as we have seen before, on the relative

¹⁵ Revisiting a point made in chapter 1.2, notice how the top of the tree is actually the ‘root’ of its structural description.

hierarchical importance of different pitch events, as illustrated in Example 2.2-1. These hierarchical pitch relations can be measured in terms of the relative RD values of pitch events. But since these RD values are initially determined in a TSR description of a passage in Lerdahl and Jackendoff's approach, the prolongational structure of a passage will therefore be determined by TSR-derived RD values – which in Katz and Pesetsky's opinion strongly couples pitch-based grammatical structure to rhythm-based TSR structure. In other words, the higher a pitch event is in the grammatical hierarchy of a passage (as seen in a prolongational reduction, and as determined by RD value), the more rhythmically prominent it will be too, since RD value also determines prominence in a TSR tree. This, in essence, leads to a Region Condition for music, paralleling the prosody-grammar relation in language. Katz and Pesetsky state this as follows:

“For every pair of distinct PR events (α , β), such that α and β are the heads of sisters ... if an event e linearly intervenes between α and β , both α and β exceed e in prosodic prominence.” (Katz and Pesetsky (2011): 31)

In other words, if an event e is lower in the syntactic hierarchy of a passage (which is the only way it can linearly intervene between two hierarchically-superior events, in this case α and β), it is *also* lower in the rhythmic hierarchy of the passage, i.e. it has lower prominence.¹⁶ In this manner, Katz and Pesetsky argue for an analogous mapping between rhythm and grammar in music and language, and thus for the identity of the two systems in this domain.

There is one major problem with this argument, which has to do with how Lerdahl and Jackendoff map rhythm *to* pitch (i.e. from TSR to PR), which is the opposite of the grammar to prosody mapping in language. Katz and Pesetsky do not think the opposite directionality of the TSR-PR model is a problem for asserting the Region Conditions for language *and* music, and therefore their identity, because for them

¹⁶ Notice that the property P shared by α and β in the Region Conditions for language and music is the property of being a phase marker in language, and of being heads of sisters in a musical tree diagram. There is a lack of isomorphism in the way the Conditions have been formulated for the two domains; whereas property P for language is an actual syntactic structure (viz. a phase), property P for music is formulated more generally, and not in terms of its being an actual musical syntactic structure, e.g. a chord. This non-isomorphism leads to problems for an assertion Katz and Pesetsky make about lexicons, which I deal with in footnote 18.

all that seems to matter in the TSR-PR relation is the notion of RD, which stays the same no matter which direction the mapping takes, i.e. whether from TSR to PR or PR to TSR. But does the mapping between TSR and PR just boil down to the RD relationships among the musical events within them?

I believe the answer to this is “no” for an important reason. Let us look at this in the context of the musical example that both Lerdahl and Jackendoff, and Katz and Pesetsky, use to make their points on this matter, viz. the 4-bar antecedent phrase from the main theme of Mozart’s K. 331 piano sonata, shown in Example 2.2-11.

Example 2.2-11. Mozart, Piano Sonata #11, K. 331/i: Main theme, mm. 1-4

The argument of both authors depends crucially on the relative structural importance of the vi^7 F#-minor harmony in the first half, and the V^6 E-major harmony in the second half, of measure 3 of the example – which have been labeled P and Q respectively. Both authors believe these two harmonies reveal a conflict between metrical accent and pitch structure in this measure. This is revealed by the relative merits of the two reductions of this passage shown in Example 2.2-12 (adapted from Katz and Pesetsky (2011): 24). Example 2.2-12A maintains the Q harmony of measure 3, while reducing out the P, whereas Example 2.2-12B keeps the P and reduces out the Q. The reductions reveal which harmony is considered structurally more important in that measure; so keeping the Q harmony in the reduction, as in Example 2.2-12A, implies that this chord is structurally more important – therefore, it would have a smaller RD value and would come higher up in the hierarchy of the tree diagram of this passage.

Following Lerdahl and Jackendoff, Katz and Pesetsky think that Example 2.2-12A is actually a bad reduction of the passage, but only in TSR terms. That is, when one takes rhythmic (and specifically metrical) information into consideration, as one does in a TSR analysis, the P of measure 3 seems to be more structurally important, considering that it falls on the *metrically*-accented downbeat of the measure,

Example 2.2-12. Mozart, Piano Sonata #11, K. 331/i: Katz & Pesetsky's reduction of mm. 1-4
A: Bad reduction. B: Good reduction.

which itself parallels the treatment of the structurally-important tonic and dominant harmonies of the previous two measures. So, if one hears a harmonically reduced version of this passage, a reduction that omits the P (such as Example 2.2-12A) would sound odd, whereas one that keeps it in favor of the Q (such as Example 2.2-12B) would sound acceptable. But both authors argue that this preference obtains only when we take the metrical information of measure 3 into consideration; they maintain that on purely pitch-grammatical terms, Example 2.2-12A would actually be a better reduction because the Q chord has greater structural weight relative to the preceding predominant harmony, as traditional harmonic theory usually agrees (e.g. see Aldwell and Schachter (2011): 161-178).

Lerdahl and Jackendoff argue that the greater TSR acceptability of Example 2.2-12B versus the greater PR acceptability of Example 2.2-12A leads to a conflict between these aspects of musical structure – and essentially a conflict between metrical structure and pitch structure. They argue that the conflict is resolved in the mapping of TSR to PR, in which the greater TSR prominence of the P harmony is overruled in favor of the greater structural stability of the Q harmony in the PR analysis by an exception clause. This exception clause allows the less prominent Q to have greater structural weight in the PR analysis of the piece, only because doing so would lead to a more prolongationally stable analysis of the complete passage. In other words, the exception clause allows the Q to attach directly to the tonic harmony of measure 4 to give the passage a more prolongationally stable structure, even though doing so violates the Region Condition for music because Q, being less prominent than P, would not normally be allowed to directly attach to that tonic harmony without first merging with the P – it would not be allowed to linearly intervene between the more prominent P and the tonic of measure 4 without first attaching to the P harmony. This exception clause therefore resolves the conflict between grouping and meter in measure 4.

The problem here is that though one can defend mapping TSR to PR, in order to stress the importance of rhythm in determining pitch grammatical structure as Lerdahl and Jackendoff do, the opposite mapping of PR to TSR (to parallel what actually happens in language) cannot be so easily defended. If we were to map PR to TSR, we would start off with an analysis of the passage in which event Q has greater structural weight in PR, and we would have to derive the TSR analysis from this – which would involve promoting P over Q in the process by means of an exception clause, analogous to the one Lerdahl and Jackendoff propose for promoting Q over P in the opposite mapping of TSR to PR. But no such PR-to-TSR exception clause exists, making the mapping from PR to TSR problematic – and for good reason too. As we have discussed in some detail now, pitch grammar and rhythm have an asymmetrical relationship – so, if a pitch structure is treated as having greater structural importance relative to another during grammatical generation, it will normally receive a stronger rhythmic interpretation than the other structure too. And the *opposite* influence, i.e. of rhythm over pitch, does *not*

normally happen – which is why an exception clause for *promoting* a rhythmically prominent event over a tonally important one is hard to defend. (Of course, weaker pitch structures are often given strong rhythmic interpretations, as happens in metrically unaccented cadences. But this just falls out from the usual way in which musical grammar maps to rhythm – not from the workings of an exception clause.) Therefore, the conflict between Example 2.2-12A and B would not be resolved and PR would not successfully map to TSR.

In other words, reversing the mapping from TSR-to-PR to PR-to-TSR is not as trivial as Katz and Pesetsky seem to believe. And since the PR-to-TSR mapping is the better parallel in music for what happens in language, the difficulties inherent in justifying this mapping, as argued above, destroy the possibility for the identity of this mapping in both music and language, contrary to what Katz and Pesetsky try to demonstrate.¹⁷ Note that this is really a problem only for a TSR-based system, like Lerdahl and Jackendoff's, which requires a grammatical description of a musical passage to have both a pitch analysis and a rhythmic analysis. It is not an issue for a primarily pitch-based theory, such as Schenkerian theory, in which assigning a rhythmic analysis to a passage is not a central goal of grammatical description – or rather, in which a rhythmic analysis falls out from the usual mapping of grammar to rhythm.

A TSR-based approach for describing the mapping between pitch and rhythm is even more problematic given some of the complications inherent in its very definition as a musical analogue of

¹⁷ There is a caveat to this argument though, which is that in the Mozart passage under discussion here Q linearly intervenes between P and the following tonic chord, but P does not linearly intervene between Q and the tonic, since it appears to the left of both these chords in the passage. Therefore, a region condition that governs the mapping of PR to TSR (and the promotion of P over Q in the process, which only an exception clause can allow), does not apply to this passage in the first place because P does not linearly intervene between the relevant α and β here. This renders my critique of Katz and Pesetsky's argument moot. However, we can easily find cases in music where a rhythmically prominent structure like P *does* intervene between two adjoining, tonally stable structures. Take suspensions for instance, in which the dissonant, and therefore tonally weak, suspension occurs between two consonant, and therefore tonally strong, pitch structures, called the "preparation" and the "resolution". According to Katz and Pesetsky's Region Condition, for a suspension to linearly intervene between its preparation and resolution, it has to be lower in (rhythmic) prominence than those two pitch structures. But this is the *opposite* of what happens in canonical suspensions, which are by definition metrically strong events (Salzer and Schachter (1969): 78). Therefore, for a suspension to be given its proper, rhythmically prominent, analysis in Katz and Pesetsky's model, an exception clause has to exist, which would allow for the suspension to be given a more prominent analysis than the standard mapping from PR to TSR would allow. But as I just argued above, the asymmetry between pitch and rhythm makes it hard to defend such an exception clause.

prosody in language – a definition to which both Lerdahl and Jackendoff, and Katz and Pesetsky, subscribe, as we have seen above. This is because, rather than being the rhythmic component in the mapping from pitch to rhythm (as the musical analogue of prosody), TSR, in my opinion, is *already* a mapping between pitch and rhythm in itself. As we have noted before, TSR is meant to inform the structural analysis of a musical passage in *both* pitch and rhythmic terms, given Lerdahl and Jackendoff's unwillingness to perform such an analysis from a purely pitch-based perspective. So, in order to see how pitch maps to rhythm, all we have to do is see how this happens *within* TSR structure – not in how TSR maps *to* PR. Moreover, if there is an asymmetry in this mapping, *from* pitch *to* rhythm, in TSR, then this would also reveal a clear parallel with how this mapping takes place in language. So if we want to compare how grammar maps to rhythm in music *and* language, as Katz and Pesetsky wish to do, we have to look no further than at what happens within TSR itself and, again, not in how it maps to PR.

To understand this, consider the fact that the actual branching structure of TSR is *different* from the hierarchical branching structure of pitch in a significant way. As Lerdahl and Jackendoff say, TSR structure is a *layered hierarchy*; i.e. a structure with different levels of groups (specifically *subgroups*, *uncadenced groups*, and *cadenced groups*) ordered from smallest to largest, such that there is no interaction among them. This means that a larger group, e.g. a cadenced group, can contain a smaller group, e.g. a subgroup, but cannot itself be contained within that smaller group. This layered structure makes TSR different from hierarchical pitch structure because in the latter a structure that occurs higher up in the hierarchy can occur at a lower level of structure too – a dominant-tonic progression can occur within a larger dominant-tonic progression.

But the higher layers of TSR are strictly binary branching (Lerdahl and Jackendoff (1983): 327). This is surprising because the structure of linguistic prosody (which TSR is supposed to be the musical analogue of) is generally *not* considered to be binary branching. Instead, linguistic prosody is usually described in terms of tree structures that can have any number of branches. (In other words, prosodic structure is *n*-ary branching (Katz and Pesetsky (2011): 32-34).) In contrast, it is the hierarchical structure of grammar (whether linguistic or musical) that is binary branching, as we have seen earlier in

this chapter. So, how can TSR be both an analogue of prosody then, and also binary branching (at least in its higher layers)?

This is why I believe that TSR is not really the analogue of linguistic prosody, but rather a mapping between pitch and rhythm in itself – which implies that it will have both pitch and rhythmic components in it. Along these lines, it is the non-rhythmic *pitch-grammatical* component that leads to the binarity of the higher layers of TSR trees. If we take this out, we are left with the rhythmic component – and this will deprive TSR of its binary-branching structure too, since rhythmic (and specifically metrical) structure in music is not usually described in binary-branching terms, but rather in terms of a metrical grid with one head (the downbeat) and any number of other branches, corresponding to the number of beats in that grid. And as we have seen earlier, there can be virtually any number of beats in a measure too. So, a purely rhythmic TSR structure, i.e. one deprived of pitch information, could very well be an *n*-nary branching structure. This, in fact, is exactly what one sees in the lower, subgroup, layer in TSR too, which has an *n*-nary branching structure – primarily because this layer is determined by metrical structure alone, as Lerdahl and Jackendoff state themselves (Lerdahl and Jackendoff (1983): 317) – unlike the higher layers, which seem to reveal the influence of pitch structure as well.

In turn, this makes TSR a true analogue of linguistic prosody, given prosody's *n*-nary structure, *but only* when we conceive of TSR in terms of its (*n*-nary) rhythmic components. This is not surprising, because it is the (metered or unmetered) rhythmic component of music (divorced from pitch) that, after all, is the true analogue of prosody, which I took to be a working assumption earlier in this chapter. However, the internal structure of TSR, when seen in the above terms, also reveals the *asymmetry* between pitch and rhythm in music, i.e. the fact that pitch maps to rhythm and not vice-versa. This is because TSR, as formulated, is not made up of only rhythmic components, but pitch components as well – and the larger structure of TSR is governed by its *pitch*, and not its rhythmic, components, since TSR is a *hierarchical* structure (albeit a layered one) in which the *higher* layers are influenced by binary-branching *pitch* phenomena. (Which therefore influence the entire TSR structure taken as a whole.)

In consequence, we can see how TSR is separate from, but still reveals the influence of, pitch grammar – which is exactly the kind of mapping one sees between grammar and prosody in language. So, the mapping of pitch to rhythm within TSR might be the place to look for a musical analogue to the mapping of grammar to prosody in language. However, neither Lerdahl and Jackendoff, nor Katz and Pesetsky, attempt to do this. Instead, Lerdahl and Jackendoff go to the extent of trying to derive pitch grammar (in the PR component of their theory) *from* TSR – which seems to confuse the whole matter altogether. This makes the very notion of TSR problematic, since it is unclear whether it represents a musical analogue of linguistic prosody (as Lerdahl and Jackendoff see it, and which might justify their attempt to derive PR from it), or whether it represents a musical analogue of the *mapping* between grammar and prosody (as I see it, and which problematizes the attempt to further derive PR from it). This suggests a major flaw with Lerdahl and Jackendoff’s description of how pitch and rhythm relate in music – or more specifically, a flaw with Katz and Pesetsky’s TSR-based description of how grammar and rhythm map to each other in both music and language.

But to my mind, this also suggests why a Schenkerian approach to rhythm – the subject of this chapter – ultimately seems to be the best way to understand the mapping between grammar and rhythm in music (and why Lerdahl and Jackendoff’s critique of this approach ultimately seems unwarranted). This is because Schenkerian theory clearly shows us how pitch grammar maps to musical rhythm in an asymmetrical way; a description that is implicit in TSR structure but never explicitly acknowledged by either Lerdahl and Jackendoff, or Katz and Pesetsky – and which is ultimately rejected in the attempt to derive PR from TSR. Therefore, only Schenkerian theory seems to bring the true relation between pitch and rhythm into relief, whereas a more TSR-based approach just obscures it.

To understand this further, let us revisit the TSR explanation for why the reduction in Example 2.2-12B is preferable over 2.2-12A. Katz and Pesetsky, following Lerdahl and Jackendoff, argue that the P harmony that is kept in the reduction in Example 2.2-12B is kept for metrical reasons, primarily the fact that it occurs on the downbeat. It is also retained because it realizes a parallelism with the harmonic events in mm. 1-2. But are these arguments good enough? That is, would the argument for P’s importance

disappear if it were not on the downbeat of measure 3? Exchanging the positions of the P and Q harmonies in measure 3 easily shows that this is not the case – P would still be the preferred harmony in a reduction of this passage, even if it fell on the second, weaker half of measure 3 and Q occurred on the downbeat instead. So, the structural importance of P does not necessarily have anything to do with its rhythmic prominence. (Katz and Pesetsky themselves show how this is the case for the pitch events occurring in the first two measures of the passage, i.e. how exchanging the positions of pitch events in those measures so that events previously on weak beats now fall on strong beats, and strong on weak, does not change our judgments regarding the structural importance of those events.)

For this reason, I think the preference for P holds even on pitch grounds, so that Example 2.2-12B is the better reduction not merely for TSR reasons but *also* for prolongational reasons. (Or, in light of my above critique of the definition of TSR, for reasons that have to do with the, under-acknowledged, pitch component of TSR.) Let us examine Heinrich Schenker's own reduction of this passage, shown in Example 2.2-13, to see why this might be the case. (I have adapted his reduction here from Allan Keiler's discussion of it (Keiler (1983-84): 204).) First of all, notice how Schenker labels the levels of his reduction from (c) to (a), suggesting that the direction of analysis here is from bottom to top – which is very much a generative approach to the passage, as opposed to Lerdahl and Jackendoff's reductive perceptual approach. Now, consider measure 3 in level (b) of Schenker's reduction. Notice here how the dissonant seventh interval between the bass F#3 and the tenor E4, which is present on the downbeat of the measure in the surface level (a) of the reduction, has been removed from level (b), turning the previously dissonant vi^7 into the more stable vi chord. Moreover, the slurs in this measure show how Schenker conceives of the apparently more (prolongationally) stable V^6 as really a *dissonant passing chord* between the vi and the tonic harmony on the first beat of measure 4. Finally, consider that Schenker takes this whole slurred measure to be a prolongation of the tonic harmony at the beginning of measure 4, so that the harmony that really occurs on the downbeat of measure 3 at the deepest levels of structure is neither the P nor the Q of Examples 2.2-12A and B, but rather a tonic chord, which is displaced onto the downbeat of measure 4 by the left-branching prolonging work done by the P and Q harmonies.

What this implies is that Schenker's idea of a good reduction of this passage is neither Example 2.2-12A nor B, but rather level (a) of Example 2.2-13. According to this view, the first three measures of the passage prolong a tonic harmony that harmonizes a *Kopfton* E5 all the way to the sixteenth-note figure of measure 4, at which point it descends to B4 through D5 and C#5, via a half-cadential gesture (which is

Example 2.2-13. Mozart, Piano Sonata #11, K. 331/i: Schenker's graph, mm. 1-4

The image displays Schenker's graph for the first four measures of Mozart's Piano Sonata #11, K. 331/i. It is organized into three levels:

- Level (c):** The original musical score for the first four measures. A large brace above the staff spans all four measures, labeled with a circled 5 ($\hat{5}$). A smaller brace above the staff spans the last two measures, labeled with circled 4, 3, and 2 ($\hat{4} \hat{3} \hat{2}$).
- Level (b):** A reduction of the score showing the prolongation of the initial tonic harmony. It features a single melodic line with a long slur over the first three measures, and a shorter slur over the last two measures.
- Level (a):** The Ursatz reduction, showing the fundamental structure. It consists of a single melodic line with a long slur over the first three measures and a shorter slur over the last two measures. Below the staff, Roman numerals are indicated: **A: I** under the first three measures, **II/IV** under the fourth measure, and **V** under the final two notes of the fourth measure.

illustrated by the level (c) graph at the top of the example). The prolongation of the initial tonic harmony takes place via the parallel tenths voice-leading progression between bass and soprano seen in the first three vertical sonorities of the level (a) reduction, viz. (written here in bass-soprano form): C#4-E5 – B3-D5 – A3-C#5. Each of these sonorities is itself prolonged by a left-branching progression, shown, as we

just saw, by the slurs within each measure in mm. 1-3 of the level (b) graph, and also by the brackets in each measure in mm. 1-3 of the level (c) graph.

Two points emerge from this Schenkerian perspective that are relevant to our discussion of the problematic third measure of the passage. First, we should not consider the P and Q harmonies in that measure as separate chords but rather as *linear*, left-branching sonorities that prolong the tonic harmony of measure 4. (This point just reinforces the argument I made in our discussion of Indian music in chapter 1.3, about confusing harmonic structure with the structure of vertical sonorities in a musical passage.) This is precisely why Schenker reduces out the dissonant tenor E4 pitch in the vi ‘chord’, to show how the vertical sonority vi⁷ in the surface of the piece, should not be taken as an actual dissonant seventh chord that resolves to the following V⁶, as Example 2.2-12A suggests. In this light, it is not clear how one would go about assigning TSR or PR values to P and Q – or specifically that P should have a lower PR value, but a higher TSR value, than Q.

The second point has to do with the P sonority of measure 3, the vi ‘chord’. From the linear perspective that I am advocating here, notice how the soprano pitch A4 of that sonority forms the interval A4-C#5 with the soprano of the tonic chord on the downbeat of measure 4, which is nothing but the major third interval of the A major triad. So, what the P sonority in measure 3 does is to linearly *compose out* (Schenker’s *Auskomponierung*) the A-major triad in measure 4. Moreover, this A4, when taken together with the bass F#3, forms a series of inner voice parallel tenths with the sonorities on the downbeats of mm. 1 and 2 – i.e. the progression: A3-C#5 – G#3-B4 – F#3-A4. So, a listener would prefer to hear this A4 in measure 3 *over* the following B4, even if we reversed the A4 and B4 because the A4 forms a *tonally* prominent, descending line with the previous measures.

With these two points we can now see why the P sonority on the downbeat of measure 3 has greater structural weight compared to the following Q on purely pitch-grammatical, prolongational grounds. So, if the reduction in Example 2.2-12B has any relative merit vis-à-vis 2.2-12A, it is *not* for metrical reasons – or any reason that necessarily defends a TSR view of that passage. This is why a TSR

analysis of this passage, as opposed to a Schenkerian one, just seems to obscure the relation between pitch and rhythm here.

Katz and Pesetsky, following Lerdahl and Jackendoff, also defend a TSR preference for Example 2.2-12B on grounds that measure 3 parallels mm. 1-2, and so should be analyzed in the same way. There is nothing obviously wrong with this argument, but it does betray the reductive, perceptual bias of Lerdahl and Jackendoff's approach. Parallelism is a very important cue for helping us perceive both musical and linguistic surfaces (see Frazier et al. (1984), and Temperley and Bartlette (2002), for a discussion of how parallel constructions facilitate parsing in language and music respectively). But parallel structures often cease to be quite so parallel when examined from a truly generative (and not parsing) point of view, since structures that seem to be parallel on the surface are often in very non-parallel, asymmetric relationships with each other in the hierarchical grammar of a musical or linguistic passage. In our Mozart example, the surface parallelism of mm. 1 and 2 obscures the fact that measure 2 *prolongs* measure 1 and is therefore in an asymmetric relationship with it, most likely as the sister branch of a binary pair in which measure 1 is the head. In contrast, if two musical passages or linguistic sentences really have parallel structures they might not possess any significant grammatical relationship, and might even belong to different grammatical trees (e.g. if they happen to be two different, yet similarly constructed, sentences).

So, in sum, a TSR-based approach just seems to obscure the true mapping of pitch to rhythm in music, by insisting on the importance of rhythmic phenomena in descriptions of this mapping. However, these rhythmic phenomena often seem to have a pitch-based explanation, given the asymmetry of pitch and rhythm in music – to which only a Schenkerian approach seems to do justice. Therefore, Katz and Pesetsky's reliance on a TSR-based approach rather than a Schenkerian one seems to do their argument for the identity of language and music more harm than good.¹⁸ In contrast, a Schenkerian approach might

¹⁸ Another problem with Katz and Pesetsky's hypothesis is their insistence on the absence of a lexicon in music. In general, I hold that the presence or absence of a musical lexicon is an open question, with good reason to believe

just be the way to justify such an identity, given that it *does* suggest a mapping of grammar to rhythm in music that parallels descriptions of this phenomenon in Chomskyan linguistics. But even if Schenkerian theory correctly describes the *direction* of mapping between grammar and rhythm in music in a way that

that music should have a lexicon – otherwise there would be no (lexical) inputs for musical grammar to operate on, and thus no point in having a musical grammar at all. However, Katz and Pesetsky clearly seem to believe that music cannot have a lexicon, and this insistence arises specifically in the matter of defining property P in the Region Conditions for language and music. As we saw earlier, phase markers are one example of property P in language that, in turn, illustrates how the Region Condition works.

But Katz and Pesetsky also state that property P could be defined as that of not being a clitic (i.e. an unstressable word that is the branching sister of a stressable word) (Katz and Pesetsky (2011): 39). In other words, if α and β have property P, then they are not clitics (i.e. they are stressable) – and, as per the Region Condition, if event e linearly intervenes between them, it will be lower in prominence than either. (In fact, e will often itself be a clitic, which means that it would not be stressable and would therefore be lower in prominence than α and β – which is what the Region Condition says anyway.) Now, if the Region Condition applies to *both* music and language, i.e. if music and language are identical, then we have a problem – viz. that the specific Region Condition for music would have to apply to language, and that for language to music. But why should a condition that maps grammar to prosody in language have to be sensitive to the Region Condition for music, which has to do with matters of musical pitch hierarchy and the like? Katz and Pesetsky’s answer has to do with the crucial *non-isomorphism* in the formulation of the Region Conditions for music and language, as noted in footnote 16. That is, in the case of music, the Region Condition does not actually deal with musical structures, such as hierarchically-organized pitches, but rather with the more general requirement that the musical events α and β be “heads of sisters”. This makes it easy to apply the Region Condition for music to language because two events can be heads of sisters in a linguistic tree too. In fact, Katz and Pesetsky even give an example of a sentence where a heads of sisters condition seems to be at work (see their footnote 34 on p. 41), though they conclude that the specific Region Condition for language normally overrules the heads of sisters condition. Katz and Pesetsky go on to argue that even though the Region Condition for music might apply to language, the Region Condition for language does not apply to music (or rather, is irrelevant to it) because music does not have a lexicon – so, it is insensitive to whether the inputs to its grammar contain phase markers or clitics.

This argument seems problematic to me. For one, if we make the Region Conditions for music and language isomorphic, which they should be if music and language are identical, then the Condition for music should be stated in terms of phenomena that play an analogous role in the generation of musical structure as phase markers or clitics do in language. So, defining the Region Condition for music in terms of “heads of sisters” will not do – it must refer to the chords with their specific pitch constitutions that allow for the “heads of sisters” condition to obtain for music in the first place. (After all, musical grammar works not because it is implicitly made up of headed hierarchies, but because it operates on pitch structures that have hierarchical relationships – which can then, in turn, be described with headed, hierarchical trees.) But if we redefine the Region Condition for music in such, pitch-structural, terms, there is no reason why it should apply to language. In that case the music Region Condition’s inapplicability to language would be similar to the language Region Condition’s inapplicability to music, the latter being something Katz and Pesetsky accept. Would that imply that *language* does not have a lexicon just because the Region Condition for music does not apply to it? Of course not (although see Halle and Marantz (1993)) – so, why should we deny a lexicon for music then? Such a lexicon would obviously be different from the linguistic lexicon (being made up of pitch structures and such), yet it would be crucially relevant for musical grammar in the same way that the linguistic lexicon is for linguistic grammar.

The problem here seems to be one of defining the word “lexicon” too narrowly, in terms of clitics and such, which preempts musical structures from being lexical. In doing so, Katz and Pesetsky are actually being inconsistent in their own labeling practice. This is because they do not restrict their definition of “grammatical tree” to language, as they do for “lexicon”, even though it has been asserted that language trees are *not* identical to musical trees because the former encode X-bar relationships and the latter tonal tension relationships (Jackendoff (2009): 200-202, Jackendoff and Lerdahl (2006): 59). So if Katz and Pesetsky are willing to overlook that distinction in favor of a unified description of musical and linguistic grammar, they should be willing to do so for a unified, ‘musicological’ description of “lexicon” too.

parallels language – and without referencing the problematic notion of TSR in the process – we still have not seen how this approach explains the actual mapping between grammar and rhythm itself, apart from the fact that they seem to align in certain ways. And we still have not seen whether this mapping takes place in a manner that is consistent with Katz and Pesetsky’s Region Condition that governs this mapping in linguistic descriptions of Phonetic Form. So, to this final goal we now turn.

2.2.4. Conclusion: Musical PF - The Mapping of Structural to Rhythmic Accent?

To begin this last section, I would like to deal briefly with an issue that complicates the mapping of grammar to rhythm in language and music. In language, grammar maps to prosodic rhythm, which may or may not be metered. Yet our musical discussion has unwaveringly focused on the mapping of grammar to musical *meter*. In this light, does it even make sense to talk about a musical equivalent of PF in language, i.e. a level of pitch-grammatical structure where only information relevant to the *rhythmic* structure of a musical surface is preserved?

To answer this question, consider the fact that meter is only a special case of rhythm. Only when rhythmic accents are placed in such a way that they are separated by equal spans of time and alternate between strong and weak accents does rhythm actually become meter (see footnote 3 in this regard). In other words, only when one rhythmic accent is followed by at least one more rhythmic accent, which helps set up a periodic division of time, does meter arise. This is the standard definition of meter in the music-theoretic literature – viz. a phenomenon that involves equal time spans ‘book ended’ by a strong beat that alternates with a weak beat – a definition found in Lerdahl and Jackendoff’s *Metrical Well Formedness Rules* (Lerdahl and Jackendoff (1983): 69-74), Victor Zuckerkandl’s metaphor of the ‘metric wave’ (Zuckerkandl (1971): 114-119) and Edward Cone’s metaphor of the bouncing ball (Cone (1968): 32-56). Schenkerians like Carl Schachter (Schachter (1999c): 80-82) and William Rothstein (Rothstein (1989): 8) subscribe to it too. (Christopher Hasty also talks about meter in terms of alternating patterns of strong and weak beats; however he does not require that these patterns demarcate equal time spans for meter to arise (Hasty (1997): 103-147.)

So, unless we have at least two rhythmic accents, neither of them can really be called a metrical accent. Consequently, the alignment that happens at the beginning of a musical group is between pitch grammar and *rhythm*, not meter, unless that initial rhythmic accent is followed by another rhythmic accent that sets up a periodic meter. So, a better way of saying that structural accents and metrical accents line up at the beginning of groups would be to say that structural and *rhythmic* accents line up at the beginning of groups.¹⁹ Once the initial rhythmic accent is made it may or may not set up a meter, and if it does set up a meter this may or may not align with the grouping structure of pitch in a passage, as we have discussed before. In other words, rhythmic accents are nothing but structural accents that are assigned certain durational values – which is why musical pitch can map to rhythm.

This last point brings us full circle, since the mapping between non-durational structural accents and durational rhythmic accents is nothing but the relation between tonal and durational rhythm in Carl Schachter's (Schenkerian) rhythmic theory – which is what started our discussion of musical rhythm earlier in this chapter. So, we see from the above that if we take away meter from our discussion of musical rhythm, we can still explain it within a grammatical, Schenkerian framework. However, this involves showing the link between pitch and rhythm, as opposed to pitch and meter.

Now, if pitch grammar maps to (metered or unmetered) rhythm in a Schenkerian description of musical structure, does this make music and language identical in this respect? That is, does a Schenkerian description of the mapping of grammar to rhythm also comply with the Region Condition for music – which would allow music to have something like a Phonetic Form too, and which would cement the connection between Schenkerian music theory and Chomskyan linguistic theory?

¹⁹ In their discussion of the types of musical accent, Lerdahl and Jackendoff discuss a third type of accent in addition to structural and metrical accent, which they call “phenomenal accent” (Lerdahl and Jackendoff (1983): 17). “Phenomenal accent” seems to be a catch-all term, since it includes any kind of accent that is not structural or metrical, such as accents due to register, dynamics, timbre and so on. However, all phenomenal accents are rhythmic accents, since all events in a musical surface occupy a certain temporal duration – all events in a musical surface have a certain rhythm. Now, Lerdahl and Jackendoff say that phenomenal accents can cue the listener to the meter of a passage (see also their MPR 4). Since phenomenal accents are rhythmic accents too, this makes perfect sense, as we have seen how rhythmic accents give rise to meter when they are evenly spaced.

Well, in light of the points just made, we should first state the Region Condition for music slightly differently: If two distinct events α and β are higher up in the pitch hierarchy of a musical passage (i.e. they are structural accents), then the event e that linearly intervenes between them will be lower in *unmetered rhythmic* prominence than both α and β . And given the way pitch maps to rhythm in Schenkerian approaches to this topic (i.e. the *asymmetrical* way in which it maps to rhythm), this Region Condition will generally hold true – suggesting an identity between music and language in this domain. This is why areas of rhythmic prominence in music tend to be tonally strong, as seems to be the case with structural beginnings (which, like language, are the normal places for one to ‘breathe’ or for pauses to appear).

Also, areas of rhythmic prominence that are not tonally strong, and which therefore violate the Region Condition, tend to violate the Condition in one specific way (i.e. a rhythmically prominent, but tonally weak, event e normally appears, i.e. linearly intervenes, between tonally strong events α and β , in only one specific way) – which is when e 's rhythmic prominence is realized *metrically*, i.e. when it appears on a strong *beat* relative to α and β . Without this metrical appearance e will not normally be seen as more rhythmically prominent than α or β – it will appear to be a mere *phenomenal* accent. (Maybe this is why rhythm has been studied in overwhelmingly metrical terms in music theory, because it is in this form that rhythmic structure conflicts with pitch structure, creating the problematic events that make for interesting theoretical study, including those that encourage the separation of pitch and rhythm, e.g. in Lerdahl and Jackendoff's separation of grouping and *meter*.)

But when a tonally weak event e does appear in a metrically strong position, it only does so because a *prevailing* meter allows it to – which is itself usually governed by hierarchical pitch properties, as in the alignment of structural beginnings with quadratic hypermeter. To understand this, let us briefly explore the concept of syncopation in music. Example 2.2-14 displays a rhythmic passage that is in phase with the meter suggested by 2.2-14A, but syncopated against the meter suggested by 2.2-14B. Now, the third attack of this passage (labeled x) is metrically weak in 2.2-14A but strong in 2.2-14B. So, if x is tonally weak compared to two (hypothetical) tonally strong events on either side of it, this would conform

to the Region Condition in A, but not in B. This is because in A *x*'s metrical weakness conforms to the Condition, but in B its metrical strength violates it. In other words, *x*'s presence in the passage would violate the Region Condition only if B were the prevailing meter. Also, if A were the prevailing meter then B would be heard as syncopated against it.

Now, is there any way of deciding whether the prevailing meter *is* A or B, and therefore whether *x* violates the Region Condition or not? Alternatively, is there any way of deciding which of the two

Example 2.2-14. Well-Formedness analysis of syncopation by Lerdahl and Jackendoff (1983)

The musical notation shows a sequence of notes on a staff. The notes are: quarter, eighth, quarter, eighth, eighth, quarter, eighth, quarter, quarter, eighth, quarter, eighth. An accent 'x' is placed above the fifth note (the eighth note). Below the staff are two rows of dots, labeled A and B, representing metrical positions. Row A has dots at positions 1, 3, 5, 7, 9, 11. Row B has dots at positions 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11.

meters is syncopated against the other one? Not on purely metrical grounds.²⁰ But we can make this decision on *pitch* grounds – and this is why the asymmetry of pitch and rhythm is so important. That is, pitch structure tells us that A should be the given meter of a passage, since it aligns with the structural beginning of the passage. Additionally, pitch structure would make B feel syncopated against the prevailing meter.

²⁰ On purely metrical grounds, there is a preference for *x* to be metrical weak in A because a strong metrical accent on *x* would ruin the evenly spaced meter that A currently has. This would violate the well formedness of the meter. In contrast, there is a preference for *x* to be metrical strong in B because this would conform with MPR 3, the preference for pitch and metrical accents to coincide (Lerdahl and Jackendoff (1983): 76-78). But both of these situations depend on which meter, A or B, is prevailing (i.e. if A is prevailing *x* should be metrical weak, and if B is prevailing it should be strong) – and, again, we do not have any purely metrical reasons for deciding in favor of either meter.

So, even a metrical phenomenon like syncopation can have a pitch-based explanation – especially when pitch governs the global structure of a musical passage, as it does in the instance of *structural* beginnings. And it is this influence of pitch on the global structure of a passage that shows us how the Region Condition for music might be met, in the mapping of grammar to rhythm in music. This is because the global influence of pitch will either make x in Example 2.2-14 a metrically weak event in A (which fulfills the Condition as noted above). Alternatively, it will make x a metrically strong event in B, but which is itself syncopated against A – making x syncopated against it too, and thus metrically weak (which then fulfills the Region Condition as well). So, in the specific condition under which the Region Condition might possibly be violated, i.e. when a tonally weak, but *metrically* (and not just rhythmically) strong event x appears to linearly intervene between two tonally strong but metrically weak events α and β , global pitch structure resolves the conflict by determining what the prevailing meter is, and therefore what the real, *global*, metrical strength of event x is.

Of course, in the *local* context of the musical surface, x might still appear to be metrically strong. So, if we understand meter B above locally, i.e. as a local prevailing meter, and not as syncopated against the globally prevailing A, then x will appear to be metrically strong. This is the situation that appears, for example, in a chain of suspensions. A chain of suspensions is usually seen as resulting from syncopation – e.g. when one voice in a polyphonic texture is rhythmically *displaced* against another, creating what appear to be a chain of dissonances that are metrically accented (Salzer and Schachter (1969): 78). But as we have just seen, tonally weak (i.e. dissonant) events that appear to be metrically strong locally, as a result of syncopation, are actually metrically *weak* in a global context, and thus meet the Region Condition. (In fact, a durational reduction, which shows how local rhythmic events relate to the global, tonal, context of a musical passage, will reduce suspensions out of it, to align the tonally strong consonances that surround it with the global meter – which is what the Region Condition requires.)

There is a challenge for the above argument though, which is that it is possible that the Region Condition really applies to the local context of a musical *surface*. Meaning that local violations of the Condition

cannot be just explained away in terms of global, pitch-grammatical considerations, as I have attempted to do in the previous paragraphs. Remember that in the case of language, the mapping of grammar to rhythm (specifically prosody), and phonology more generally, occurs in order for a sentence to be *interpreted* at the *surface*. Grammar must meet the well formedness conditions imposed by phonology in order for a *surface*- or S-structure to be generated that is also *pronounceable*. This ‘meeting’ takes place in the surface level of structure known as Phonetic Form, where grammar maps to prosody to generate an acceptable sentence for a speaker to pronounce. And as we saw in Chapter 1.2, this is why a *transformational* grammar is so important, because it is through the various grammatical operations known as transformations that the grammatical elements of a sentence can be moved around to meet the well formedness conditions imposed by phonology on the generation of a surface structure. So, how can one defend an isomorphism between how grammar maps to rhythm in *both* music and language, when the Region Condition seems to be routinely violated in music in common *surface* events such as tonally weak, yet rhythmically strong, suspensions?

Rather than being a criticism of my attempt to explain local violations of the Region Condition in global, pitch-grammatical terms, I think the answer to the above question actually shows the deepest connection between how grammar maps to rhythm in music *and* language, and thus the strongest connection between the Schenkerian approach to rhythm being explored here and the Chomskyan approach to these matters in language. Specifically, the answer to this issue takes us beyond Region Conditions into the very basis for a grammar to rhythm mapping in music and language.

Understanding this will require our revisiting a distinction made earlier in the chapter, viz. between rhythm as the result of tonal events and rhythm as an active force in music. We have seen earlier how the active role of rhythm in music often leads to structural and rhythmic accents being thrown out of phase with each other in shallower levels of musical structure, even though they seem to line up at the deepest levels of structure, at the level of tonal rhythm – normatively at the beginnings of pitch groups as I argued. But is the misalignment between pitch and rhythm at shallower levels of structure really the result of the active force of *rhythm*?

I actually think that this misalignment is really the result of the active role of *pitch* itself. Through various grammatical operations pitch itself is prolonged in shallower levels of structure, and when it attains duration in the musical surface, phenomena like phrase expansions and suspensions result. Moreover, these grammatical operations lead to surface *pitch* structures that then do not align with the evenly spaced metrical accents of a well formed meter – and therefore do not meet the well formedness conditions imposed by musical ‘phonology’, i.e. *rhythm*. And this is what leads to the conflict between pitch and rhythm, the breakdown of the mapping between pitch grammar and rhythmic structure, and ultimately the ‘anomalous’ phenomena that seem to violate the Region Condition, such as the tonally weak, but rhythmically strong, ones we have examined in some detail now – what Lerdahl and Jackendoff call “the counterpoint of structures that underlies much of the rhythmic richness of tonal music” (Lerdahl and Jackendoff (1983): 34).

However, the constraints placed on pitch by rhythm often helps resolve some of these conflicts even at local, surface, levels of structure, but under the influence of more global pitch properties, so that the interplay between rhythm (i.e. pitch) as an active force and as the result of tonal structure still allows *specific* surface structures to be generated in tonal music – akin to how transformational processes in language allow grammar to map to phonology via specific S-structures in language. This conflict resolution at the musical surface might not be amenable to a Region Condition-type explanation, but it does show how grammar does ultimately map to rhythm even at the surface – which is the ultimate purpose of the Region Condition anyway. So, in describing how this mapping occurs we can see the deep connection between this phenomenon in both music and language, and therefore, possibly, their identity in this domain, more so than the Region Condition might reveal. I would like to end by briefly discussing the two ways in which this interplay actually occurs in tonal music:

1. “Too many beats” (Expansions): We saw in Example 2.2-6 that when a phrase is expanded, the structural accents of that phrase do not remain equally spaced. In that example, measure 2 is expanded to two measures in duration, so the structural accent in that measure is not equally spaced between the metrical accents in mm. 1 and 3. Now, if structural accents were equivalent to metrical accents such

phrase expansions would lead to an ill formed meter, because a metrical structure is not well formed when it does not have equally spaced beats. However, since structural and metrical accents are not equivalent, the metrical and pitch structure go their own separate ways here, leading to a phrase that is seemingly out of phase with the meter. One could call this conflict between meter and pitch the problem of “too many beats” because the reason the pitch structure is out of phase with the metrical structure is because the expanded part of this group contains too many beats – its expanded length makes it too long for adjoining structural accents to be equally spaced, thus throwing the pitch structure out of phase with the meter.

But notice something interesting about this phrase. Even though pitch and meter are out of phase at the level of the whole phrase, rhythmic structure still constrains tonal structure *within* the phrase. Even though the phrase expansion puts the surface meter and pitch structure out of phase with each other, the expansion of the second measure takes up exactly 2 measures, not 2.5 or 1.5. So, after the second measure has been expanded, the next structural accent occurs on the *downbeat* of the next measure, even though it could appear anywhere now that pitch has asserted itself as an active force and made itself out of phase with the metrical structure of this passage. In this manner, pitch aligns with rhythm at the musical surface – it meets the well formedness conditions imposed by the meter. We can envision this as the placement of an extra beat in between the second and third structural accents, even after the expansion threw them out of phase with each other, which allows the surface meter to remain well formed without throwing it completely out of phase with pitch grouping structure. Moreover, this allows even the bar-level meter at the surface to have a coherent structure that aligns with surface pitch structure, what William Rothstein calls “surface hypermeter” (Rothstein (1989): 98).²¹

The interaction between rhythm and pitch in phrase expansions does not just apply to musical structure at levels of a measure or more; they even apply to the sub-measure level. For example, Carl Schachter talks about a four + *half* measure expansion of supertonic harmony in his durational reduction of Mendelssohn’s Op. 102 No. 4 “Song Without Words” (Schachter (1999c): 114-115). Here the

²¹ Incidentally, all theories of expansion, including Johann Philipp Kirnberger’s famous “echo expansion” (discussed in Rothstein (1989): 65-66), pace out the expanded measures in exactly such a way to match them to a well formed metrical surface.

alignment between surface pitch and rhythm after the expansion has put them out of phase cannot be resolved at a measure-level hypermetrical level. However, it can be resolved at a *hypo*-metrical level, i.e. at the half-measure level of meter. In other words, the requirement that pitch map to rhythm makes them align at the half-measure level of metrical structure – a half-measure musical PF, so to speak – after they have been thrown out of phase, which can be thought of as adding extra beats at the half-measure level.²²

2. “Too few beats” (Deletion): The opposite of the “too many beats” conflict between pitch and rhythm occurs commonly in music too. It is often seen where the opposite of phrase expansion takes place, viz. in cases of phrase *deletion*. In a deletion, a part of a phrase – usually the ending – is deleted, and replaced by the beginning of the next phrase. One of the most famous examples of this phenomenon can be found in the theme from Brahms’ Haydn Variations. Brahms’ own two-piano arrangement of this is shown in Example 2.2-15. The A section of this theme is comprised of a ten measure period, in which both the antecedent and the consequent phrases last five measures each (which is notated with the Arabic numerals in the middle of the stave systems).²³ Following this A section, an 8-measure B section ensues, which is divided into two quadratic, 4-bar phrases in mm. 11-14 and 15-18 that have the appearance of a statement and response pair.

In measure 19, the consequent phrase from the A section period (i.e. mm. 6-10) is heard once again. However, before the last measure of this phrase can be sounded, in measure 23, a *new* 4-bar phrase commences. The onset of this new phrase is heralded by the orchestra with crashing chords played *forte* (by piano II in the example) and by the sustained tonic B-flat pitch class played across four octaves (by

²² Such ‘hypometer’ actually seems to be a rather frequent occurrence in tonal music, and is often used to resolve the metrical conflict arising from syncopation as well. For example, in mm. 73-105 of the last movement of his Op. 24 “Spring” violin sonata, Beethoven changes the notated *alla breve* meter to a 12/8 in the violin part. At this point the notated *alla breve* meter is taken over by the piano; however it plays this rhythm only in syncopated fashion against the violin’s 12/8 rhythm. The only way one can resolve the metrical conflict here is by going into the half-measure level.

²³ Note how the unusual 5 + 5 metrical grouping of this period results from expanding mm. 2 and 7 of the period to two measures each. That is, the passage is generated from a normative 4 + 4 quadratic structure being expanded to a 5 + 5 structure in which mm. 2-3 and 7-8 each represent one (expanded) structural accent. The validity of this explanation can be seen by looking at the voice leading of the top voice in mm. 2 and 7. In both measures the top voice descends from E-flat5 to D5 – which would be normatively followed by the C5 in mm. 4 and 9. However, this large-scale voice leading is delayed by the local ‘inner voice’ descent to C5 and B-flat4 in mm. 3 and 7.

Example 2.2-15. Brahms, Variations on a Theme of Haydn, Op. 56b: Metrical deletion in the theme

The image displays a musical score for Brahms' Variations on a Theme of Haydn, Op. 56b, specifically focusing on the theme. The score is written in 2/4 time and marked *Andante*. It consists of two systems of staves, labeled I and II. Each system includes a treble and bass clef staff. The first system (I) begins with a piano (*p*) dynamic. The second system (II) begins with a forte (*f*) dynamic. The score is divided into measures, with measures 1 through 5 explicitly numbered. Above the first system, a bracket labeled "Expansion" spans measures 3 and 4. Above the second system, a bracket labeled "Expansion" spans measures 3 and 4. Above the third system, a bracket labeled "Expansion" spans measures 2 and 3. The notation includes various rhythmic values, rests, and articulation marks such as accents (>) and slurs. The key signature is one flat (B-flat major or D minor).

Example 2.2-15 (contd). Brahms, Variations on a Theme of Haydn, Op. 56b

15 *pp*

15 *pp*

19 *f*

19 *f*

24 *dim. smorzando* *p*

24 *dim. smorzando* *p*

Expansion

piano I, depicted with the long slurs in mm. 23-26). Finally, in measure 27, the final tonic chord is sounded and then expanded for three measures. In light of the *diminuendo* from *forte* to *piano* in the dynamics of these measures, this expansion can be seen as a “written-out fermata” that ends the piece.

As we see from the above, measure 23 has been metrically reinterpreted as the first measure of a new phrase (or specifically hypermetrical group) from its previous durational interpretation as the last measure of a 5-bar group. The “5 = 1” sign in measure 23 illustrates this. There is obviously a conflict of some sort here because an expected ending has been replaced by an unexpected beginning. The conflict is one of too *few* beats because both the last structural accent of the 5-bar group that would have spanned mm. 19-23, and the first structural accent of the new 4-bar group, have to compete for the same metrical accent (the downbeat of measure 23) for their actual, durational realization in the score. Obviously both accents cannot appear simultaneously (except in the special case of phrase elision, which I will not discuss here), so the former accent is ‘deleted’ – the phrase thus contracting from five measures to four (mm. 19-22), the opposite of what happens in expansions. This then aligns the phrase with the surface meter – thus allowing pitch structure to meet the well formedness conditions imposed by the meter.

Interestingly, even though the kinds of conflict between pitch and rhythm in such deletions is the opposite of what happens in phrase expansions, these conflicts are both resolved in the same way, i.e. by means of *surface* resolution of conflict between pitch and rhythm – conflict that was generated in the deeper structure of the passage by pitch taking on a more active role in it. In expansions, the deeper conflict between expanded pitch structure and rhythm is resolved by the surface expanding in a way that seems to introduce extra beats. In deletions, the deeper conflict between a 5-bar and a 4-bar phrase competing for the same metrical accent is resolved by a surface deletion of one of the competing structures.

So, we see once again how conflicts and resolutions between pitch and rhythmic structure arise from their interaction in the multilevel hierarchical grammar of a piece of music – a grammar that is ultimately determined by pitch structure, but which maps to rhythm at the surface to meet the well formedness conditions imposed by rhythm on it. Therefore, understanding these levels of structure can

help us explain the rich interactions between pitch and rhythm in music. But ultimately, it helps illustrate, in my opinion, the deep connection between the mapping of grammar to rhythm in both music and language, in what appear to be strikingly similar PF levels of structure in both. It also helps illustrate how a generative approach is the best way to understand the connection between grammar and rhythm, whether it be a Schenkerian approach to music or a Chomskyan approach to language, since this is the only approach that reveals the nuanced, asymmetrical, but ultimately consistent, way in which grammar maps to rhythm in both music and language.

One last point. The above interactions between pitch and rhythm at different levels of structure lead to the rhythmic richness of tonal music, and are therefore an important ingredient in creating the stylistic variety we see in music too. This is because the choice of metrical level, as seen in Example 2.2-14, helps set up or resolve conflicts that often impact stylistic choices made by composers as well, which then leads to the wonderful stylistic or idiosyncratic creativity that marks the global practice of music. In the first two measures of Mozart’s “Champagne” aria from the opera *Don Giovanni* (shown in Example 2.2-16), the pitch-constrained preference for a strong first measure meter is offset by the rhythmic accent that occurs on the fourth syllable in a line of Italian verse, known as the *Accento commune* (Webster (1991): 134-135).

Example 2.2-16. Mozart, “Champagne” aria from *Don Giovanni*, K. 527: Italian *Accento commune*

Accento commune

Presto

f

etc.

Text: *Finch' han del vino*

Syllable: 1 2 3 4

We know that metrical well formedness considerations will not allow both the first measure accent and the *Accento commune* to be stressed equally. So, the Italian accent ends up being normatively placed on the downbeat of the *second* measure, as syncopated against the prevailing hypermeter of the passage. This not only resolves the rhythmic conflict between the two accents but also becomes a defining feature of musical style, since the opposition of a first measure ‘German’²⁴ metrical accent to Italian prosodic accent is a hallmark of the High Classical style, seen in much of the music written by Mozart and his contemporaries in 18th-century Vienna.

²⁴ “German” because of the metrical emphasis given to beginning accents in ‘Germanic’ music theories such as the Schenkerian one being discussed in this chapter.

Chapter 2.3

From Description to (Internalist) Interpretation: Chopin Mazurkas, A Case Study¹

“Chopin was a genius of universal appeal. His music conquers the most diverse audiences. When the first notes of Chopin sound through the concert hall there is a happy sigh of recognition. All over the world men and women know his music. They love it. They are moved by it. Yet it is not “Romantic music” in the Byronic sense. It does not tell stories or paint pictures. It is expressive and personal, but still a pure art. Even in this abstract atomic age, where emotion is not fashionable, Chopin endures. His music is the universal language of human communication.” (Attributed to Arthur Rubinstein)

2.3.1. Music Analysis as Experiment

In the previous chapters we explored an internalist theory of musical meaning and rhythm, and we also saw how the internalist concepts of Logical and Phonetic Form proposed in Minimalist linguistics might have analogues in the ideas of contemporary Schenkerian music theorists like Kofi Agawu and Carl Schachter. In this chapter I will explore a practical demonstration of such a theory.

A practical demonstration of a theory invariably involves some sort of data, which can be studied empirically and which can serve as evidence for the theory. In the case of music theory, musical compositions and performances can serve as data for individual music theories, just as a corpus of linguistic constructions can serve as data for a linguistic theory (Uriagereka (1998): 95-97). The practice of music analysis, which examines the data of musical compositions and performances to discover patterns and regularities within them, is therefore a form of empirical study – a study that can give empirical validity to a music theory. Of course, what such an empirical study demonstrates depends on the scope of the theory associated with it. If the theory is a poetic or critical one, then an analysis will not yield any scientific observations about the music being analyzed. This is one of the reasons why the immense body of music-analytical literature that engages in music criticism is often ignored by the practitioners of musical science in their investigations of music. On the other hand, a scientific musical analysis, of the kind that depends on listeners’ responses to or judgments of certain musical passages, is

¹ Parts of this chapter were presented at the 11th International Congress on Musical Signification in Krakow, Poland, under the title, “Towards a Formal Theory of Musical Meaning.” This was also published as “Against Musical Function: Formal Signification in Chopin”, in *Music: Function and Value: Proceedings of the 11th International Congress on Musical Signification: Volume 2*, (Eds.) Teresa Malecka and Małgorzata Pawłowska. Krakow: Akademia Muzyczna w Krakowie and Musica Iagellonica, pp. 401-411.

often seen as irrelevant for more aesthetically-oriented theories of music – which explains the derision for music cognition studies often expressed in certain branches of musicological scholarship.

There obviously has to be a balance between the two though – if a critically- or culturally-oriented music theory ignores objective facts about music, in whatever way we define “musical fact”, it will cease to have the explanatory force of a true theory (see Agawu (1992b, 2004, 2006) for some thoughts in this regard, specifically concerning the New Musicology). On the other hand, if a scientific theory of music ignores musical phenomena that are relevant for a scientific understanding of music, even if they appear to be stylistic or aesthetic ones – such as those having to do with musical meaning – then the theory will be deemed inadequate or lacking in explanatory power as well. As we saw in chapter 2.1, it is precisely objections of this kind that make a study of musical meaning so important for the kind of scientific music theory being explored in this dissertation. In turn, a project in musical analysis can help provide a practical, empirical demonstration of such a theory. To this end I now turn.

2.3.2. Chopin: Formalist or Functionalist?

The music of Frédéric Chopin provides a wonderful analytical case study for a variety of different music theories. One of the reasons for this is the variety of approaches that can inform such a study. From an externalist perspective, Chopin’s music has proved irresistible to many because of the way it seems to demonstrate the externalism of musical meaning. For one, his Polish heritage has provided a resource for those seeking to explore issues of nationalism or authenticity in music. Some have even tried to trace his music back to Polish folk sources, even though such attempts have always been mired in controversy (e.g. see Abraham (1939), Dahlhaus (1989), Milewski (1999), Swartz (1983)). More generally, his reputation as a virtuoso pianist, his lifelong ill-health and death at the young age of 39, combined with his relationship with George Sand, have helped paint a picture of him as an enduring, tragic Romantic figure, which scholars have tried to situate in his unique musical style, with its ubiquitous emphasis on *rubato* (e.g. see Rosen (1995): 410-471). Finally, his mastery of compositional genres derived from dance forms, such as the polonaise, mazurka and waltz have inspired some authors (e.g. McKee (2004a)) to investigate

Chopin's music from this functional perspective, e.g. to see how his music conveys the expressive gestures of various dances, and is thus able to function as dance music.

However, despite the various functional and normative views of musical meaning that his music has, and continues to, inspire in some, others have noted the 'universal' quality of Chopin's music, which seems to cut across norms and social customs. In other words, to understand the significance of Chopin's music, in this perspective, has less to do with meaning, and more with how his compositions engage the very essence of tonal grammar (and therefore meaning in an *internalist* sense), as a result of which Chopin's music seems to epitomize, the universal, language-like nature of music – something that has been the focus of this dissertation from its very outset. This is what the celebrated pianist and Chopin interpreter Arthur Rubinstein seemed to be alluding to in the quotation cited at the beginning of this chapter, when he speaks of Chopin's music being “the universal language of human communication”.

Indeed, it is this attribute of his music that made Heinrich Schenker a fan of Chopin's music too. Notorious for his dismissal of folk music, and generally any music that does not originate from the major-minor system – and universally 'Germanic' in his preference of composers – Schenker still allowed Chopin into his pantheon of master composers:

“If the writer elevates the name of Frederic Chopin for inclusion in the roll of great German masters, this is because, despite the fact that his masterworks do not stem directly from Germany but are indirectly bound to it, he wishes them, too, to be accessible as a source of the highest operations of genius, and in this most exalted sense also to place them newly at the service of the German youth.” (Schenker (2004): 21)

“As was the case in Italy with Scarlatti, so too in Poland Chopin stands isolated – a musical genius. And, like Scarlatti, he too was forever estranged from his own people, albeit he came from the people, felt their spirit within him and expressed that spirit with fervour. ... For the profundity with which Nature has endowed him, Chopin belongs more to Germany than to Poland. May German musicians at long last give him their attention and understanding.” (Schenker (1994): 81)

As can be seen from the second quote above, the Italian Domenico Scarlatti was the only other non-German granted inclusion in Schenker's list of master composers. But Chopin's inclusion in this list is more striking than Scarlatti, since the latter composed in a late Baroque/early *galant* style that influenced much of the 18th century Viennese music adored by Schenker, most notably that of Wolfgang Amadeus

Mozart. So, what is it about Chopin's music that transcends his Polishness – that makes his music the 'universal language of human communication'?

I will attempt to answer this question in the next few pages by exploring the structure and meaning of some of his best known, and most characteristically Polish, works, the piano mazurkas. However, I will use the grammar-centric Schenkerian approach to studying musical structure and meaning discussed in the previous chapters in this venture, through which I will attempt to show that even these most culturally-laden of Chopin's pieces can be understood in terms completely internal to the pieces themselves, i.e. in terms of their grammar, and how this grammar relates to their semantic and rhythmic structure. In this manner, I hope to show not only how Chopin's music reveals the profound internal qualities of music, but how this further demonstrates music's deep connection to language too, as aspects of human nature – which has been the main goal of this dissertation.

2.3.3. Internalism, Generation and Interpretation in Chopin's Mazurkas

To commence this study, let us revisit how the music theorist Kofi Agawu describes the internal qualities of grammar and meaning in music, in his exploration of the Bach chorale "Ach Gott, wie manches Herzeleid". (See section 2.1.4 in chapter 2.1.) In that study, Agawu discusses the plays of forms in the chorale, which reveal both the chorale's internal grammatical structure, and how the chorale's meaning results from this. Specifically, we can imagine the surface of the chorale, with all its melodic and rhythmic nuance, as arising from a simple harmonic progression I – V – I, i.e. Schenker's *Ursatz*, and the most prototypical of chord progressions in Western tonal music. Agawu goes on to describe how Bach elaborates this simple progression with other chords and contrapuntal techniques (i.e. the 'play of forms'), and that is how the meaningful, articulated surface of the piece arises. In various parts of this dissertation, we have had a chance to compare this with what happens in the generation of sentences, or specifically S-structures, in language, and how this leads to the comprehension of linguistic meaning, and the articulation of linguistic speech sounds and gestures too.

Bach's "Ach Gott, wie manches Herzeleid" chorale was composed in 1724 in Leipzig, as part of a church cantata to be performed at the traditional Sunday Lutheran service. Agawu's analysis shows us how music suitable to such a function can be generated from the 'conditions of possibility' afforded by tonal grammar. Let us travel from Leipzig in 1724 to Warsaw over a hundred years later where, in 1835, Chopin composed his mazurka in C major, the third of his Op. 67 set. Example 2.3-1 gives a full score for this piece. Let us now explore this piece, to see whether we can describe its grammatical structure, and understand how its meaning and rhythmic articulation can be interpreted.

The simplicity of the Op. 67 No. 3 mazurka makes it a particularly good place to start an exploration of the grammatical structure and meaningfulness of the Chopin piano mazurkas. It is a relatively short piece with a conventional rounded binary (i.e. A-A-B-A) formal design. The exposition is made up of an A section that consists of the main theme of the mazurka (mm. 1-16), which is repeated once without variation. This theme itself is a 16-bar compound period, composed of an 8-bar sentential "compound antecedent" phrase (mm. 1-8, repeated in mm. 17-24), and an 8-bar sentential "compound consequent" phrase (mm. 9-16, repeated in mm. 25-32), the consequent phrase being quite similar to the antecedent, apart from the fact that it doubles the melodic material of the antecedent in 3rds and 6ths, and replaces the imperfect authentic cadence at the end of the antecedent with a perfect authentic cadence.²

The A section of the recapitulation is essentially identical to the A section of the exposition, the only difference being that it does not repeat the main theme in the way the exposition does. Therefore, the vast majority of the music of this piece consists of just the 8-bar sentential compound antecedent phrase of the main theme, given how similar the compound consequent phrase, and the various repeats of the compound antecedent, are to the original statement of the antecedent in mm. 1-8. The only real contrast to

² The use of terms like "compound antecedent" and "compound consequent" above invokes William Caplin's theory of musical form (Caplin (1998)). Some might object to my using Caplin's theory to discuss the formal aspects of Chopin's music, as I will continue to do in this chapter, given that it was developed to deal with the music of the High Classical style, specifically the works of Haydn, Mozart, and Beethoven. But as we will see later, Chopin was noted for his use of conventional Classical forms too, so the above objection is one I will ignore for our present purposes.

Example 2.3-1. Chopin, Mazurka in C major, Op. 67 No. 3: Score

Allegretto ♩ = 144

p rubato

tr

tr

tr

f

tr

tr

a tempo

cresc.

< sf

ff poco rit.

p

tr

tr

tr

tr

f

cresc.

< sf

Ped. * Ped. * Ped. * Ped. * Ped. * Ped. *

Ped. * Ped. * Ped. * Ped. * Ped. * Ped. *

Ped. * Ped. * Ped. * Ped. * Ped. * Ped. *

Ped. * Ped. * Ped. * Ped. * Ped. * Ped. *

Ped. * Ped. * Ped. * Ped. * Ped. * Ped. *

Example 2.3-1 (contd). Chopin, Mazurka in C major, Op. 67 No. 3

The image displays a musical score for Chopin's Mazurka in C major, Op. 67 No. 3, covering measures 31 to 52. The score is written for piano and features a variety of musical notations and performance instructions.

Measure 31: The right hand begins with a trill (*tr*) over a half note chord. The left hand plays a steady accompaniment. Dynamics include *ff* and *poco rit.* Below the staff, the instruction *ped.* is followed by an asterisk.

Measures 32-37: The right hand continues with chords and trills. The left hand features a melodic line with dynamics *sf* and *pp*. The tempo marking *a tempo* is present. *ten.* markings are placed above the right hand notes. *ped.* instructions with asterisks are placed below the staff.

Measures 38-44: The right hand has a melodic line with a trill at the end. The left hand has a melodic line with dynamics *sf*, *pp*, *rit.*, and *p*. The tempo marking *a tempo* is present. *ped.* instructions with asterisks are placed below the staff.

Measures 45-51: The right hand features a melodic line with a trill and dynamics *f*. The left hand has a steady accompaniment. *ped.* instructions with asterisks are placed below the staff.

Measures 52: The right hand has a melodic line with a trill and dynamics *cresc.* and *sf*. The left hand has a steady accompaniment with dynamics *ff* and *poco rit.* *ped.* instructions with asterisks are placed below the staff.

this comes from the short, eight-measure B section (mm. 33-40), which consists of an equally simple 4-measure ‘statement-response’ idea that is repeated once. Given this minimalist design, with the additional fact that it is in the most accessible of keys, C major, this mazurka provides us a particularly clear and concise model for how structure and meaning arise in the piano mazurkas of Chopin.

Example 2.3-2 describes how the basic pitch structure of mm. 1-8, where the main content of this piece lies, is derived. Just as in Agawu’s description of the Bach chorale’s final phrase in Example 2.1-5, we start with the fundamental harmonic structure of Western Classical tonal music, viz. I – V – I. This is shown in level A of Example 2.3-2. This level is identical to Agawu’s level A, including in the way the chords have been voiced. Level B of the example adds the same ‘enriched’ predominant sonority that Agawu does, the $V^{6/5}/V$ chord, with the same voicing as well. Therefore, until now the derivation of Example 2.3-2 has proceeded identically to Agawu’s derivation of mm. 13-16 of Bach’s chorale “Ach Gott, wie manches Herzeleid”.

Agawu says that from this level to “Bach’s four-bar phrase is but a short step”. In our example, the path to the surface chord progression of Chopin’s mazurka is equally short albeit different. Level C of Example 2.3-2 demonstrates this. The V chord of levels A and B is now transformed into a V^7 , and then expanded by means of a cadential 6-4 sonority. Importantly, this 6-4 sonority is derived *from* the V rather than the other way around, in spite of the fact that the pitches of this sonority form a C-major triad – the tonic of this piece. Since a C-major triad initiated the derivation of this passage at the deepest level of structure, the derivation of the same sonority at a different, ‘shallower’ level of structure later in the derivation demonstrates the recursive nature of chord progressions discussed in chapter 1.1.

In addition to the expansion of the dominant, level C also shows how the initial tonic of the passage is expanded. Here the recursive nature of tonal music is demonstrated even more clearly, for within the larger passage being generated in Example 2.3-2, we have a smaller nested progression that begins and ends with the tonic chord, with two intermediate chords that serve to expand tonic harmony. (This expansion is illustrated by the slurs in level C that join the two tonic chords that bookend this

Example 2.3-2. Chopin, Mazurka in C major, Op. 67 No. 3: Harmonic generation, mm. 1-8

A

I V I

B

I $V^{\frac{6}{5}}/V$ V I

C

I $V^{\frac{6}{5}}/V$ V I

8 - 7
6 - 5
4 - 3

Chopin Mazurka, Op. 67 #3, mm. 1-8: LH Score with RH Melodic Reduction

D

mm. $\hat{3}$ $\hat{2}$ $\hat{5}$ $\hat{2}$ $\hat{1}$

nested phrase.) Since the two intermediate sonorities are not functional chords, but are merely derived by voice leading, they have been represented with quarter-note symbols to distinguish them from the structurally-important chords (represented with the usual whole-note notation) surrounding them.

The end result is that we have eight chords that have been derived from the initial I – V – I progression and that act as the harmonic backbone for this passage. That the path from here to the surface of mm. 1-8 of the mazurka is short is indeed evident when one compares the chord progression of level C with the actual left hand part of the mazurka depicted in level D of Example 2.3-2. Apart from a few differences in chord voicing, these two levels are identical. Level D also gives a melodic reduction of the right hand part of the passage. Only the reduction has been provided here in order to highlight the structural notes of the melody and show how they relate to the harmony. The actual right hand melody will be derived shortly.³ We can see from this melodic reduction how the melodic ‘backbone’ represented by the right hand part relates to the underlying harmony of the passage. For one, the overarching melodic progression of the passage is a descending line involving scale degrees 3, 2 and 1 (indicated by the upward-pointing stems), which are none other than the soprano pitches of the three chords in level A from which this passage was generated. Therefore, these three melodic pitches constitute the Fundamental Line (Schenker’s *Urfinie*) of the passage, and when combined with the bass notes that harmonize them, they form the Fundamental Structure (Schenker’s *Ursatz*) of the passage.

The last pitch, the C5, is given in parentheses because it is not actually there in measure 8 of the score, but is merely implied by the harmonic structure of the passage (review section 1.2.4.v in chapter 1.2, in this regard). I will discuss how this C5 is substituted with another pitch in my discussion of how the right hand part is derived, but for the time being let me just point out that this substitution prevents the passage from achieving closure – which a return to the tonic C would have accomplished. The music being left incomplete, more music is thus necessitated, which allows the music to move forward. This is a brief example of how composers can play with larger, more ambitious compositional designs by leaving

³ Since this is just a reduced representation of the melody, and not the actual score, the right hand part lacks a time signature. Therefore, the barlines and note heads here do not represent actual durations. They have only been provided so that these representative pitches can be related to the actual score.

passages incomplete or by generating expectations for more music. It also shows how Chopin took the simple form of a mazurka dance and turned it into a more ambitious form of ‘art’ music, not necessarily intended as dance music.⁴

Let us conclude our discussion of Example 2.3-2 by examining the rest of the melodic reduction in level D of the graph. In mm. 1-4 the melody of the piece follows a descending line E5-D5-D5-C5. This demonstrates how Chopin plays with the melodic structure of the piece, since this descending line, although similar in pitch structure to the *Urlinie*, is not the same as it. This is because these pitches represent a motion ‘into an inner voice’ (a technique discussed by Schenker (1979): paragraph 203), which delays the structural descent of the melody in the *Urlinie*. The inner-voice nature of the pitch motion just described can be seen in its representation with black note heads with downward pointing stems in the example, which distinguishes it from the white note heads with upward pointing stems of the *Urlinie*. Chopin continues to follow an inner voice in measure 5, rather than the structural D5 of the soprano (hence the representation of this D5 in parentheses, as an implied tone), before resuming the structural melodic line in measure 6 (i.e. E5).

These inner-voice motions are described more clearly in the next example, Example 2.3-3. This example shows how the right hand part of the main melody of the Op. 67 No. 3 mazurka arises. Importantly, this step is not a derivational step – the harmony for this passage has already been derived, so no further derivation (or generation) is required. What happens now is an act of interpretation, specifically a ‘phonetic’ interpretation, which results in the articulation of the musical surface of the passage. In other words, now the generated pitch structure of the passage has to be given a specific durational realization. Moreover, it has to be given a specific sonic realization – the abstract harmonies of the passage have to be sounded in a particular way. This aspect of phonetic interpretation in music is

⁴ One important point to note here is that even though the above passage is left open in terms of the design of the piece, the passage itself is grammatically complete (which is implied by the complete chord progression from which it is derived). This exemplifies the distinction between grammatical structure and poetic structure that was discussed in chapter 2.1.

Example 2.3-3. Chopin, Mazurka in C major, Op. 67 No. 3: Interpreting the theme

Chopin Mazurka, Op. 67 #3, mm. 1-8: Melodic Middleground

mm. ② ⑤ ⑧

A

This musical score shows the piano accompaniment for the first eight measures of Chopin's Mazurka, Op. 67 No. 3. The right hand (treble clef) features a sparse melodic line with notes and rests, while the left hand (bass clef) provides a rhythmic accompaniment of chords. Measure numbers 2, 5, and 8 are indicated above the staff.

Chopin Mazurka, Op. 67 #3, mm. 1-8: Melodic Prolonged Middleground

mm. ② ⑤ ⑧

B

This musical score shows the piano accompaniment for the first eight measures of Chopin's Mazurka, Op. 67 No. 3. The right hand (treble clef) features a more continuous melodic line with slurs, while the left hand (bass clef) provides a rhythmic accompaniment of chords. Measure numbers 2, 5, and 8 are indicated above the staff.

Chopin Mazurka, Op. 67 #3: Melodic Surface
(a. mm. 1-8, 17-24 and 41-48, b. mm. 9-16, 25-32 and 49-56)

a

b

C

This musical score shows two different interpretations (a and b) of the piano accompaniment for the first eight measures of Chopin's Mazurka, Op. 67 No. 3. The right hand (treble clef) features a melodic line with trills (tr) and slurs. The left hand (bass clef) provides a rhythmic accompaniment of chords. Measure numbers 2, 5, and 8 are indicated above the staff.

something we have not discussed much in previous chapters, but it usually involves revoicing the harmonies of the derivation in order to highlight certain aspects of the melody (particularly inner-voice

ones) over others. This can be done by arpeggiating these chords or elaborating them with non-structural pitches.

Level A of Example 2.3-3 just shows us what level D of Example 2.3-2 already did, viz. how the melody relates to the harmony of the passage. In addition it shows the inner-voice origins of the melody in mm. 5-7 (the A4-G4-F4 of these measures come from the tenor voice, shown in level C of Example 2.3-2). This is why the structural D5 of measure 5 is but implied in the surface melody – it is substituted by an inner-voice motion. In measure 6, however, the E5 of the structural upper voice is brought back into play in the following way: the surface melody in that measure consists of three notes G4-C5-E5 in what looks like an arpeggiation. Actually, this is an unfolding (Schenker's *Ausfaltung*), a technique in which two pitches of a triad are 'spread out' horizontally, i.e. the interval between these pitches is 'unfolded'. We have seen from level C of Example 2.3-2 that the cadential 6-4 chord in the basic harmonic progression of the passage consists of a soprano E5 and a tenor G4. These two pitches are unfolded in measure 6 to give the impression of an arpeggiation. The unfolding aspect of this measure is illustrated by the zigzag line in measure 6 in level B of Example 2.3-3. Just as measure 6 unfolds the interval between E5 and G4, the interval between the soprano D5 and tenor F4 of measure 7 is unfolded in the same way, as can be seen by the zigzag line in that measure. The left hand part of this level is identical to level A and does not require further explanation.

Level B also shows us something interesting about the preceding tenor-voice motion A4-G4-F4 of mm. 5-7. Where does this voice leading come from? After all, the first four measures saw a motion into an inner *alto* voice, so how did the melody switch to the tenor voice all of a sudden? It didn't – this is because the $V^{6/5}/V$ chord that contains the A4 prolongs the *following* dominant harmony of mm. 6-7. Moreover, we have seen the unfolding relationships that connect the tenor pitches of mm. 6-7 with its soprano pitches. So, the tenor voice motion of mm. 5-7 does not arise from the preceding alto voice but from the following, structural V harmony. This harmony, as we have seen earlier, is expanded by the preceding cadential 6-4 sonority that contains the soprano E5. Therefore, the E5 continues to the D5 of this dominant harmony, thus giving back the role of the surface melody to the structural top voice – from

whom it was taken away in the motion to an inner voice in measure 2. All these intricate inner-voice contrapuntal maneuvers, however, give the sense of an elegant descending melody E5-D5-C5-A4-G4 in mm. 1-5, which then rises back up to the E5 in measure 6 – a play of forms indeed! This illusion is further heightened by the intricate neighboring motions that decorate each pitch of the descending line E5-D5-C5-A4-G4, illustrated by the unstemmed note heads in mm. 1-5.

But the play of forms is not over yet. As we have seen earlier, the final C5 of measure 8 is also implied in the actual melody of the passage. Instead, we hear a high G5 in its place, which is elaborated by an appoggiatura A5 that appears before it. Where does this G5 come from? Well, it has an inner-voice origin too – it belongs to the *alto* voice of the background harmony. If you look at level C of Example 2.3-2, you will notice that this alto voice is made up of the line C5-C5-B4-G4 in mm. 5-8. In the actual surface melody, shown in level B of Example 2.3-3, the last G pitch class is realized an octave higher as G5, by ‘reaching over’ (Schenker’s *Übergreifen*) the structural soprano line. This is shown by means of the diagonal line connecting the G to the preceding B4 in the alto. This preceding B4 is not actually heard in the surface melody though (hence its notation in parentheses, as an implied tone), which is why the A5-G5 motion in the last measure of the surface melody seems to continue the soprano line of the preceding measure instead.

What this implies is that the reaching over motion prevents the music from coming to a close at the end of measure 8, and this allows the music to flow forward. In other words, the ‘phonetic’ realization of the harmonic background of this measure, by means of a specific articulation of its pitch structure that involves a reaching over motion, allows one to *semantically interpret* this measure as a middle (which builds on our discussion in chapter 2.1 of the significance of beginnings, middles, and endings in a grammatical theory of musical meaning). The same can be said of the inner-voice motion that happened in mm. 1-4 – meaning that in the specific pitch articulation of the surface, the *sound* of this surface is paired with its *meaning*, as a middle. In this manner, the phenomenon of interpretation in music, which itself results from the phenomenon of grammatical generation, shows us how music can be seen as a system of sound-meaning pairings too, just like language.

Finally, the two graphs a and b of level C give us the various versions of the compound antecedent phrase of mm. 1-8 that Chopin wrote in the actual score. C(a) shows us the score for the original statement of the phrase in mm. 1-8, which is repeated at mm. 17-24 and 41-48. As we can see, the pitch content of this score is almost identical to that of level B, with the exception of some structurally non-essential notes that were left out of level B. (These include the turn figure E5-D5-C#5-D5 of measure 7, which is represented in level B by just the structural D5 this figure elaborates, and the passing tone E5 between the D5 and F5 of the same measure.) Graph C(b) shows us the score for the compound consequent phrase that follows the compound antecedent in mm. 9-16, 25-32 and 49-56. This phrase basically doubles the melodic line of the antecedent phrase in parallel 6ths (apart from a few notes in mm. 5-6, where it is doubled in 3rds), the pitches of the doubling voice being members of the chords from which this passage was derived, of course. In addition to this variation, C(b) shows us how the last measure of the compound consequent avoids the reaching-over gesture at the end of the compound antecedent phrase, and just descends to the structural tonic C5 instead, as predicted by the harmonic structure of the passage, to give closure to the A section of the piece. (This means that the specific surface pitch structure of this measure, as opposed to the last measure of the compound antecedent phrase, can be semantically interpreted as an “ending”.) A final modification that Chopin makes to the compound consequent relative to the compound antecedent can be found in the bass line. Whereas the bass line for the first half of the antecedent and consequent (the “basic idea” part) is identical, Chopin replaces the $V^{6/5}/V$ in the bass line of the second half of the antecedent with a IV chord in the second half of the consequent. This switch from a chord applied to V to a *diatonic* predominant harmony (i.e. IV) keeps the ending of the passage more firmly in C major, and therefore allows it to close the A section more convincingly.

The above discussion shows how the surface structure of Chopin’s C-major mazurka Op. 67 No. 3 is arrived at by a series of generative and interpretive stages. Since the passages thus generated make up all but eight measures of the piece, our simple generative procedure is able to show consequently how much

a characteristic Polish dance piece written in the mid-nineteenth century shares with a German Lutheran hymn written over a hundred years earlier. Now anyone familiar with Chopin's music will not be surprised that his music has a lot in common with the music of the German Baroque, after all J. S. Bach was supposedly his favorite composer! And we certainly have not finished discussing all the characteristic aspects of Chopin's mazurkas. (In particular, I have not yet discussed the rhythmic aspects of the Op. 67 No. 3 mazurka, which is related to the phenomenon of "durational interpretation" in music. I will return to this later in the chapter.) But the point I would like to make for now is that, insofar as the surface of a musical piece is the bearer of information that can be interpreted, we have been able to show how the surface of this Polish dance piece can be generated from a basic musical 'lexicon' (i.e. some chords), and then semantically and phonetically interpreted, without any reference to Chopin's Polishness, or indeed any aspect of his music external to the notes themselves. In other words, we have shown how the celebrated, 'universal' qualities of at least one of Chopin's masterpieces can be explained in a manner that is contained completely within a grammatical description of its generation. This demonstrates the *internalism* of Chopin's music, or more specifically, our ability to theorize it.

One particularly salient aspect of Chopin's music, and indeed of many composers with alleged or actual folk/nationalistic influences, is its use of modal melodies. Since the use of such melodies, or the modal scales with which they are associated, seems to depart from use of the major-minor musical system by composers in the Western Classical "common-practice", they are often seen as the clearest indicators of folk influence, and thus some sort of external (i.e. cultural or nationalist) signification in the music. Even Schenker, who was highly critical of such practices (given his commitment to the major-minor system) noticed this aspect of Chopin's music. Regarding the latter's use of Lydian inflections in the Op. 24 No. 2 mazurka, Schenker speaks of Chopin's "literal quotation" of "nationalistic melodies" and his "genre imitations":

"Chopin by no means intends to establish the old [i.e. modal] systems as equivalent [to the major-minor diatonic system] and as independent; this is sufficiently clear from the refined artistry he uses in the introduction as well as the harmonization in general to provide the listener with the absolute certainty of

only C major and F major (in this connection, compare in particular the ingenious conclusion of the Mazurka, which orients the listener beyond any doubt!). Thus, the passage in question simply contains a few features of artistic archaism, a highly ingenious trick, such as could befall Chopin occasionally in the midst of his fantastic improvisations. It is merely a literal quotation, a curious genre-imitation, from that golden age when people still believed in the “Lydian” system, and sang and played irresponsibly, especially “nationalistic” melodies, because actually they did not know how to play and hear better.” (Schenker (1987) Book 1: 57-58)

Given its Schenkerian influence, many of the discussions in this dissertation have focused on music that is essentially diatonic. Of course we are not committed to only music that is diatonic, given the wider universalist and Minimalist influences on this dissertation, and as was even demonstrated in the investigation of Indian music in chapter 1.3. So, Schenker’s (largely ideological) commitment to the major-minor system, as expressed in the above quote, can be safely ignored for our purposes. But his opposition to extra-musical ‘quotations’ or genre imitations *is* worth considering here, given its resonance with the internalist approach to musical structure and meaning that we have been exploring. In particular, it seems that much of the externalism that the use of folk or nationalistic elements in music is supposed to signify can be subsumed within an internalist approach to music.

Let me elaborate on this by exploring the very mazurka passage under question here, mm. 21-36 of the Op. 24 No. 2 mazurka in C major. We have already seen how (some of) the structure and meaning of a mazurka surface can arise from a basic harmonic progression in C major. Example 2.3-4 revisits this for the above measures of Op. 24 No. 2. Level A of the example just shows the basic chord progression of mm. 21-24, which is repeated in mm. 25-28, 29-32 and 33-36. If we read this passage as being in the F-major (rather than Lydian) scale, as is implicit in Schenker’s comments, then it can be derived in exactly the same way as levels A and B of Example 2.3-2 were derived, i.e. by starting with the tonic chord and then deriving a dominant, and then a predominant chord, from it. The first thing we can see after this progression has been generated is that it is identical to the actual left hand part of the score of this passage, as depicted in level B of Example 2.3-4. So, the surface content of this left hand part is determined directly by the harmonic grammar from which it was generated.

Example 2.3-4. Chopin, Mazurka in C major, Op. 24 No. 2: Pitch structure. mm. 21-36

A

F major: I V^6/V V I

Chopin Mazurka, Op. 24 #2: LH part mm. 21-24, 25-28, 29-32 and 33-36

B

Chopin Mazurka, Op. 24 #2: Melodic Interpretation. a. mm. 21-24 b. 25-28 c. 29-32 d. 33-36

a

b

c

d

Now, notice the B-natural on the downbeat of the second measure of level B. B-natural is the very note that is supposed to signify the Lydian folk influence of this passage. However, in this context the downbeat B-natural is anything but a Lydian scalar pitch. Rather, it is clearly a member of the first inversion G⁷ chord that prolongs the following C-major triad, and as such no one has ever claimed that this pitch signifies a Lydian meaning for the measure. So, whatever external folk/nationalistic signification we would like to find in this passage, this specific B-natural can only convey an internal one, governed by the harmonic context in which it occurs. Of course, it is not this B-natural but the one two measures later in the soprano that has been associated with a Lydian meaning. But let us understand how the melody that contains this B-natural arises – which is especially important given the melody’s soaring, flowing quality that also allegedly signifies the vocal traditions of Italian opera, which Chopin was known to admire. Level C of Example 2.3-4 therefore explores the genesis of this melody. Since the passage in which the melody occurs is made up of the same four-bar chord progression repeated three times, the melody can be seen as a theme with three variations, each represented in graphs a, b, c, and d of level C.

The main thing to point out here is that the flowing quality of the melody results from the wonderfully creative way in which Chopin voices the underlying chord progression of this passage. The lower stave in graph C(a) just repeats the chord progression given in levels A and B of the example, but also transposes the pitches of these chords to higher registers – so it just revoices these chords by registral transposition. Once this is done, we can easily see how the melody (shown in the upper stave of C(a)) arises from the voicing of the chords in the lower stave. For example, the F5 of measure 21 skips up an octave to F6 in measure 22 (instead of staying in the same register, as shown in level A), and then descends to an implied E5 in measure 23, before returning to its starting register in measure 24. In measure 22, this voice is also decorated by an upper neighbor G6.

By following all the diagonal lines in graphs a-d of level C, we can see how all the voices arise in this way; often by leaping up or down an octave or moving to pitches that are not realized in the actual score but are implied by the harmonic progression of levels A and B (and thus marked in parentheses). All these leaping and skipping motions give the melody a wonderfully fluid quality. This might seem to

signify folk or operatic vocal writing, which it might very well do – but that only makes sense in the grammatical context of harmonic progression, chord voicing and voice leading that it occurs in.

In fact, it is only in this context that the soprano B-natural, the ‘intimate Lydian B’, makes sense too. As we can see in graphs C(b) and C(d), the relevant voice leading in the last measures of these graphs is the motion from the ‘tenor’ G5 in measure 23 to the ‘soprano’ A5 in measure 24 – which is *elaborated* by the B-natural as an incomplete upper neighbor note. This makes the prolonging, non-structural role of the B-natural clear. If the B-natural were a B-flat (and the passage unequivocally in F major instead) this would make its non-structural role even clearer, since it would form the dissonant interval of a tritone with the pitch class E in the chord. (Hence Schenker’s comment that the presence of F major here is made certain by the harmonization). So, the B-natural may very well have a folk-derived Lydian meaning but this is completely subsumed under its meaning as a non-structural upper neighbor in the grammar of the passage, because that is how the pitch arises in the first place.

As was also discussed in Example 2.3-3, the melodic and chord voicing gestures seen in the mazurka Op. 24 No. 2 passage are all matters of interpretation, they are not generated from the underlying chord grammar of the piece. And since interpretation is subject to all kinds of performance factors, including composer idiosyncrasy, one can choose to interpret the melody of a piece in whatever way one wants to. But interpretation also depends on grammar – not only would there be nothing to interpret if no surface structures are generated by grammar, whatever interpretation takes place would not even be possible if the grammar did not produce something *interpretable* in the first place. Just as the grammatically incorrect sentence “The drank boy the milk” is not interpretable (whereas “The boy drank the milk” is), a grammatically incorrect musical ‘sentence’ cannot be interpreted either, whether in internal terms, or in external terms, as signifying something nationalistic or operatic or dance-based etc. Of course, one could argue about what a grammatical ‘sentence’ in music is (as discussed in section 1.2.2 of the second chapter), but there can be no doubt that the phrases of the Op. 24 No. 2 mazurka are grammatical – at least they conform to standard common-practice notions of grammaticality. And the important point is that it is only because of this that these phrases are interpretable – musical

interpretation can occur only when grammar maps successfully to extra-grammatical systems, such as those pertaining to musical meaning or rhythm.

There is one aspect of internal structure generation and interpretation in music that we have not dealt with yet, in our current exploration of Chopin's mazurkas. As mentioned before, this is the issue of musical rhythm, and its associated phenomenon of rhythmic (or more generally "durational") interpretation, which is something we discussed in the last chapter. And given the importance of rhythm in dance forms like the mazurka, an internalist discussion of mazurka structure and meaning that does not deal with its rhythmic character will be considered quite unsatisfactory. This is especially true given that it is in its rhythmic aspects that the mazurka seems most functionally or culturally signified too, especially as a social dance form. So, I now turn to a discussion of rhythm and meter in the mazurkas. For this purpose, I will revisit the above generative procedure that demonstrated how passages from the Op. 67 No. 3 and Op. 24 No. 2 mazurkas are generated. Interestingly, Kofi Agawu's similar generation of phrases from the Bach chorale "Ach Gott, wie manches Herzeleid" did not examine the rhythmic aspects of that piece. As discussed in chapters 2.1 and 2.2, the reason for this lies in the different kinds of interpretation inherent in Agawu's generative approach, which deals with musical meaning, in a manner similar to how linguists discuss Logical Form – which is separate from the generative approach that deals with rhythm and meter, such as Carl Schachter's, which seems more akin to the idea of phonetic and prosodic interpretation in language, inherent in the notion of Phonetic Form. Therefore, we will now add a discussion of rhythm to our earlier semantic investigation of the Op. 67 No. 3 mazurka, within the larger generative description of this piece.

In chapter 2.2 we saw that rhythm and meter in music can be discussed under the umbrella of *durational interpretation*. That is, they can be understood in terms of how the pitch-grammatical content of a musical surface is interpreted durationally, which then allows the surface to be performed in real time. Metrical durational interpretation can affect fairly large sections of the tonal content of the piece because of the way the durational aspects of pitch structure interact with hypermeter (i.e. in phrase rhythm), of which we saw an example in our study of the last movement of Beethoven's "Tempest" piano

sonata in the previous chapter. However, rhythmic durational interpretation is usually of interest only within small spans, usually just a metrical cycle, because the rhythmic value of pitches that span more than a measure is usually not considered to be of much interest, or is subsumed within a study of meter. For this reason, I shall begin the discussion of durational interpretation in the mazurkas with a discussion of metrical interpretation. However, the surface rhythms of a mazurka are considered one of its most characteristic aspects, as just mentioned, so I will not end without discussing that aspect of mazurka structure too.

We saw in the last chapter how the binary-branching nature of tonal grammar, and the inherent rhythmic properties this endows tonality with (in the form of tonal rhythm), influences how durational rhythm emerges and metrical structure is interpreted. How does this affect the metrical structure of the mazurkas? Example 2.3-5 shows how this happens for mm. 1-8 of Chopin's Op. 67 No. 3 mazurka. We have seen in Examples 2.3-2 and 2.3-3 that the harmonic basis for this passage is the headed tree-structure shown in level A of Example 2.3-5. At this level of structure, the two chords in level A do not have durations – their sole rhythmic properties are the ones associated with tonal rhythm. So, when it comes to assigning durations to the passage the most tonal-rhythmically faithful one will be the one depicted in level B of the example – namely, a structure in which the entire passage is divided into two durational parts, corresponding to the two halves of the binary tree-structure that generated this passage. Importantly, the two parts do not have a specific duration yet. At a given tempo, they could each last for as many measures as the composer chooses to expand them for. However, as we have seen earlier, the number of measures that one part lasts for has to be equaled by the number of measures the other part lasts for too, to meet conditions of *metrical* well formedness.

This is the intuition behind the notation “whole note = $2 \times n$ measures in any time signature” in level B. If the duration of the entire passage is notated as a whole note (divided into the two half-note tonic chords from which the passage is generated), then however many (i.e. n) measures each chord is

Example 2.3-5. Chopin, Mazurka in C major, Op. 67 No. 3: Metrical interpretation, mm. 1-8

A

I I

Durational Interpretation: Background ($\circ = 2 \times n$ measures in any time signature)

B

I I

Durational Interpretation: Middleground ($\circ = n$ measures in any time signature)

C

I ————— V⁵/V V I

Durational Interpretation: Foreground

D

expanded for, the other chord, i.e. the other durational event, must last the same number of measures. Therefore, if a phrase lasts for four measures, say a typical antecedent phrase, the consequent phrase will also be four measures – and the whole passage will be $2 \times 4 = 8$ measures. Of course, the consequent phrase might be longer than four measures – but as discussed in the last chapter, this is often the result of a phrase expansion at a ‘shallower’ level of structure, whose deviation from the norm can be shown by a durational reduction.

As we saw in Example 2.3-2, the initial tonic of the C-major mazurka is expanded from one vertical sonority to four by means of the voice leading process described in level C of that example. So, in the mazurka passage, we have four structural accents corresponding to the four vertical sonorities generated by tonal grammar, which would imply four durational spans (and four is a multiple of two!). This is what level C portrays. The barlines between the chords represent durational spans of $n/4$ measures since there are four chords in the initial tonic prolongation, and the prolongation itself is n measures long. (Remember that duration, in this case the number of measures, is not determined by tonal structure, only structural accents are. So, the initial tonic prolongation could last any number of measures, as long as it distributes the accents within it evenly to align with meter. If the passage is four measures long, each chord would last a measure each. If it is eight measures long, each chord would last for two measures each, and so on. Also, “multiples of two” includes fractions, so if the passage is two measures long, each chord would occupy only half a measure.)

Moreover, the durational length of the initial tonic prolongation must be matched in length by the second half of the passage, at least at the deeper levels of structure. Unsurprisingly, the second half of the mazurka passage is made up of four chords as well, which form the basis for an expanded cadential progression that closes the passage. This cadential progression should be *ex hypothesi* of the same length as the initial progression and each chord in it should have a duration of $n/4$ measures too. As we see in level D of Example 2.3-5, this is exactly what happens in the metrical surface of the piece – each half of the passage (which prolongs one each of the two tonic chords that generate this passage) has a duration of four measures, and each chord within it a duration of $4/4$, i.e. one, measure.

Let us now turn to a description of metrical interpretation in the other mazurka we have looked at so far, the C major mazurka, Op. 24 No. 2. Even though the process of metrical interpretation is the same here as the previous example, mm. 5-20 of this piece show us a few new things that the last piece did not. Example 2.3-6 illustrates the metrical structure of these measures. As usual, tonal generation here starts with two tonic C major chords in a headed binary-branching relationship, as shown in level A of the example. However, these chords are prolonged in a slightly different way than they were in Op. 67 No. 3. Though the latter chord is expanded by means of a standard dominant harmony, the initial tonic chord is expanded by means of an actual functional submediant harmony this time, not the voice leading progression we saw in Op. 67 No. 3. This is shown in level B of the example.

When it comes to assigning durations to these chords, we know from our preceding discussion that we should expect the whole passage to be divided into durational spans in multiples of two, which is predicted by the binarity of the tonal structures that generate the passage. In Example 2.3-5, each chord was assigned a measure's worth of duration (at any given tempo) by Chopin. However, in Op. 24 No. 2, he assigns each structural chord *two* bars worth of duration, as seen in level C of Example 2.3-6. This is done by means of further prolonging operations, depicted in level D. In the upper system of level D, we see how the I chord is given two measures worth of durational rhythm by prolonging the tonic with a $vi - IV^6 - vii^0 - I$ chord progression. The following vi is then prolonged for two measures by means of a $ii - ii^6 - V^7/vi - vi$ chord progression. The second half of the passage sees V being prolonged for two measures with a $V - ii - V - ii$ chord progression, and then tonic harmony is prolonged for the final two measures by means of a $ii - V^7 - I$ cadential progression (see the second system in level D). But the metrical play of forms is not yet done. Chopin chooses to repeat the first half of the passage immediately after it is played for the first time *before* going on to the second half. Again, following the "2 x n" rule of Example 2.3-5 (which links tonal and durational rhythm), Chopin ends up repeating the second half of the passage too, presumably to resolve the asymmetry that repeating the first half caused. This is illustrated by the "x 2" symbols in level D of Example 2.3-6.

Example 2.3-6. Chopin, Mazurka in C major, Op. 24 No. 2: Metrical interpretation, mm. 5-20

A

I I

B

I vi V I

C

I — vi — V — I

8 — 7

D

I — vi —

piu f

V — ii — V⁷ — I

Something particularly important emerges from the above description of the metrical structure of mm. 5-20. This concerns in particular the final, tonic-affirming cadential progression in it. Note how the durational interpretation of this cadential progression, i.e. as a 2-bar structure within the 2 + 2 + 2 + 2 metrical design of mm. 5-20, pairs up with its *semantic* interpretation as an ending. We already saw how such a pairing seems to happen in the Op. 67 No. 3 mazurka, and here, in the Op. 24 No. 1 mazurka, we see how the durational aspects of music participate in this phenomenon. Which makes sense, given that the durational aspects of music were described in the last chapter as being part of the ‘sound’ (i.e. phonological) aspects of music, to which musical grammar maps at the musical equivalent of a PF level of structure. And the convergence between PF and LF structures is just another way of describing a pairing between *sound* and *meaning* – so to the extent that durational and semantic interpretation are paired in music, as seems to happen in the mazurka being discussed, this acts as just more evidence for there being a sound-meaning pairing in music, just like in language.

So far we have explored the relation between tonal and durational rhythm, and how this leads to the emergence of surface metrical structure, particularly structures that are organized in multiples of two. We have also explored the more general idea that this characterization of how meter arises, and is durationally interpreted, in music, conforms to certain Minimalist theses about structure generation, interpretation, and sound-meaning pairings in music and language

However, this discussion has a specific significance in the context of Chopin’s music, because the existence of a simple, binary metrical organization has been seen as major property of Chopin’s *style* – and not just a general property of grammar-to-rhythm mapping in tonal music. William Rothstein has noted that Chopin’s music largely exhibits regular four- or eight-bar phrases and tends to avoid phrase expansions or other devices that disrupt the propensities of tonal rhythm. Instead, rhythmic variety in his music comes more from playing with patterns of articulation on the musical surface to weaken phrase regularity (Rothstein (1989): 214-248). Several authors note the simple 2-bar, repeated motivic structure of Polish dances like the *oberek*, which were appropriated by Chopin in the mazurkas too. Finally, Charles Rosen says that Chopin’s mazurkas employ:

“...the most conventional forms of art music [such as the A-A-B-A form of Op. 67 No. 3], and, ... a rigid sectional structure ..., characteristics quite normal and proper for dance music. Both these characteristics are associated with the mazurkas throughout Chopin’s life, but he begins very early on to play with both the conventional form and the sectional phrase structure in increasingly subtle ways. It is, in fact, in studying the mazurka that one finds the greatest sense of gradual progress and stylistic change in his work.” (Rosen (1995): 423)

In fact, in some of his (especially earlier) mazurkas, Chopin often lets a simple formal and metrical design, created by just a few generative steps, ‘speak for itself’ – i.e., he does not adorn the piece in the imaginative melodic or rhythmic ways that we have encountered in the mazurkas discussed so far. Such simplicity might lull someone into believing that the piece is incomplete in some way – and this is often understood in externalist terms, i.e. as a sign of its folk origins. However, Carl Schachter says:

“As is well known, many Chopin mazurkas reflect the folk origins of the genre through their lack of a strong sense of closure at the end. Mostly these open-ended pieces have in fact achieved a structural cadence and open up only after having done so ... even the amusingly folk-like and fragmentary C major Mazurka, Op. 7, No. 5, marked “Dal segno senza Fine,” has a normal cadence in bar 12 that would sound more or less like an ending even if the pianist decided to stop somewhere else.” (Schachter (1999d): 303)

He also remarks that such perceived open-endedness (in another mazurka):

“...keeps alive the notion of the dance as a pattern that never really ends, even though any particular manifestation of it will have to stop at some point. In a very general way, therefore, the Mazurka’s suppressed cadence is programmatic.” (Schachter (1999d): 306)

This suggests that the externalist assumption about the folk origins of the mazurkas seems to be contingent on how their *grammatical* structure conditions the interpretations we assign to them. Which is fully consistent with the internalist approach to grammar and meaning in the mazurkas we have been exploring so far. Let us look at how grammatical generation actually occurs in the piece Schachter cites, viz. the C major mazurka, Op. 7 No. 5, and thus fosters the above impressions of the piece. Example 2.3-7 describes this phenomenon.

As before, level A starts with the basic I – V – I progression in C major we have seen in the earlier mazurka derivations. Level B replaces the Gs in the two tonic chords with Cs, thus tripling their roots, and revoices all the chords. In fact, we could have started the derivation at this level – level A was

Example 2.3-7. Chopin, Mazurka in C major, Op. 7 No. 5: Tonal generation, mm. 5-20

A

I V I

B

I V I

C

$\begin{matrix} \hat{3} & \hat{2} & \hat{2} & \hat{1} \\ 7 & 6 & 7 \\ 5 & 6 & 7 \\ 3 & 4 & 3 \end{matrix}$

I V⁷ I

mm. 5/6 7 8 9/10 11 12

D

* > Ped. * > Ped. *

Example 2.3-7 (contd). Chopin, Mazurka in C major, Op. 7 No. 5: Tonal generation

mm. 13/14 15 16 17/18 19 20

E

G major: _____ I V _____ I

C major: _____ V V/V _____ V

really provided to show the link between this derivation and the previous ones. The next step of the derivation, level C, shows a compositional technique we have not seen before, viz. the off-tonic opening. In this technique, a piece begins with a non-tonic harmony, in this case the dominant, which we only realize *is* the dominant later, when it resolves to the real tonic of the piece. In level C, we see that the initial tonic harmony is not prolonged by chords that follow it, as in the previous derivations, but by a G^7 harmony that comes *before* it, in measure 5 (after a 4-bar introduction prior to this). Of course, the 7th (i.e. the F-natural) of this G^7 chord clarifies the role of this entire G^7 harmony as a *dominant* seventh chord that prolongs the following C-major *tonic* harmony. However, the 4-bar introduction prior to the G^7 harmony (not shown in the graph) consists of open octaves in the left hand that just play the note G, and therefore add weight to a non-tonic understanding of the opening of the piece. (Schenker terms such a phenomenon an “auxiliary cadence” or *Hilfskadenz*, see Schenker (1979) : 88-90.)

Even before we realize that the G^7 prolongs the following C-major harmony, the latter being the initial tonic of the *Ursatz* in level A of Example 2.3-7, the G^7 is itself prolonged, via a neighboring 6-4 motion over the bass of the G^7 chord. During this neighboring motion, one of the upper voices of the G^7 chord, viz. the pitch G4, is transposed up an octave to G5. Level C of Example 2.3-7 shows this with the

diagonal line at the beginning of the system. So, when the G^7 chord eventually resolves to the C-major triad that initiates the *Ursatz*, this G5 descends to the E5 pitch in the soprano of that C-major chord, which we now understand to be the *Kopfton* of the piece. (This E5 is in the soprano of the initial C-major chord of the *Ursatz* in levels A and B too, which confirms its *Kopfton* status.)

After the initial tonic chord has been reached, the rest of the derivation in level C proceeds in the standard manner. The G^7 chord in measure 9 that appears after the initial tonic triad is the ‘structural’ dominant chord of the piece, since it is derived from the G-major triad in the *Ursatz* of levels A and B, and supports the scale degree 2 pitch D5 in the *Urlinie*. In level C it is prolonged by means of a similar 6-4 neighboring motion as the first G^7 chord in the piece was, but given its deeper structural status as a member of the *Ursatz* (and having been derived earlier in the generative process), the 6-4 sonority that prolongs it has a greater structural status than the earlier 6-4 sonority too. Therefore, the “6” of this 6-4 sonority, i.e. the soprano pitch E5 in measure 9 that prolongs the D5 of the *Urlinie*, is notated with an open note head with flagged stem, which symbolizes its greater structural, yet neighboring, status. Also, the G^7 chord in measure 9 again has an upper voice G4 that is transposed up an octave to G5, as happened with the G^7 chord at the beginning of the piece. However, this G5 does not descend to the E5 of the *Urlinie* any more, since the *Urlinie* has already descended to D5 at this point, the D5 being supported by the same structural G^7 harmony that supports the G5 pitch. Therefore, the G5 becomes part of an inner-voice descending line G5-F5-E5 that reaches over the *Urlinie*. You can see this in Level C in the slur over the final scale degree 2 to 1, D5-C5, descent in the *Urlinie* at the end of the system. Finally, this descent in the *Urlinie* is harmonized by the structural G^7 harmony of measure 9 and its resolution to the final tonic triad of the *Ursatz* in measure 12. With this, the *Ursatz* is complete and the passage has reached a perfect authentic close.

Level D of Example 2.3-7 shows the actual music of mm. 5-12, whose derivation we explored in levels A, B, and C. The *Urlinie* of the piece, which traces the descending line E5-D5-C5, has been sketched into level D for comparison with level C. Notice something curious about the actual music – the music for mm. 5-8 is almost exactly the same as that of mm. 9-12, apart from the last measure in each

group – but even that measure is almost identical if you see how the last three notes of mm. 8 and 12 are the same, but transposed an octave. This seems to illustrate the simple, folk-like structure of the piece.

But this simplicity is deceptive though. Since mm. 5-8 and 9-12 are nearly identical, it is possible that on hearing the entirety of mm. 5-12, one will hear it as two near-identical passages, each harmonized by a $V^7 - I$ progression, and therefore equally open or closed – rather than as a large $I - V - I$ progression with a clear structural beginning at measure 8 (with the initial, *Kopfton*-harmonizing, C-major triad of the *Ursatz*), and a clear ending at measure 12 (with the final tonic triad of the *Ursatz* that supports the C5 in the *Urlinie*).⁵ This indecision is heightened when we examine the remaining measures of the piece, shown in level E of Example 2.3-7. These measures repeat almost exactly the music of mm. 5-12 discussed in levels A-D of the example, except that they are transposed down a fourth. So, do we interpret these measures in a manner similar to mm. 5-12, i.e. as a $I - V - I$ progression, but transposed down to G major, and therefore as closed in G major as mm. 5-12 are in C major – or do we hear the whole piece in C major, and therefore this new passage in mm. 13-20 as an (open) $V - V/V - V$ progression in C major too? In light of the ambiguity between C and G major that exists from the beginning of this piece, especially given its non-tonic opening, these questions seem almost unanswerable – all of which adds to the apparent open-endedness of the piece and, most importantly, its interpretation as a never-ending dance.⁶

However, as just the previous sentence demonstrates, these questions all have to do with how we *interpret* the structure of the mazurka. That is, given the surface of this passage, how do we *understand* its structure on perceiving it? But interpretation is subject to all kinds of vagaries, as we have discussed before. And

⁵ The latter reading implies that mm. 5-12 can be interpreted formally (and therefore semantically, following the discussion in chapter 2.1) as an 8-bar period, with a 4-bar antecedent phrase (mm. 5-8), and a 4-bar consequent phrase (mm 9-12), the former ending with an imperfect authentic cadence in measure 8, and the latter with a perfect authentic cadence in measure 12.

⁶ Notice that the question of how to relate mm. 13-20 to mm 5-12 is a question of ‘inter-sentential’ discursive meaning. After all, these two passages are well formed in and of themselves, as shown by the well-formed derivation in Example 2.3-7. So, the question of how to relate the two is a discursive one. This brings up something significant – even if we do interpret the piece as being open-ended discursively, we can only do that because the two phrases are closed grammatically, and can therefore be related to each other as either bringing the piece to a close or leaving it open.

moreover, the above questions are from ‘the top down’ – asked *after* the fact of the piece’s generation. That is, they are not *cognitive* questions that deal with our *knowledge* of the piece, particularly of how it arose and of the tonal grammar that generated it. But having knowledge of a piece’s grammatical origin is critical when interpreting it, because it is its grammatical structure that makes a piece *interpretable*. So, when one asks *these* questions, about grammar and how this affects *interpretability* (as opposed to *interpretation*, which is indeterminate) much of the confusion about a piece’s structure disappears – and it is in this context that we can understand statements like Schachter’s claim that the Op. 7 No. 5 mazurka *does* have a clear ending in measure 12, because of the specific grammatical structure it has, and as we explored in our generation of this structure from levels A through D of Example 2.3-7. This is why a grammatical approach to music is so important – it can clarify issues of musical structure and meaning in a way that a theory based on extra-musical signifiers never can.

In chapter 2.1, we saw how the sentence “Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo” is perfectly interpretable, being perfectly grammatical, but is still hard *to* interpret without a knowledge of the sentence’s grammatical structure. Part of this has to do with the complicated structure of the sentence, generated recursively from just one word, which makes it hard to interpret. The remarkable thing about the Op. 7 No. 5 mazurka is that it presents exactly this kind of a situation. It has a simple structure, made up of essentially two chords (a C and a G major one), prompting much talk of folk roots and such. However, just these two chords are prolonged at different levels of derivation by *themselves*, to give the piece the simple, yet highly recursive structure it has – which is exactly why it is so hard to interpret, leading to claims of open-endedness and eternal ‘dancibility’ – even though it is perfectly interpretable, as long as one *knows* the grammatical origins of the piece, which Example 2.3-7 has now illustrated for us in detail.

While we are on Op. 7 No. 5, let us look at its metrical structure too. For brevity’s sake, I will just discuss the meter of mm. 5-8, which contains the main melodic figure of the piece, which is repeated in mm. 9-12, and then transposed in mm. 13-20 as described above. Example 2.3-8 describes how the structure of

these measures can be metrically interpreted. Level A of the example shows us the binary tree that generates it, which corresponds to the first half of level C in Example 2.3-7. The middle voice F3 in the initial V^7 chord has been left out of this level, since the F pitch class of this harmony, as is true of most sevenths in seventh chords, arises at a more surface level of generation, normally through 8-7 motion with the bass. For this reason, this 8-7 motion is depicted in level B of the example instead, where we see it in the soprano. As a result of this, a *V triad* is shown to be what is really prolonged in mm. 5-6, with this triad being transformed into a V^7 only in measure 7 with the descent of the soprano G5 of the chord to F5 (not F3), on the way to the E5 *Kopfton* of the piece in measure 8, which initiates the *Urlinie* of the piece (which was described in levels C and D of Example 2.3-7).

The result of all of the above is that the four measures of mm. 5-8 that we have been considering have a 2 + 2 division – the first two bars prolonging V, and the next two prolonging I, the latter by means of the V^7 harmony in measure 7 that resolves to I in measure 8. And this is what the tonal rhythm of the passage (and the “2 x n” rule) predicts too, as shown in level B of Example 2.3-8. Which means that it is easy to interpret the metrical structure of this passage in binary terms, as the principles of metrical well formedness would have us do anyway. So, from here, to level C, where the actual score is shown, is a short step. And given the repetitive, recursive structure of this piece, this simple metrical interpretation just reflects how easily the rest of the meter of the piece arises – which is why it is also easy to see why so many have speculated about this piece’s dance-like character and folk origins.

The long discussion we have had so far about musical generation and interpretation – whether metrical, melodic, or semantic/formal – and all seen from an internalist perspective, was but a necessary step to bring us to the next, and probably most significant part of our journey through Chopin’s mazurkas, viz. a discussion of rhythm in the mazurka. As mentioned before, rhythmic interpretation is normally a surface phenomenon, something that is only of interest within bar lines. So, to have a discussion of what happens

Example 2.3-8. Chopin, Mazurka in C major, Op. 7 No. 5: Metrical interpretation, mm. 5-8

The image displays three staves of musical notation for Chopin's Mazurka in C major, Op. 7 No. 5, measures 5-8.
Staff A: Shows the harmonic structure. The treble clef has a whole note chord (V) in measure 5 and a whole note chord (I) in measure 8. The bass clef has a whole note chord (V) in measure 5 and a whole note chord (I) in measure 8. A line connects the V chords, and another line connects the I chords.
Staff B: Shows the metrical interpretation. The treble clef has a whole note chord (V) in measure 5 and a whole note chord (I) in measure 8. The bass clef has a whole note chord (V) in measure 5 and a whole note chord (I) in measure 8. A line connects the V chords, and another line connects the I chords. Below the staff, a line indicates the 8-measure phrase and a 7-measure phrase.
Staff C: Shows the melodic and harmonic details. The treble clef has a melody with triplets and accents. The bass clef has a harmonic accompaniment with triplets and accents. Dynamics include *fz* and *fz*. Articulation includes *Red.* and **.*

within bar lines, we needed to have the discussion of how bar lines come about first – tonally and metrically. Having had that discussion, I think the proper context has been now created for discussing surface mazurka rhythm.

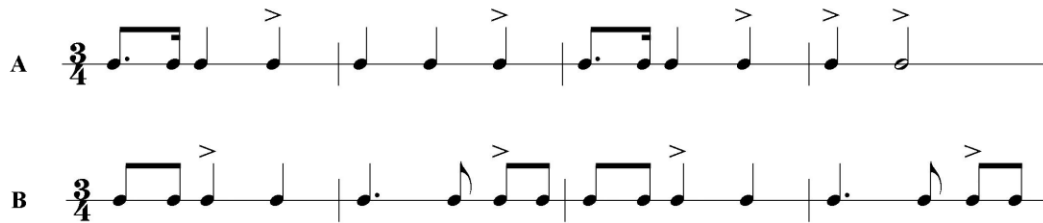
The undeniable feature of mazurka rhythm is that it is a dance rhythm. There also seems to be a consensus that it has much in common with the rhythm of a number of other, specifically Polish, folk dances too, such as the fast *oberek*, the medium-tempo *mazur* and the slower *kujawiak*, Jane Bellingham going so far as to say in her Grove article on the mazurka that the term “mazurka” encompasses all these

dances. (Indeed, Chopin's mazurkas have elements of all of these dances in them, though they only go by the name "mazurka".) In his Grove entry on the mazurka, Stephen Downes actually specifies a basic mazurka rhythm of two eighth notes followed by two quarters in triple time, and says that mazurka rhythms are to be found in the above dances – thus defining all of them as mazurkas implicitly.

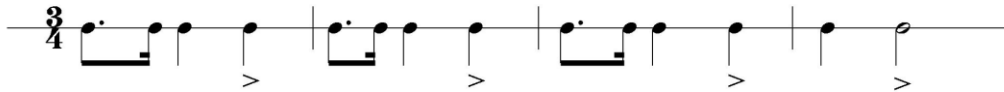
Even though the geographical origin and the folk dance exemplars of the mazurka are pretty well documented, defining the concept of a mazurka rhythmically is a different matter all together. It is without question in triple *meter*, but how the three beats of this meter are divided in the prototypical mazurka is a question with no definitive answer. Charles Rosen even says, "accounts of mazurka rhythms are comically confused" (Rosen (1995): 412). However, Rosen believes that the division of beats or accents is not unsystematic, and both Stephen Downes and Rosen give accounts of what they consider to be common mazurka rhythms, reproduced here in Examples 2.3-9 and 2.3-10. So, the concept "mazurka" seems to be based, as concepts usually are, on the family resemblances between the various 'field' exemplars of this dance in which the above common mazurka rhythms can be found. But we saw the problems inherent in such conceptual definitions in chapter 2.1 – and as we also saw then, this will not help us answer the question about what the various exemplars of the mazurka do have in common. Just looking at Downes' and Rosen's examples of mazurka rhythm below, we see that Rosen's 'systematic' account of the stress patterns in the dance rhythm (i.e. stresses on the third beat, except for the last measure where it brings the dance to a close by falling on the second beat) are immediately contradicted by Downes' accents in level A of Example 2.3-9, where stresses occur on all three beats. Even the division of the beats in Rosen's account, in which divisions less than a quarter-note long all occur on the first beat, is contradicted by level B of Downes' example, in which eighth-note patterns occur on the third beat. And all of this in spite of Downes' clear characterization of the basic mazurka rhythm as eighth-eighth-quarter-quarter in 3/4 meter, and the following statement he makes about this rhythm:

"The basic mazurka rhythm shifts the accent to the weak beats of the bar within a triple metre. Triple metres became dominant in Polish folk music in the 17th and 18th centuries, while the displacement of the accent may have its origins in the paroxytonic accent in the Polish language."

Example 2.3-9. Characteristic mazurka rhythms from Downes' Grove entry "Mazurka"



Example 2.3-10. Common mazurka rhythm from Rosen (1995)



The confusion is not just limited to Downes' and Rosen's accounts. Gerald Abraham (1939) says that the "three major types of mazurka [i.e. *oberek*, *mazur*, and *kujawiak*] are all characterized by the accent on the third beat", though he adds that, "the *mazur* takes the accent on the second beat" according to some sources. Anne Swartz (1983): 417-418, says,

"Folk mazurkas share certain common musical characteristics: triple meter with accents on the weaker second or third beats of the measure, short repeated rhythmic and melodic motives, ornamented melodies with grace notes, trills and mordents, and use of drone instruments such as the *duda*, a bagpipe-like instrument, to accompany the dances."

However, she also says,

"Probably the oldest Polish folk mazurka is the *mazur* which originated near Warsaw. This mazurka is distinguished from the other folk dances by its very free pattern of accents. According to earlier studies by Polinski, the accents in the *mazur* may fall anywhere in the measure, and any number of accents may be found in each measure."

That there is so much disagreement over how to define accentual patterns in a mazurka is probably not very surprising given the nature of rhythmic accents. Unlike structural or metrical accents, rhythmic

accents are the most ‘surface-level’ accents of them all, i.e. they are the most marked and most susceptible to interpretive choice. Lerdahl and Jackendoff call such accents “phenomenal accents” (Lerdahl and Jackendoff (1983): 17), and treat them as less important for musical structure than structural or metrical accents. So, Charles Rosen’s assumption that mazurka rhythmic accents are systematic, and Stephen Downes’ assertion of what this systematic mazurka rhythm is, do not have to be true.

Given all of this confusion over the definition of “mazurka”, or at best the ‘family resemblance’ nature of the concept, is there any point in theorizing mazurka rhythm at all then – or for that matter any social aspect of music external to the music itself, such as the very idea that certain musical pieces signify certain dances? In the internalist context I have been following here, I do not think there is. In fact, Ludwig Wittgenstein, with whom the notion of family resemblances is most associated, had this to say about “dance music”:

“Are we supposed to imagine the dance, or whatever it may be, while we listen? If seeing the dance is what is important, it would be better to perform that rather than the music. But that is all misunderstanding.” (Wittgenstein (1998): 69e)

So, it seems that we may never find any consensus about how to define “mazurka”. This seems inevitable given the problems inherent in defining concepts in general, whether in music or language. Therefore, it is not surprising that mazurka rhythm, when defined in terms of some division of triple meter, manifests itself in reality in countless ways, and indeed almost every page of Chopin’s mazurkas show some new way of dividing the triple meter, which endows almost each and every mazurka theme he composed with a different rhythmic character. So, the standard mazurka pattern (the top figure in Example 2.3-11) described in several places in the literature – including in Stephen Downes’ Grove article – is quite probably not prototypical at all.

But the above does not mean that we cannot *use* the word “mazurka” in the above context of dancing, meter, and phenomenal accents. As was the case with the word “field” discussed in chapter 2.1, we can easily use a word in certain communicative contexts, without necessarily having a *theory of* what that word means. I am only challenging the idea of a theory of mazurka, which attempts to define the term

in a specific way (especially in terms of its rhythmic characteristics), outside the internalist description of how mazurka phrases are generated, and then semantically and ‘phonetically’ interpreted. In fact, a lot of the extra-musical phenomena that are discussed in the context of music are quite possibly worth studying, but probably not within a *theory* of music. So, the cultural practices that are often described as signifying music extrinsically might better be studied in the form of a history, sociology or anthropology of music – or even within the context of music appreciation, as I argued in chapter 2.1 for the extra-musical phenomena known as musical *topoi* or “topics”. Our present study of mazurka rhythm is a valuable study too but perhaps better suited to a theory of dance than a theory of music.

Example 2.3-11. The ‘Generic’ Mazurka

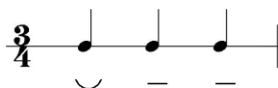
Standard Mazurka Pattern



Holupiec Pattern



'Generic' Mazurka



In this respect, consider the fact that many accounts of the mazurka describe a foot stomping gesture on the second and third beats known as the *holupiec* (shown in the middle system of Example 2.3-11). In his Grove article, Stephen Downes says one of the characteristics of mazurka dancing is, “accents and dynamics emphasizing the gestures of the dancers, especially stamping or heel-clicking leaps (*holubiec*) on the displaced accents on the second or third beats of the bar”. Barbara Milewski seconds this, saying:

“In an *oberek* this *holupiec* gesture – a measure of three eighth notes sung or played in the same pitch in 3/8 meter – accompanies the heel stomping of dancers that can mark the beginning or ending of a dance. It is one of the more recognizable features of indigenous Polish music because it is inherently dramatic; its three punctuated, repeated notes at the beginning of the dance signal the listeners and dancers to attention, while at the end of a dance these same repeated notes articulate closure.” (Milewski (1999): 119-120)

Milweski’s description differs from Downes’ in that she points out the marking of the first beat in a *holupiec* too. However, both these accounts of the *holupiec* intersect with one idea in Charles Rosen’s account above, viz. the observation that a heel click or stomp on the second beat can signal the end of the dance. This is even more significant since the end of the dance would be marked tonally, by a cadence – and an example of a traditional folk *holupiec* Milewski discusses in her above text stresses the two most important pitches of the scale, scale degree 1 and 5, by the foot-stomping at the end of the dance. In addition to all of this, the third beat in all but the last measure signifies the continuation of the dance, as Rosen suggests, often through a hopping dance step (which Swartz describes as well, in addition to describing a characteristically *holupiec*-like gesture at the end of the dance (Swartz (1983): 419).

So, in consideration of how dancers actually dance the various forms of the mazurka, certain rhythmic stress patterns make sense, particularly given how they can cue the dancers to what part of the dance they are in. Lerdahl and Jackendoff say that phenomenal accents have one important role in musical structure, that of cueing a listener in to the *metrical* structure of a piece (see their Metrical Preference Rule 4 (Lerdahl and Jackendoff (1983): 79)). So, if there is any case in which rhythmic accents occur systematically in the mazurka, it would be at least partly to cue dancers to the meter of the dance. (I described this in a slightly different way in chapter 2.2, when I suggested that a musical equivalent to the notion of PF in language might lie in the mapping of pitch grammar to rhythmic accents in music – which then makes a generated musical surface *metrically interpretable*.) Of course, mazurka rhythms never stress the downbeat of the bar to the exclusion of the two other beats, which would be Lerdahl and Jackendoff’s ideal way of phenomenal accents cuing one to the meter. And even if they happened systematically somewhere else in the measure that would cue the listener to a syncopation against the meter – but syncopation is not a major feature of mazurkas either. But the fact that mazurka

stresses can act as important indicators of how to dance, especially in light of the observation that stresses invariably occur on either the second and/or third beat (depending on which source we are reading), suggests that a definition of “mazurka” has teeth within a theory of dancing, and within that of meter as it relates to a dancer’s footsteps while dancing.

With this in mind, it might be worth thinking of mazurka, and specifically mazurka rhythm, more generally, i.e. not within a theory of mazurka, as a tangible musical concept, but as a word of indeterminate meaning, which we happen to use in discussing certain extra-musical social activities (such as dancing) – which could then help us understand certain characteristics in musical pieces such as the Chopin mazurkas, insofar as these pieces relate to dancing in some way. This is opposed to a theory of mazurka, which attempts to define “mazurka” in a specific way, and therefore expects that any musical piece that is referred to by this name have special, identifiable characteristics, such as those that supposedly underlie mazurka rhythm. In this light, it might even make sense to speak of a ‘generic’ mazurka, a pattern in which a triple meter is not divided in any specific durational way, but is marked in a way that reflects how dancers dance. Importantly, this means that such a generic mazurka *does not have to correlate with how musicians make music at all* – a generic mazurka does not generate anything, it does not govern how the surfaces of certain musical pieces that happen to be called “mazurka”, such as those by Chopin, are derived. And given our above inability to define “mazurka”, a generic mazurka does not even govern how we (semantically) interpret such pieces, i.e. as being “mazurkas”. In other words, it is not even necessarily a mazurka, in the sense of being a semantic construct with a certain meaning. Rather, it is more of a *pragmatic* construct – that might help us understand how and why the word “mazurka” is used to discuss certain musical pieces.⁷

⁷ This point refers back to a point made in chapter 2.1 about how our inability to define terms such as “field” in language (and “mazurka” in music), leads to the possibility of there being so such thing as a “semantics”, i.e. a theory of meaning for language, and presumably music too. This might lead one to assert that the study of language or music should just involve a study of grammar, and a study of pragmatics, i.e. how grammatically-generated structures are used in a variety of extra-linguistic or extra-musical contexts – an assertion that Noam Chomsky has explicitly made for language (Chomsky (2000): 132).

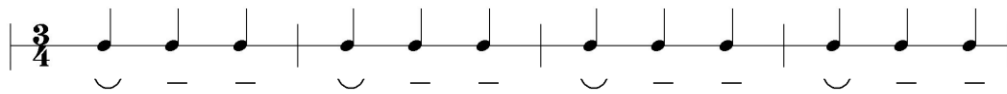
The third figure in Example 2.3-11 depicts such a generic mazurka, as a pattern with a weak or short accent, and one or two strong or long accents, the division between the types of accent occurring between the first beat, and the second and third beats. Even the standard mazurka pattern discussed in the literature can be seen as a derivative of this generic pattern, in the way it deemphasizes the first beat (by dividing it into two eighths) and stresses the next two beats in quasi-*holupiec* fashion. When we tie these two beats together (as shown in the top figure of Example 2.3-11), we even get the second-beat-stressed closing pattern that ends a dance, as described by Charles Rosen in Example 2.3-10.

As mentioned above, this might help us understand how certain musical pieces, specifically the Chopin mazurkas we have been discussing, might be understood pragmatically, i.e. in dancing – and more importantly, this might shed light on the elusive rhythmic characteristics of these mazurkas, for which we have still not provided a compelling account. Example 2.3-12 illustrates how we might understand the Op. 7 No. 5 and Op. 67 No. 3 mazurkas in this respect. At the top of this example, we just see the generic mazurka described in Example 2.3-11. Level A underneath this discusses the Op. 7 No. 5 mazurka. The lower stave of this level shows the score of the 4-bar main idea of the mazurka (i.e. mm. 5-8), and the upper stave depicts the specific rhythm this idea realizes, i.e. the specific way it realizes the rhythmic pattern of the generic mazurka. (As mentioned in an earlier footnote, this 4-bar idea can be understood formally as an antecedent phrase, ending with an imperfect authentic cadence in measure 8. But this requires understanding mm. 5-12 as a complete structure, viz. an 8-bar period, which has problems of its own, given the ‘open-ended’ interpretation some might give these measures – hence I will persist with the “4-bar main idea” label for this passage. Also, our discussion here will only concern these four measures, because the rest of the mazurka has the same surface rhythmic structure as these measures, in addition to a near-identical tonal structure, achieved through recursion as discussed earlier.)

As we see in both staves of level A, the 4-bar idea divides into two smaller 2-bar groups, which have the appearance of a statement and a response. (Again, these 2-bar groups could be seen as manifesting the basic and contrasting ideas of a 4-bar antecedent phrase, but I will desist from using this

Example 2.3-12. Rhythmic structure in Chopin Mazurkas Op. 7 No. 5 & Op. 67 No. 3

Generic Mazurka



Chopin Mazurka, Op. 7 #5, Main Idea

A

Chopin Mazurka, Op. 67 #3, Compound Antecedent phrase

a. mm. 1-2 & 5-6 b. 3-4 & 7-8

B

description for the reasons just mentioned – especially given the non-tonic opening of the piece, which leads to a deviant non-tonic prolonging harmony for the basic idea.) This specific 2-bar grouping is particularly amenable to the quick foot switching required in fast dances, and is indeed characteristic of the fast *oberek* dance (Swartz (1983): 423). However, the two 2-bar groups each realize the generic mazurka pattern in considerably different ways. Their accents are most idiosyncratic – the accents in the statement are on the *second* beat, while the accents in the two measures of the response are on the *first* and *last* beats, respectively – the opposite of what we have come to expect from mazurka accent patterns.

On the one hand this shows the truly phenomenal nature of stress accents, but on the other hand it speaks to a general expressive feature of Chopin's music, how he creates rhythmic variety in his music from "playing with patterns of articulation on the musical surface to weaken phrase regularity" (Rothstein (1989): 214-248). This is especially significant in this mazurka, since we have had reason to remark on its deceptively simple structure before. In fact, the only regular rhythmic occurrence here is the dotted rhythm on the last beat of each measure. This rhythm gives the music a sense of forward propulsion, again consistent with a fast dance, and which just aids the sense of this piece being unending too. But the unusual stress accents also make this a rather difficult piece to dance to, which supports the competing, internalist, belief that Chopin's mazurkas were not meant as pieces for dancing.

Level B of Example 2.3-12 takes us to the Op. 67 No. 3 mazurka. Again, the analysis here is restricted to the 8-bar compound antecedent phrase of mm. 1-8, since most of the music of this piece, as we have seen, just repeats and/or varies this phrase. The rhythm of this passage is much more flowing, even lilting, which speaks to the piece's *allegretto* tempo, which is in stark contrast with the Op. 7 No. 5's *vivo* tempo. Unlike that mazurka, where irregular stress patterns greatly animate the rhythmic surface, here they are virtually absent – the only apparent rhythmic stresses are the left hand downbeat stresses effected by the pedal signs (not shown in this graph). No surprise then that this piece resembles the slower *kujawiak* dance in its rhythmic character – in fact, its surface rhythm (the one shown in level B) is almost identical to that of the *kujawiak* dance in Example 3 of Anne Swartz's paper (Swartz (1983): 421). So, pragmatically speaking, the music for this mazurka might be understood as music for a *kujawiak*-like dance. But that is probably best left for a dance theorist to decide.

The rhythmic structure of mm. 1-8 creates some issues of formal interest though, which are relevant to our semantic interpretation of the piece. Basically, the rhythmic structure here corresponds to, and therefore emphasizes, the compound antecedent structure of these measures. This compound antecedent is sentential, which means that it can be divided into an initial 4-bar presentation phrase, and a

final 4-bar continuation phrase.⁸ I refer to these in level B of Example 2.3-12 as a “statement” and “response” respectively, for ease of comparison with level A. The presentation phrase or “statement” can be further divided into a 2-bar basic idea and its 2-bar repeat, which are labelled in staves a and b of level B as Part 1 and Part 2 of the statement. Paralleling this, the continuation or “response” is also divided into a Part 1 and Part 2. Part 2 of both statement and response varies from Part 1 a little bit, particularly in its addition of a trill.

Not only does the division of rhythmic patterns in these eight measures correspond to its formal structure, the rhythmic structure of these measures reveals some interesting details about the distribution of stresses within this formal structure. Specifically, the second measure of both Part 1 and Part 2 in the statement, and the second measure of Part 1 in the response, reverses the long-short-short rhythm of the first measure to the generic short-long-long rhythm. As we saw in Charles Rosen’s discussion of mazurka rhythm above, this reversal helps close each part.⁹ But interestingly, where we would most expect this phenomenon to occur, i.e. in the second measure of Part 2 in the response – in other words, in the last measure of the passage – we see a long half note spanning the first two beats followed by a quarter note. Meaning that the closing short-long-long rhythm does *not* occur in the very measure that is supposed to close the phrase. The reason for this seems to lie in the fact that it happens in the last measure of a compound *antecedent* phrase, which has the formal function of a *beginning*. So, the last measure of this phrase, though an ending locally, is not an ending in the larger structure of the compound period that contains this phrase. And this is reflected in its rhythmic structure – whereas the closing pattern of short-long-long is absent from the rhythmic structure of the last measure of the antecedent, it *does* occur in the last measure of the following compound *consequent* phrase that closes the larger compound period (see measure 16 of the piece, in this regard).

⁸ This final 4-bar continuation phrase is what William Caplin would refer to as having a continuation \Rightarrow cadential formal function, given that these measures are harmonized exclusively by what is known as an “Expanded Cadential Progression” or “ECP” (Caplin (1998): 19-20).

⁹ The response part, at least in measure 6, ties the two long notes together in the manner described by Rosen too, as can be seen by comparing this to Example 2.3-10 and the second half of the standard mazurka pattern in Example 2.3-11.

What this reflects is the fact that the way in which rhythm is *articulated* in this passage, has implications for its form-functional *interpretation*. In other words, the structure of this passage reveals a pairing between its ‘prosodic’ (i.e. rhythmic), and more generally ‘phonetic’, aspects and its ‘semantic’ (i.e. form-functional) aspects – of the kind we see in the pairing of PF and LF in language. This just seems to confirm, yet again, how we can think of music as a system of sound-meaning pairings akin to language.

We finally come to the Op. 24 No. 2 mazurka. Level A of Example 2.3-13 depicts the rhythmic structure of the “Lydian” mm. 21-36 of the piece, in terms of the “theme plus three variations” nature of this passage described earlier in level C of Example 2.3-4. The top line in level A depicts the rhythm of the theme, and A(a) displays the theme itself (mm. 21-24). The three subsequent variations display almost the exact same rhythmic structure, which just reaffirms the formal division of this passage into these four parts. But the variations also *vary* the rhythm of the theme little-by-little, keeping with their nature *as* variations of the theme – level A(b) varies the third measure, A(c) the second measure in addition to the third measure, and also the last beat of the first measure, and A(d) varies all of the first three measures. Moreover, the stress patterns are rather unsystematic and appear in rather odd places, but this just reflects the rather odd character of the rhythm here – everything seems the wrong way around. The theme conforms to the generic mazurka rhythm the most, but each subsequent variation seems to deviate from this a little bit more every time, as they vary the rhythm of the theme, until variation 3 in level A(d) is reduced to a string of unstressed, unaccented eighth notes in the first two measures – the least danceable rhythm one could expect to find in a mazurka. Finally, the theme plus its variations *all* end with a final measure that displays an ‘unclosed’ long-short pattern, denying them the *holupiec*-related closing role they could have had.

The significance of all of this becomes clear when we compare this with level B of Example 2.3-13, which illustrates the rhythmic structure of the preceding mm. 5-20 of the mazurka. As discussed in

Example 2.3-13. Chopin, Mazurka in C major, Op. 24 No. 2: Rhythmic structure, mm. 5-36

Chopin Mazurka, Op. 24 #2, mm. 21-36

a. mm. 21-24 b. 29-32 c. 25-28 d. 33-36

A

Rhythm of Theme

Theme

Variation 1

Variation 2

Variation 3

Chopin Mazurka, Op. 24 #2, mm. 5-20

a. mm. 5-8 b. 9-12 c. 13-16 d. 17-20

B

Rhythm of Statement

Rhythm of Response

Statement

Response

holupiec

Example 2.3-6, we can think of this passage as having two parts, viz. an initial 4-bar structure in which two measures in C major are prolonged by two subsequent measures in A minor, all of which is repeated once, which is followed by another 4-bar structure in which a 2-bar prolongation of dominant G-major harmony is followed by a two-bar cadential progression that confirms C major again, all of this being repeated once as well. Level B of Example 2.3-13 refers to the first 4-bar structure as a “statement”, and the second 4-bar structure as a “response”, in keeping with the trend established in the previous examples. The top two, single-line, staves of level B display the basic rhythm of the statement and the response. Underneath this, staves a and b depict the actual music of the statement and its (slightly varied) repetition, and staves c and d depict the music of the response and its repetition.

Now, compare the rhythm of the statement, shown in the top staff of level B, with the rhythm of the theme of mm. 21-36, shown in the top staff of level A above this – in particular, compare the first two measures of these rhythms. You will notice that the rhythm of the first two measures in the statement in level B is switched around in the first two measures of the theme in level A. So, the unclosed nature of the theme and its variations in mm. 21-36 of the mazurka has partly to do with the way they switch around the closed mazurka rhythm of the beginning of the piece, probably to create a sense of forward motion, directed towards the recapitulation of mm. 5-20 of the mazurka *after* mm. 21-36, i.e. in mm. 37-52. This recapitulation then brings the music to a close, which gives mm. 5-52 of the piece a rounded binary formal structure.

The closed nature of mm. 5-20, i.e. the initial A section of the rounded binary form of mm. 5-52, can be seen further from the rhythm of the response to the statement in level B of example 2.3-13 too. This rhythm, shown in the second single-line staff at the top of level B, ends with a clear *holupiec* gesture, shown in the box in the last measure of the response. This *holupiec* gesture often signifies the end of a dance, as we know from our earlier discussion of this term. Moreover, this gesture coincides with the final tonic triad of the authentic cadence that ends this A section too, which confirms the closing nature of the gesture. Finally, the rhythm of the entire 4-bar response structure that ends with the *holupiec* gesture, has accents on the third beat in its first three measures, but an additional accent on the second beat in its

last measure, which contains the *holupiec* gesture, and is harmonized by tonic harmony. Following Charles Rosen's above observations about this, such an accent pattern signifies the end of the dance too – or at least the dance section that this initial A section of the mazurka represents.¹⁰

To summarize, we see from the preceding discussion that a simple generic rhythmic pattern can be realized in a number of ways, in different mazurka pieces. However, these different realizations all seem to connect with the different foot gestures and meter of dances, and also correlate with certain aspects of form in Chopin's compositions, especially in the 'theme and variations' and 'statement and response' formal designs we have seen above. This demonstrates that mazurka rhythm is not an unsystematic, random distribution of stresses and accents, despite the fact that these stresses and accents do not lend them themselves to a conclusive, semantic theory of mazurka.

What the preceding discussion also illustrates, yet again, is how a phenomenon like rhythm in Chopin's mazurkas, with all its externalist associations, still merits an *internalist* exploration, particularly in its connection to the mazurka's grammatical structure. This is evidenced by the fact that it is this internalist approach that has allowed us to describe the 'opening' and 'closing' aspects of this problematic concept, and therefore its connection to the formal, melodic, harmonic, and metrical structure of Chopin's mazurkas. Moreover, it is only because of this focus have we been able to make sense of one of the characteristic features of mazurka rhythm, viz. the *holupiec* gesture, since it is through a formal and internal explanation of mazurka structure that we can understand the closing (and sometimes opening) function of this gesture. Nowhere have we seen Chopin place a *holupiec* where the tonal activity suggests a 'middle' – it is only at cadential endings (and occasionally at structurally-important, tonic-prolonging beginnings) that the gesture appears.

¹⁰ Note that both the second and fourth measures of the statement's rhythm at the top of level B have their second beat accented, and decorated with a grace note too in the actual music in the a and b staves below. This reveals the mini-closing role of these measures. In the case of the second measure, this closes the initial C-major tonic prolongation of the statement, which is also the part that is switched around in mm. 21-36. In the case of the fourth measure it closes the statement itself, which leads to the music of the response.

Finally, we have explored the above aspects of the mazurka, whether it is rhythmic structure, metrical structure, form-functional structure, or harmonic structure, in pieces that range from Chopin's early Op. 7 No. 5 work, stretching all the way to his posthumously-published Op. 67 No. 3 mazurka, meaning that our internalist attempt to describe the grammar of the mazurka, and interpret its semantic and phonetic aspects, has helped us understand a substantial body of Chopin's life's work. In this manner, Chopin's mazurkas can be seen as evidence for a scientific theory of music, especially one that attempts to describe and explain how the workings of the musical mind result in the generation and interpretation of actual musical structures by people in different musical idioms.

At this point it is also worth revisiting the idea, though, that such a theory only describes and explains how music is generated, and how this generation makes music *interpretable*. It does not attempt to describe or explain how music is *interpreted*. This is because the interpretation of music is subject to a variety of pragmatic, performance factors – as a result of which our above analyses of how various Chopin mazurka passages are interpretable (in light of how their grammatical structure maps to their metrical, formal, and other aspects), is not how others *will*, necessarily, analyze these passages. This means that musical surfaces can be interpreted in a variety of ways, and for a variety of reasons – leading many music analysts to understand their analyses of musical pieces not as objective experiments on those pieces (as this chapter does), but rather as something subjective, akin even to how musicians subjectively interpret pieces in a musical performance. (See Whittall (1987) for a discussion of this metaphor.) This is a perfectly valid attitude to have towards analysis, of course, given the pragmatic nature of musical interpretation. But I think it is worth re-emphasizing the distinction between musical interpretation, on the one hand, and how musical grammar makes music interpretable on the other, if for no other reason than to highlight the objective quality analysis can have, when it is situated within a formal, internalist investigation of musical structure.

In the next, and final, section of this chapter, I will explore this distinction between interpretation and interpretability in more detail. In particular, I will focus on a passage from Chopin's F major mazurka

Op. 68 No. 3, to see how its contrapuntal structure makes its interpretation challenging, in spite of the fact that its grammatical structure makes it perfectly interpretable.

2.3.4. Interpretation vs. Interpretability in Tonal Music

So, let us begin by examining the F major mazurka, Op. 68 No. 3. The relevant passage in this mazurka for our discussion on interpretation is mm. 1-8. The first 32 measures of this piece have the form of a small rounded binary structure, so mm. 1-8 of the piece make up the first half of the A section of this rounded binary structure. (The exact form of these first eight measures, which I will not discuss just yet, will be relevant later for how we interpret these measures.)

Example 2.3-14 depicts the underlying grammatical structure of this passage, and therefore the aforementioned contrapuntal complications involved in its generation. The passage is in F major, and ends with a half cadence, so its background harmonic structure is a large-scale I – V motion. This is shown in level A of the example, which also reveals that the pitch A4 of the initial tonic triad is the *Kopfton* of the *Urlinie* of the passage, which ends the passage by descending to the scale degree 2 G4 pitch of the final V chord of the passage.

As in the previous derivations, level B adds an intermediate sonority to the fundamental progression of level A, in this case an applied predominant V/V chord. Level C then prolongs the initial tonic sonority of the fundamental progression through a contrapuntal operation – and here we begin to see the contrapuntal complexity of the passage emerge. This is because the operation that prolongs the initial tonic triad involves taking this triad and prolonging it rightward by means of a contrapuntal (or “voice-leading”) progression known as a “sequence”. A sequence is a voice-leading progression whose constituent sonorities either ascend or descend in a patterned manner. In this case, the constituent sonorities are triads, whose roots descend by third, generating the progression I – vi – IV. The contrapuntal nature of such sequences can be seen in the fact that the constituent chords of the progression do not have any grammatical function – therefore the IV in it does not have subdominant

Example 2.3-14. Chopin, Mazurka in F major, Op. 68 No. 3: Tonal generation, mm. 1-8

The image displays four systems of musical notation (A, B, C, D) for Chopin's Mazurka in F major, Op. 68 No. 3, measures 1-8. Each system shows the tonal generation in both the treble and bass staves.

- System A:** Shows the initial I chord (F major) in the bass staff and a triad with a third (^3) in the treble staff. The final chord is V (C major) in the bass staff and a triad with a second (^2) in the treble staff.
- System B:** Shows the I chord in the bass staff. The treble staff shows a triad with a third (^3) and a V/V chord (D major) in the bass staff. The final chord is V (C major) in the bass staff and a triad with a second (^2) in the treble staff.
- System C:** Shows the I chord in the bass staff. The treble staff shows a triad with a third (^3) and a V/V chord (D major) in the bass staff. The final chord is V (C major) in the bass staff and a triad with a second (^2) in the treble staff. Fingerings are indicated as 10 - 10 - 10 in the treble staff.
- System D:** Shows the I chord in the bass staff. The treble staff shows a triad with a third (^3) and a V/V chord (D major) in the bass staff. The final chord is V (C major) in the bass staff and a triad with a second (^2) in the treble staff. Fingerings are indicated as 10 5 - 10 6 - 10 5 - 10 6 - 10 5 - 10 6 - 10 5 in the treble staff. Chords are labeled as I, vi, IV, V/V, and V in the bass staff.

Example 2.3-14 (contd). Chopin, Mazurka in F major, Op. 68 No. 3: Tonal generation

E

8 5 10 8 10 8 10 8 10 8 5 10

10 5 10 5 10 5 10 5 10

I vi IV V/V V

function. (For this reason, there can be sequences whose constituent chords appear in an order that violate the way in which functional harmonies are supposed to appear in a grammatical progression. An example of this is the ascending 5ths sequence, the roots of whose constituent chords ascend by fifth, leading to progressions like I – V – ii – vi – iii and so on, which violates the standard descending “circle of fifths” pattern in which harmonies appear in grammatical progressions in tonal music.)

The voice leading basis for the descending thirds sequence that prolongs the initial tonic triad in level C of Example 2.3-14, can be seen further in how the bass and alto voices of the chords in this sequence form the interval of a 10th (i.e. a compound 3rd). As we move through the sequence, descending by a third each time (given the descending 3rds nature of the sequence), these 10ths in each chord descend by a 3rd too, leading to the parallel 10ths bass-alto progression F – A, D – F and B-flat – D, shown by the “10 – 10 – 10” label in level C. Such a pattern of parallel intervals is known in Schenkerian theory as a “linear intervallic pattern” (Forte and Gilbert (1982): 83-99). As we continue the derivation of mm. 1-8 of the F-major mazurka in level D, a tenor voice is added to the progression too. But this leads to a bad linear intervallic pattern of parallel 5ths between the I, vi, and IV chords, which can be seen by the bass-tenor pairs F – C in the I chord, which leads to D – A in the vi chord, which then leads to B-flat – F in the IV chord. To fix this, first inversion C-major and A-minor triads are inserted between the I and vi, and vi and IV, chords respectively. The interval between the bass and tenor of these first inversion triads

is that of a 6th, which therefore breaks up the aforementioned parallel 5ths between the surrounding chords. This can be seen in the “5 – 6 – 5 – 6 – 5” label in level D, as a result of which this version of the descending 3rds contrapuntal sequence we have been exploring is also known as the “descending 5-6” sequence.

Notice that the addition of the first inversion triads to the sequence in level D leads to more bass-alto parallel 10ths being added to the parallel 10ths linear intervallic pattern of level C. Moreover, notice how the parallel 10th and parallel 5-6 patterns continue beyond the IV chord, into the following V/V chord that was added to the derivation in level B. This ensures a contrapuntally smooth connection between the first and second halves of the passage. But this is exactly the kind of contrapuntal nuance that will lead to problems of interpreting this passage, as we shall see in a moment.

Finally, we move to level E of Example 2.3-14, where we complete the generation of the surface of mm. 1-8 of the F major mazurka. If you compare this level with the previous level, you will notice that the only difference between them lies in the voicing of the chords in the two levels. As previously discussed, such a difference really lies in the realm of the ‘phonetic’ interpretation of the surface, i.e. in how the harmonic structure of the surface is actually articulated with specific pitches in specific registers, and which have specific durations. But I would like to consider the voicing of the chords in level E of the example within our current discussion of how the surface is generated anyway, because of the way this voicing relates to the background structure, i.e. the *Ursatz*, of the passage. The main feature worth noting here is how the *Urlinie* of the passage, which involves a descent from the A4 *Kopfton* of the passage to the G4 in the final V chord of the passage, does not manifest itself as the actual soprano of the passage, i.e. in its surface manifestation, as it does in the more background levels of structure. Instead, the alto voice of levels C and D, beginning with the F4 pitch in the initial tonic triad of the passage, manifests itself as the actual soprano of the passage in level E, by being transposed up an octave as the upward-pointing shows. Moreover, the tenor voice of levels C and D, beginning with the C4 pitch in the initial tonic triad of the passage, is also transposed up an octave, where it becomes the actual *alto* voice in level E. This means that the *Urlinie* of the passage manifests itself as the *tenor* voice in the passage’s surface.

Finally, the actual bass of the passage is realized in a different way from the abstractions of levels C and D too, because the first inversion C-major and A-minor triads that were introduced in level D, to break up the parallel fifths in that level, are now realized as root-position chords in the surface – which changes the bass line of the passage. (This transforms the descending 5-6 sequence of the passage into its “root-position variant” too (cf. Aldwell and Schachter (2011): 318).) Due to this altered bass line, the intervals between the bass and the upper voices are now altered as well, so that the parallel 10ths linear intervallic pattern now becomes a 10-5 pattern instead, realized now in the tenor voice of the surface due to all the voice transpositions mentioned earlier. Similarly, the 5-6 linear intervallic pattern now becomes a 5-8 pattern, realized in the alto voice of the surface, and the surface’s soprano voice (not to confused with the structural soprano, i.e. the *Urlinie*), realizes an 8-10 linear intervallic pattern.

So, the derivation of this mazurka passage shows us how complicated its surface is, due to all the contrapuntal intricacies that play a role in this derivation. Let us briefly examine this surface itself, to consider how it might be interpretable, before we move on to the aforementioned challenging task of actually interpreting it. Example 2.3-15 illustrates the surface structure of mm. 1-8 of the mazurka passage. Image A in the example shows how the surface realizes the pitch content of level E in the previous example with actual durations, specifically within the triple meter structure characteristic of mazurkas. It also shows how Chopin divides the triple meter into patterns of short and long notes, which have connections to the accents inherent in the mazurka dance, as we saw above. As a result, what image A provides is a durational interpretation of mm. 1-8 of the F major mazurka, in both rhythmic and metrical terms. However, image A also illustrates how this passage is durationally interpretable – but for this to be evident, we need to explore the rhythmic and metrical structure of this passage more closely, to see how the grammatical structure of the passage maps to it. I shall do this in a moment.

If you look at the first measure of image A again, you will notice that the downbeat sonority in the piano’s right hand part is the dyad A4-C5, the C5 being replaced by an F5 on the third beat of the measure. You might remember that both the C5 and F5 are actually registrally-transposed inner voices –

Example 2.3-15. Chopin, Mazurka in F major, Op. 68 No. 3: Surface structure, mm. 1-8

Durational Structure: Chopin Mazurka, Op. 68 #3, Surface

mm. 2 5 8

A

Pitch Structure: Chopin Mazurka, Op. 68 #3, Surface

mm. ³ 2 5 8 ²

B

I ————— vi ————— IV ————— V/V V

we saw in level E of Example 2.3-14 how these pitches, and in fact the complete voices they belong to, are transposed up an octave from their inner voice positions in levels A through D. This means that, on the downbeat of the first measure in image A, the voice that contains the A4 pitch is the structural upper voice, or the *Urlinie*, of the passage, and the A4 is indeed the *Kopfton* within this *Urlinie* – although it appears to be an inner voice in the surface of the piece shown in image A because of the way the actual inner-voice C5 and F5 pitches have been transposed above it. This is made clearer by image B in Example 2.3-15, which illustrates the surface pitch structure of our F major mazurka passage. Here we clearly see how the A4 in the first measure is the scale degree 3 *Kopfton* of the piece, which descends to the G4 of the *Urlinie* over the V harmony at the very end of the passage. Interestingly, this G4 is but implied by the structure of the *Urlinie*, and is not actually there in the surface of the passage, which is

why it is shown in parentheses. The reason why it is not present in the surface is because it has been transposed down an octave, and is therefore present in the surface, but in the piano's left hand part, as the arrow in the last measure of image B shows.

Finally, the actual top two voices of the passage, beginning with the aforementioned C5 and F5 of the first measure – transposed up from their inner voice status in the deeper structure of the passage – are realized in an interesting way in the surface of the passage too. As mentioned above, the C5 occurs on the downbeat of measure 1, whereas the F5 occurs on the third beat of that measure. But in the next measure, we see the E5 that follows F5 in the actual top voice of the surface occurs on the *downbeat* instead, and the C5 that was on the downbeat of the previous measure, is now pushed back to beat 2 of this measure. And this pattern continues into the second last measure of the passage, which means that the actual two top voices of the passage are never sounded simultaneously throughout the passage, but are staggered against each other instead. Another way of saying this is that in each measure, the interval between the two top voices is *unfolded*, which is why they are not sounded simultaneously. We have encountered the phenomenon of unfolding (i.e. Schenker's *Ausfaltung*) previously, in our discussion of the C major Op. 67 No. 3 mazurka, and we learned about the zigzag notation used to represent this phenomenon in a Schenkerian graph – and we see this again in image B in Example 2.3-15, which now explains the staggered nature of the top two voices of the passage as a series of unfoldings between them, which lasts into the second-last measure of the passage. (Notice how the last unfolding, between D4 and G4 in measure 7, involves an implied tone, viz. the D4, because of the way this pitch has been transposed down into the piano's left hand, as the arrow in that measure shows – and just as happened to the G4 in the following measure.)

What example 2.3-15 shows us, yet again, is the complicated structure of this F major mazurka passage, which has much to do with its complicated contrapuntal origins, which, in turn, leads to its nuanced surface structure – something image B in the example illustrates particularly well. Because of this complicated contrapuntal structure, which contains instances of voice-leading phenomena as varied as

sequences, linear intervallic patterns, chord re-voicings, and voice transpositions, we are now faced with a particularly gnarly problem of describing how this passage can be interpreted – which is the very issue that made this passage worth discussing in the first place. For on the one hand, the above contrapuntal structure of this passage might make us see much of the passage as a descending 3rds prolongation of the *initial* tonic triad, because of the sequential nature of this progression – but on the other hand, the very sonorities that make up this descending 3rds progression, especially the vi and IV chords, might be seen as prolonging the *final* V triad, as dominants to this V chord.

To understand this better, consider Example 2.3-16, which displays the two above ways of understanding the prolongational structure of the passage with grammatical tree diagrams. In both of the images in this example, the final V harmony of the passage prolongs the initial tonic harmony of the passage, which can be seen in the way the branch that represents the V chord branches rightwards off of the initial tonic triad's branch.¹¹ Also, the final V harmony is itself prolonged by the V/V harmony introduced in level B of Example 2.3-14, which is therefore shown as branching leftward off of the final V harmony in both images in Example 2.3-16.

The difference between the two images in Example 2.3-16, however, lies in how they represent the sonorities that lie in between the initial tonic harmony of the passage, and the V/V and V harmonies at its end. Image A on top privileges the complicated contrapuntal structure of the passage in this regard, and sees these intermediate sonorities as prolonging the initial tonic harmony, which is why they are shown as *right-branching* off of the initial tonic triad. That is, the descending 5-6 sequence, which is built on the more fundamental descending 3rds progression I – vi – IV, is seen as prolonging the initial tonic triad, which is why the vi and IV chords are represented by branches in the tree diagram that branch

¹¹ In section 1.2.4.ii of chapter 1.2, I had discussed the possibility that *back-relating* dominant chords, like the final V chord in the above example, really arise from a movement operation involving the V chord and the *final* tonic triad of the *Ursatz*, which follows the V chord in the *Ursatz*. This final tonic triad does not appear in the F-major mazurka passage we are currently discussing, and is therefore not shown in the above examples. However, if the above movement hypothesis is correct, the final V chord in the mazurka passage should be branching *leftward* from that final tonic triad, rather than rightward from the initial tonic triad as shown in the above examples.

Example 2.3-16. Chopin, Mazurka in F major, Op. 68 No. 3: Prospective/Retroactive trees

Prospective Prolongation

A

I vi IV V/V V

Retroactive Prolongation

B

I vi IV V/V V

rightwards off of the initial tonic triad's branch. These sonorities are themselves prolonged by the chords that break up parallel fifths between them, which explains the shorter branches of the tree that right-branch off of the initial I, vi, and IV chords as well.

To put all of this in terms of Kofi Agawu's description of right-branching prolongations (discussed in section 1.2.4.vii of the second chapter), the vi and IV chords in image A can be seen as *prospectively* prolonging the initial tonic triad – which explains the title given to image A. In contrast, image B at the bottom of Example 2.3-16 sees the prolongational structure of the passage in *retroactive* terms. That is, this image privileges the harmonic structure of the passage over its contrapuntal structure, and therefore treats the vi and IV harmonies in the descending 3rds progression as predominant harmonies instead – which therefore prolong not the initial tonic triad, but the final V triad of the passage. This is why image B shows them as *left-branching*, retroactively, off of the final V triad, just as the applied predominant V/V harmony does.

So, which tree diagram represents mm. 1-8 of the F major mazurka correctly? In other words, do the intermediate sonorities of the passage, especially the vi and IV chords, prospectively prolong the initial tonic, or retroactively prolong the final dominant? The answer to these questions is where this section's focus on the distinction between "interpretation" and "interpretability" becomes so crucial. This is because, as *interpretations* of the passage, both A and B are legitimate. One could interpret the passage as having a prospective prolongational structure, a retroactive prolongational structure, or indeed some other structure altogether – as long as one can find an appropriate pragmatic justification for interpreting the passage that way. Our very discussion of the two readings of the passage in terms of a "harmony versus counterpoint" dialectic therefore suggests that both readings can be justified, as legitimate interpretations of the passage.

But this is separate from what makes the passage *interpretable*. For the passage to be interpretable, tonal grammar has to generate the surface of the passage in such a way that it can be comprehended, i.e. receive a 'semantic' interpretation – a *form-functional interpretation*, as chapter 2.1 put it, and also articulated or pronounced, i.e. receive a 'phonetic' interpretation – a *durational*

interpretation, as chapter 2.2 put it. So, one might (consciously) choose to interpret mm. 1-8 of the F major mazurka as a prospective or retroactively prolonged entity, but that is not necessarily how tonal grammar makes the passage (unconsciously) interpretable to a person who has native competence in the Western Classical tonal idiom. For *that*, we have to explore how the grammatical structure of the passage maps to its meaning and durational rhythm.

But there are reasons to believe that such an exploration would reveal a breakdown in this mapping between grammar, meaning and rhythm, in *both* the prospective *and* the retroactive interpretations of the passage – meaning that neither reveals how the passage is interpretable, even if they are both legitimate interpretations of the passage. In order to understand why this seems to be the case, consider the fact that both the prospective and retroactive interpretations of the passage divide the passage into two parts, stemming from the two harmonies from which this passage is generated. Therefore both interpretations see the first part of the passage as prolonging the initial tonic harmony of the passage, and the second part as prolonging the final dominant harmony of the passage. And for the surface of the mazurka to be generated successfully, according to either of these interpretations of its structure, these two parts of the passage have to be semantically interpretable too, perhaps as a beginning and an ending.

Now the prospective interpretation sees the first part of the passage as a prospective prolongation of the initial tonic harmony of the passage by means of the vi and IV harmonies, whereas the retroactive interpretation sees the final dominant harmony of the *second* half of the passage as being retroactively prolonged by these two harmonies instead. But can the prospective prolongation of the initial tonic harmony be semantically interpreted as a beginning? Similarly, can the retroactive prolongation of the final dominant harmony be semantically interpreted as an ending? If the answer to one or both questions is “no”, then either or both the prospective and retroactive views of the passage will fail to be accounts of how the passage is interpretable, even though they might both be perfectly legitimate interpretations of the passage.

And the answer to both of the above questions does indeed seem to be “no”. This has to do with the form of this mazurka – and as I mentioned earlier in this section, this is where the specific form of the

mazurka passage we have been considering, i.e. mm. 1-8, becomes relevant. The first 32 measures of the mazurka, as mentioned earlier, constitute a small rounded binary form, and mm. 1-8 of the piece make up the first half of the A section of this form. Rounded binary A sections tend to be made up of conventional 8- or 16-bar themes as well, so mm. 1-8 make up the first half of such a conventional, in this case 16-bar, theme, viz. a compound period. This means that mm. 1-8 form an 8-bar compound antecedent phrase within this larger compound period. According to William Caplin, compound antecedent phrases can be further subdivided into two parts, viz. a first half, equivalent to a beginning – which is usually made up of a simple antecedent phrase, a simple presentation phrase, or a compound basic idea – and a second half, roughly equivalent to an ending, which is usually made up of a continuation phrase (Caplin (1998): 63-70). (The two halves of a compound antecedent are each usually four bars long too. I will ignore this feature to focus on just the beginning + ending aspects of the structure for the time being.) So, for either the prospective or retroactive trees of Example 2.3-16 to be semantically (i.e. formally) interpretable, as a compound antecedent phrase – which mm. 1-8 of the mazurka seem to be – their first halves must map to the aforementioned first halves of a compound antecedent phrase, and their second halves must map to the second half of a compound antecedent phrase as well. In other words, the first half of both the prospective and the retroactive trees must be formally interpretable as either a simple antecedent or presentation phrase, or a compound basic idea, and the second half of both the prospective and the retroactive trees must be formally interpretable as a continuation phrase.

But it is hard to see how this can be the case. The first half of neither the prospective tree nor the retroactive tree ends with a cadence, which is required for them to be interpreted as a simple antecedent phrase. For them to be interpreted as a simple presentation phrase or a compound basic idea instead, they have to be further divisible into an initial tonic-prolonging “basic idea”, and either a subsequent repeat of the basic idea (in the case of a presentation phrase), or a contrasting idea (in the case of a compound basic idea). In the case of the prospective tree, whose first half is founded on the descending 3rds progression I – vi – IV, this means that either the basic idea would be harmonized by the initial tonic harmony of the descending 3rds progression (leaving its repeat or a subsequent contrasting idea to be harmonized by the

remaining vi – IV progression), or it would be harmonized by a I – vi progression (leaving its repeat or a subsequent contrasting idea to be harmonized by just the remaining subdominant harmony). Both possibilities are, however, very strange – I cannot think of any other presentation phrase or compound basic idea that is harmonized in this way. This means that the first half of the prospective tree does not seem to be interpretable as either a presentation phrase or a compound basic idea, i.e. as the first half of a compound antecedent phrase. Which further implies that the prospective tree itself, as a whole, does not seem to be interpretable as a compound antecedent phrase.

On the other hand, the first half of the *retroactive* tree is founded on just the initial tonic harmony of the passage – and this could be the harmonic basis for both the basic idea, and its subsequent repeat or replacement with a contrasting idea too. This means that the first half of the retroactive tree might be interpretable as the first half of a compound antecedent phrase. But if this is the case, then the second half of the retroactive tree must be interpretable as the second, continuation, half of the compound antecedent phrase too. This means that this continuation phrase would be founded on the remaining harmonic progression of the retroactive tree, i.e. vi – IV – V/V – V. Again, this is possible – but there are reasons against it. The main reason has to do with the role of *formal processes* in defining a continuation phrase. A formal process is a process that helps shape the specific form of a formal function such as a continuation phrase. We saw one example of such a process in chapter 2.2, known as “fragmentation”, which involves breaking up a passage’s parts into smaller bits. Such fragmentation gives a sense of forward motion, which accords with the ‘continuing’ nature of a continuation phrase. Other such formal processes are increase in harmonic rhythm, increase in surface rhythmic activity, and motivic liquidation – in which part of a motive (usually from the basic idea) is deleted away from the music of the continuation phrase – all of which also enhances the sense of unrest and forward motion associated with continuations.

Now, such formal processes *can* be seen in the second half of the retroactive tree, or more specifically in the way its underlying vi – IV – V/V – V harmonic progression manifests itself in the surface of the F major mazurka. However, they only seem to involve the IV – V/V – I part of the

progression. For example, in the mazurka's surface a fragment of the IV chord is sounded a measure after the IV chord is itself sounded, and this leads to an increase in the passage's harmonic rhythm as well. (Review image A in Example 2.3-15 to verify this.) So, certainly the IV chord and following parts of the retroactive tree might be interpretable as being part of a continuation phrase. But none of these things happen to the vi chord that precedes the IV chord, which problematizes *its* interpretation as part of a continuation phrase. This means that the second half of the retroactive tree as a *whole* does not seem to be interpretable as a continuation phrase, i.e. as the second half of a compound antecedent phrase. Which further implies that the retroactive tree itself, as a whole, does not seem to be interpretable as a compound antecedent phrase either.

And given that mm. 1-8 of the F major mazurka does seem to have the formal structure of a compound antecedent phrase, the above implies that neither the prospective nor the retroactive description of this passage seems to formally interpretable – although, again, they might both be perfectly legitimate *interpretations* of this passage.

In light of the above, it seems like we need to describe the grammatical structure of mm. 1-8 of the mazurka passage in a different way altogether, if we want to account for how this passage is interpretable. Example 2.3-17 depicts a tree diagram that might provide such a description.¹² Unlike the prospective and retroactive trees of Example 2.3-16, this tree generates the vi and IV harmonies of the mazurka passage in different ways. The vi harmony is derived through prospective prolongation of the initial tonic triad – but the IV harmony is derived through retroactive prolongation of the final V triad, as a predominant to that triad. This means that the first half of the tree is made up of a I – vi progression (which also happens

¹² Even though Example 2.3-17 attempts to provide an interpretable description of the passage, remember that the retroactive tree in Example 2.3-16 did have aspects to it that justified its interpretability. This goes to show that certain musical structures can be interpretable in different ways – meaning that such passages are genuinely ambiguous. Of course this does not mean that the description of how musical structures are interpretable is just a subjective activity in the way actually *interpreting* them is. But it does show that there is a place for ambiguity and conflict even within an internalist theory of grammatical generation and semantic/phonetic interpretability.

Example 2.3-17. Chopin, Mazurka in F major, Op. 68 No. 3: Alternative derivation of mm. 1-8

to support the A4 *Kopfton* of the passage, as can be seen in the soprano voice of the image). This initial tonic-prolongational branch¹³ is then followed by the dominant-prolongational second half of the tree, which is harmonized by the remaining IV – V/V – V progression of this passage. Not only is such a grammatical structure for mm. 1-8 consistent with our discussion in Part I of the dissertation about how tonal phrases begin with a tonic-prolongational branch that supports the *Kopfton* (which is followed by another branch that initiates the *Urlinie*'s descent back to the tonic), this structure also shows us how the mazurka passage is formally interpretable. This is because the first half of the structure can now be clearly interpreted as a beginning, and the second half as an ending.

¹³ The name I have given this initial branch is not arbitrary – “tonic prolongational” is meant to specifically invoke Allan Keiler’s label for the first half of the Schenkerian *Ursatz*, which also supports the *Kopfton*, and which can also be described as a binary-branching structure, when understood in tree-diagrammatic terms as we did in chapter 1.2. That is, our current discussion about Chopin’s F major mazurka, and its wider connection to the notions of “interpretation” versus “interpretability” is meant to be fully consistent with a Schenkerian, and Minimalist, description of musical grammar of the kind discussed in the first half of this dissertation – much of which built on Keiler’s ideas too.

Specifically, the first half can be interpreted as a presentation phrase, in which the basic idea is supported by the initial tonic harmony of the I – vi progression that harmonizes this half of the tree, and the repeat of the basic idea therefore being supported by the remaining vi harmony of this progression. The repeat of a basic idea that is supported by vi is referred to as a *sequential* repeat of the basic idea (Caplin (1998): 39), and presentation phrases with this constitution are common in the tonal literature (e.g. see the main theme of the second movement of Beethoven’s Op. 30 No. 1 violin sonata). This suggests that, unlike the prospective and retroactive trees of Example 2.3-16, the tree shown in Example 2.3-17 is not only formally interpretable, but has a form, and supporting harmony, that is quite conventional too.

Following this first half, the second half of the tree can be interpreted as a continuation phrase, given that it is harmonized by the aforementioned IV – V/V – V progression, in which the initiating IV harmony, as just discussed, clearly shows continuation features in the way it is realized in the mazurka’s surface. Given how the V/V – V part of this progression supports the half cadence that ends this passage, it can be clearly interpreted as a cadential idea too, as shown in Example 2.3-17, whose formal function as an ending is obvious.

In this manner, we see how a grammatical description of a musical passage, as exemplified by Example 2.3-17 can be interpretable as well. But we have only looked at one aspect of the interpretability of grammatical structure, viz. the formal (‘semantic’) side. What about the durational (‘phonetic’) side of things? That is, is the structure shown in Example 2.3-17 durationally interpretable too, or at least more interpretable in this regard than the prospective and retroactive trees of Example 2.3-16 were?

To examine this matter, we turn to Example 2.3-18, which gives a durational, and specifically metrical, analysis of mm. 1-8 of the F major mazurka. Level A of the example just depicts the I – V background progression upon which this passage is founded, which we have already examined in previous examples. But level A in this example also shows us the tonal *rhythmic* characteristics of this level, as per our discussion of this concept in chapter 2.2. That is, it shows us how there is a structural

Example 2.3-18. Chopin, Mazurka in F major, Op. 68 No. 3: Metrical analysis, mm. 1-8

Tonal Rhythm

A

I V

Durational Interpretation: Background (♩ = 1 measure in any time signature)

B

Tonic prolongation Dominant prolongation

I V V I

Prolonged Tonal Rhythm

C

I V V I V I I I

Durational Interpretation: Middleground (♩ = 1 measure in any time signature)

D

Background: I V I V

Middleground: I V I V

Example 2.3-18 (contd). Chopin, Mazurka, Op. 68 No. 3: Metrical analysis

Durational Interpretation: Foreground (♩ = 1 measure in any time signature)

Background: ■ ————— ∇ —————

Middleground: ■ ————— ∇ ————— ■ ————— ∇ —————

Foreground: ■ — ∇ — ■ — ∇ — ■ — ∇ — ■ — ∇ —

accent associated with the initial I chord, which would be interpreted in metrical terms as a downbeat (hence the two dots underneath it, as opposed to the single dot under the final V chord, which shows their relative positions in the metrical hierarchy of the passage). If we were to interpret this tonal rhythm durationally, it would lead to the pattern of down and upbeats depicted in level B of the example. That is, the initial tonic triad would be interpreted as a downbeat, and its following prolongation (by vi harmony, as shown in Example 2.3-17) would be the following upbeat. In contrast, the final V harmony of the passage would be interpreted as a downbeat, and its preceding prolongation (by IV and V/V) would be the *preceding* upbeat.

This leads to an ill-formed meter of course, given the uneven distribution of down and upbeats to which this leads. But then again tonal structure does not determine rhythmic structure – *tonal* rhythm is not *durational* rhythm, and *structural* accents are not *metrical* accents. All that tonal structure does is map to rhythm, according to certain well formedness conditions imposed on grammar by musical ‘phonology’ (review the discussion of these terms in chapter 2.2, in this regard). This is why all that needs to happen for our mazurka passage to be durationally interpretable, and specifically metrically interpretable, is that the structural beginning of the passage needs to *align* with a metrical accent – after which the requirement that a well formed meter have evenly spaced beats, and the “2 x n” rule, jointly take care of everything.

We see this happening in level C of the example, which shows the prolonged tonal structure of the passage that elaborates the I – V background of level A. Here, the first structural accent, i.e. the initial tonic triad of level A, is aligned with a metrical accent, i.e. a downbeat, as the notation of the level shows. As a result of this alignment, the tonal structure of the rest of the passage gets mapped to meter accordingly. Following the “2 x n” rule, the prolongation of the second half of level A, i.e. the prolongation of the final V triad in level A, gets accorded the same duration in level C as the prolongation of the initial tonic triad. As a result, the upbeat following the initial downbeat is placed right in the middle of the passage, i.e. at the beginning of the prolongation of the final V, at measure 5. Since this distribution of beats accords with the tonal structure of the passage, we see how this structure is durationally, i.e. metrically, interpretable as well – in addition to being formally interpretable, as Example 2.3-17 illustrated.

There is one problem with the above picture though – it maps the grammatical structure of the passage’s harmonic *background* to the passage’s meter, meaning that what level C really shows is how the passage is *hypermetrically* interpretable. This problem arises because the distribution of downbeats and upbeats in the passage, even though well formed, is only at the 4-bar hypermetrical level (which is why the 8-bar long mazurka passage we have been considering is shown to have only one downbeat and one upbeat in level C. Obviously, downbeats and upbeats occur at levels smaller than that of 4-bar hypermeter too, so this has to be accounted for in our metrical analysis of the mazurka passage. Level D of the example attempts to do this for the 2-bar ‘middleground’ hypermetrical level. This means that the two formal parts of the passage, i.e. the 4-bar presentation phrase and the 4-bar continuation phrase, now get a downbeat and upbeat each. What this shows us is that the four harmonies shown in the level – i.e. I, vi, IV, and V – all align with the pattern of evenly spaced beats that metrical well formedness requires. In turn, this allows the pattern of beats in the passage to align with its 2-bar formal structure – i.e. 2-bar basic idea (harmonized by I), 2-bar sequential repeat of the basic idea (harmonized by vi), 2-bar fragmented continuation (harmonized by IV), and 2-bar cadential idea (harmonized by V). What this also shows us is that the IV harmony, which begins the dominant-prolongational continuation phrase of the

passage – and therefore receives a structural upbeat label in the background of the example – is now metrically *re*-interpreted as a downbeat in the middleground, mainly to ensure that down and upbeats are spaced evenly in the middleground, to conform with metrical well formedness. But this also allows the structural accent represented by the IV, as the beginning of a dominant-prolongational, continuation phrase to map to a metrical accent. All of which just goes to show how (durationally-articulated) sound is paired with (form-functional) meaning in tonal music.

This finally takes us to level E of Example 2.3-18, where we see the metrical structure of the bar-level foreground depicted in addition to the 2- and 4-bar hypermetrical levels discussed above. There is nothing new to be discussed here, apart from the fact that now the vi harmony (and the repeat of the basic idea it initiates), and the V/V harmony (and the cadential idea it initiates) are metrically reinterpreted from being upbeats in the middleground, to downbeats in the foreground. Once again, this shows the pairing of formal function to metrical accent, and therefore the pairing of meaning and sound in music.

While we are on the subject of the metrical structure of the F major mazurka, it might be worth examining the *intra*-measure durational structure of the mazurka passage we have been studying too, just to see how this relates to the previous section's discussion about mazurka rhythm, dancing, and the generic mazurka pattern. Example 2.3-19 illustrates this structure. Level A of the example is just a copy of the mazurka passage's foreground metrical analysis in level E of the previous example, included here so that we may use that to discuss the intra-measure rhythmic structure of the passage. Each of the eight harmonies in this level span a measure's worth of duration, as the "half note = 1 measure in any time signature" key indicates. Level B realizes this harmonic-metrical structure in terms of the specific triple meter of the mazurka, and also divides the harmony in each measure into three quarter notes. This last action allows us to picture the intra-measure durational structure of the mazurka passage in terms of the short-long-long "generic mazurka" pattern discussed in the last section. As was mentioned then, this generic mazurka rhythm does not generate the surface rhythm of the mazurka passage, nor is it connected

Example 2.3-19. Chopin, Mazurka in F major, Op. 68 No. 3: Rhythmic structure, mm. 1-8

Metrical Analysis: Foreground ($\text{♩} = 1$ measure in any time signature)

A

Background: □ _____ V _____
 Middleground: □ _____ V _____ □ _____ V _____
 Foreground: □ — V — □ — V — □ — V — □ — V —

Rhythmic Structure: Generic Mazurka Level

B

Background: □ _____ V _____
 Middleground: □ _____ V _____ □ _____ V _____
 Foreground: □ — V — □ — V — □ — V — □ — V —

Rhythmic Structure: Chopin Mazurka, Op. 68 #3, Rhythmic Surface

C

Background: □ _____ V _____
 Middleground: □ _____ V _____ □ _____ V _____
 Foreground: □ — V — □ — V — □ — V — □ — V —

to the tonal structure of the passage. All it does is help us relate the surface of the passage, and specifically the rhythmic (or phenomenal) accents in it, to issues of dancing.

Finally, level C in the example depicts the harmonies of the passage shown in the previous levels, but now realized in terms of the actual rhythmic structure of the passage's surface. If you compare the

foreground distribution of down and upbeats in level B with this level, you will notice that the surface rhythm of the mazurka is synchronized with this pattern of beats. That is, the second beat of every second measure of the passage, the latter corresponding to every foreground upbeat in level B, is emphasized by means of a half note stress, and this serves to close the larger, 2-bar group of which every second measure forms the second half. (In other words, the 2-bar groups in mm. 1-2, 3-4, 5-6, and 7-8, are all closed by means of a second beat stress in mm. 2, 4, 6, and 8 respectively. Also, the second beat stress closes these groups because of the role these stresses have in signaling the end of a dance, as we observed in connection with Charles Rosen's thoughts on the matter in the last section.) The closing role played by every second measure in the surface therefore helps prepare the following downbeat, and also helps demarcate the 2-bar grouping of the passage. Moreover, the dotted rhythm on the first beat of each measure also helps emphasize the 'short' nature of this first beat within the generic mazurka pattern. In this manner, we see how the intra-measure surface rhythm in this mazurka passage articulates the generic mazurka pattern well, while also synchronizing with the metrical structure of the passage. And since the generic pattern just reflects how dancers dance, we see how the surface rhythm of the passage – especially its first beat lack of stress, second beat stresses, and 2-bar hypermetrical grouping structure – therefore reflects the characteristic foot gestures of dancing as well.

After having discussed the “interpretation versus interpretability” issue quite thoroughly in the context of the first eight measures of the F major mazurka, Op. 68 No. 3, let us now turn to another mazurka passage, to see what *its* structure adds to our understanding of this issue. The relevant passage is the main theme of the G major mazurka, Op. 50 No. 1. This mazurka passage has, in particular, a rather unconventional structure that complicates its interpretability.

The G major mazurka as a whole is in seven-part rondo, i.e. ABACADA form. However, there are elements in it that point to sonata-rondo form (e.g. its coda-like final refrain) – and this very ambiguity hints at the problems to come when interpreting this piece. The main theme of the mazurka, in other words the first A section, which will be our passage of interest in this piece, seems to be structured

as a conventional 8- or 16-bar theme, like most rondo main themes – in this case what appears to be a 16-bar compound sentence. A compound sentence is normally comprised of an initial 8-bar “compound presentation” phrase, and a subsequent 8-bar continuation phrase. The 8-bar compound presentation phrase is itself normally comprised of a 4-bar “compound basic idea”, which is then repeated. Finally, the 4-bar compound basic idea, and by extension its 4-bar repeat, are themselves made up of an initial, tonic-prolonging 2-bar basic idea, and a subsequent 2-bar contrasting idea that could prolong either tonic or non-tonic harmony. Importantly, this contrasting idea, and therefore the compound basic idea as a whole, does not end with a cadence, or else it would be classified as either an antecedent or a consequent phrase.

The main theme of the G major mazurka roughly has the above formal and metrical structure, which gives it the aforementioned appearance of a compound presentation phrase. In discussing this theme, though, I will only focus on its formal, and pitch, structure, and not its metrical structure, because that does not seem to be where the challenge in interpreting this theme seems to lie. So, concerning its formal structure, the G major mazurka theme seems to begin with a 4-bar compound basic idea, just as a 16-bar compound sentence should. However, even at this early stage in our formal interpretation of the theme, we run into problems, for the 2-bar basic idea that is supposed to comprise the first half of the compound basic idea prolongs *dominant* harmony here, instead of tonic harmony. (In other words, the theme, and therefore the piece, has an off-tonic opening, just like the Op. 7 No. 5 mazurka discussed earlier.) As a result, the first four measures of the G major mazurka theme are founded on a V – I harmonic progression, which is a very unusual harmonization for a compound basic idea.

To explore the unusual nature of this compound basic idea further, let us look at Example 2.3-20, which describes how this structure is generated. Level A here shows us the V – I fundamental harmonic structure of the passage. Importantly, the D-major V harmony that begins this structure might also be interpreted as a I, if the passage is viewed top down – meaning that the V – I progression might be interpreted as a I – IV progression in D major. Now such an interpretation might not accord with how the grammar of the piece makes the passage interpretable, i.e. such an interpretation might not reveal how the

Example 2.3-20. Chopin, Mazurka in G major, Op. 50 No. 1: Tonal generation, mm. 1-4

A

B

mm. ① ② ③ ④

C

D

Vivace

Detailed description of the musical example: The example illustrates the tonal generation in the first four measures of Chopin's Mazurka in G major, Op. 50 No. 1. It is divided into four systems (A, B, C, D).
 System A shows the initial chords: a dominant chord (V) in the 8th register and a subdominant chord (7) in the 7th register, resolving to the tonic (I).
 System B shows the continuation of the V8 - 7 progression, with the tonic (I) appearing in the 8th register.
 System C shows the first measure (mm. 1-4) with a 'register transfer' indicated by an arrow, moving the right-hand part from the 7th to the 8th register. The chords V8 - 7 and I are shown below the staff.
 System D shows the full score for measures 1-4, marked 'Vivace'. It includes dynamics like *f* (forte), *Ped.* (pedal), and *** (accents), as well as articulation like slurs and accents. The tempo and dynamics markings are placed above the treble clef staff.

pitch structure of the piece maps to its rhythmic and form-functional aspects. But it could be a legitimate interpretation nonetheless if some justification can be found for it – especially if it is a more convincing harmonization for the initial compound basic idea.

But when the D-major harmony that begins the passage is supplemented with a C pitch (as we see in the alto in level A) – a pitch that forms the dissonant interval of a seventh with the root of the harmony – then the harmony is unequivocally interpreted as a *dominant* seventh harmony in G major. This is because the dissonant C must resolve down by step to B according to the rules of counterpoint, the B itself being part of the G major tonic harmony that follows the dominant seventh in the background harmonic structure of the passage. But level B in the example shows us that the reality of this mazurka passage is more complicated. Rather than resolving immediately, the dissonant C pitch is first transposed up an octave, and only then resolves down by step to a B in the following G triad – this B therefore being an octave higher now too. The result of this octave transposition (which is an instance of the phenomenon known in Schenkerian theory as a *register transfer*) is that the initial V^7 harmony that contains this C pitch is extended for an extra measure in the musical surface. This means that the first two measures of the mazurka passage are exclusively harmonized by a prolonged V^7 harmony, which leads to the situation mentioned earlier in which the supposed “basic idea” of the passage, which is contained within the first two measures of the passage, is exclusively harmonized by a V^7 prolongation rather than a tonic prolongation – a very unusual situation indeed. Moreover, as level B of the example also shows, the following G major triad is revoiced over the span of the next two measures of the passage, by means of a voice exchange that replaces the soprano B4 (i.e. the pitch that resolved the preceding dissonant C pitch) with a soprano D5. This means that the “contrasting idea” of the compound basic idea of mm. 1-4, i.e. mm. 3-4, is now exclusively harmonized by a tonic prolongation.

This unusual structure for the first four measures of the G major mazurka theme, i.e. the theme’s supposed compound basic idea, makes more sense when we see a more enriched description of it in level C of the example. Here we see again the register transfer of the C and A pitches of the initial V^7 harmony, which expands this harmony into measure 2 of the piece. We also see that when these registrally-

transferred pitches occur in measure 2, an octave higher, they do not occur simultaneously – which suggests that the interval between them has been unfolded, just as we saw to be the case for the upper voices in much of the F major Op. 68 No. 3 mazurka passage we studied earlier. This unfolding is represented, as before, by a zigzag line, in measure 2 of level C. And the veracity of this description can be confirmed by comparing it with the actual music of these measures shown in level D.

What level C also shows us is that the B4 in measure 3, to which the dissonant, registrally-transferred C5 of the previous measure resolves, is now established as the scale degree 3 *Kopfton* of the passage. (Hence the upward stem with the caret symbol that is attached to it.) The G major tonic triad in which this *Kopfton* occurs is expanded in measure 3 itself, by means of a voice exchange between its inner G and D voices, and is then expanded further into measure 4, by means of the larger voice exchange in this passage mentioned earlier, in which the *Kopfton* B4 is itself exchanged with the D5 soprano pitch that follows it in measure 4, and which is shown in level B of the example. These voice exchanges are described in more detail in Example 2.3-21, where we see them as occurring between sonorities 1, 2, and 3 of image A in the example – all of these sonorities being different voicings of the G-major harmony that is prolonged in these measures. We have already discussed the first of these voice exchanges, between sonorities 1 and 2 in image A, which occurs between the inner-voice G and D pitches in measure 3 itself. This is shown in the first example of image B in Example 2.3-21. A second voice exchange, between sonorities 2 and 3, and shown in the second example in image B, transfers the *Kopfton* B4 to the tenor position, after being exchanged with the G3 pitch that was involved in the first voice exchange. And finally, all of this is subsumed under the third, larger voice exchange of the passage, which transfers the inner-voice D involved in the first exchange up to the soprano position at the end of measure 4. This is shown as the third example in image B, and was already illustrated in the level B stave of Example 2.3-20.

What all of the above suggests is that the first four measures of the G major mazurka theme seem to make sense, i.e. in the way they work towards establishing a *Kopfton* supported by tonic harmony, in a

Example 2.3-21. Chopin, Mazurka in G major, Op. 50 No. 1: Voice exchanges, mm. 3-4

The image displays two musical staves, A and B, for measures 3 and 4 of Chopin's Mazurka in G major, Op. 50 No. 1.
Staff A: Labeled 'Vivace' and 'f'. It shows a complex rhythmic texture with triplets and voice exchanges between the hands. The bass line includes markings like 'Lea.' and asterisks. Measure numbers 1, 2, and 3 are indicated below the staff.
Staff B: Shows a simplified harmonic structure with chords and voice exchange lines. Measure numbers 1, 2, 2, 3, 1, and 3 are indicated below the staff.

manner that clearly supports the interpretation of this passage as a beginning. In fact, this is hardly different from the much more conventional structures in the tonal literature that establish a *Kopfton*, supported by tonic harmony, by means of an initial ascent or *Anstieg* (as depicted in Examples 2.1-6 and 2.2-7 in this dissertation), and which are also consequently interpreted as beginnings. However, the unusual harmonic structure of our mazurka passage certainly complicates its formal interpretation as a compound basic idea. Which just goes to show the kinds of challenges for interpretation this mazurka raises.

But we have only considered the first four measures – the situation gets much more difficult with the subsequent measures. This is because for us to maintain our formal interpretation of the mazurka theme as a compound sentence, not only do the first four measures have to be interpretable as a

compound basic idea, the remaining measures of the theme have to be interpretable as the remaining constituents of a compound sentence as well. And it is not at all clear that this is possible.

To understand this, consider Example 2.3-22, which describes the grammatical structure of the entire G major mazurka theme, and not just its first four measures. Image A at the top of the example discusses mm. 1-4 of the mazurka theme (and also mm. 9-12, which differs from the structure of the first four measures in just the final chord in measure 12, shown in parentheses – which I will discuss in a bit). Its two stave systems, numbered 1 and 2, depict, respectively, the harmonic background of the four measures under consideration, and the actual score for these measures for comparison with this harmonic structure. We have already discussed both of these in levels C and D of Example 2.3-20, and therefore do not need to do so here again. The two stave systems of image B in Example 2.3-22 depict the harmonic background and score for mm. 5-8 of the mazurka theme – and these are worth discussing, for how they relate to the challenging task of interpreting this mazurka

Notice, for instance, how the D4 alto pitch in measure 5, at the beginning of the harmonic background in image B, is the same pitch that appears in the alto at the beginning of the harmonic background in image A. In image A, this D4 led to the dissonant C4 pitch that turns the initial V harmony of the passage into a V⁷, and which was then registrally-transferred up an octave, where it resolved to the following *Kopfton* B4. This *Kopfton* then participated in a voice exchange that put a D5 in the soprano at the end of measure 4. And this D5 was then registrally-transferred *down* an octave subsequently – which is how we get the alto D4 pitch at the beginning of measure 5 in image B. This means that the original, alto register of the D4 pitch is *coupled* with the soprano register of the D5 pitch (and the B4 *Kopfton* and C5 that precede it), by means of the register transfers that moved the C5 up an octave, and then D5 down an octave. Such a coupling of registers is a special instance of the register transfer phenomenon we have been exploring, known in Schenkerian theory, unsurprisingly, as “coupling” (Schenker’s *Koppelung*, see Schenker (1979): 52).

Example 2.3-22. Chopin, Mazurka in G major, Op. 50 No. 1: Tonal generation, mm. 1-16

Chopin Mazurka, Op. 50 #1: mm. 1-4 & 9-12
 1. Harmonic Background 2. Score

A

Vivace

1

2

8 - 7

V I

^ 3

m. 12

B

Chopin Mazurka, Op. 50 #1: mm. 5-8
 1. Harmonic Background 2. Score

1

2

8 - 7

V I IV

^ 3

Example 2.3-22 (contd). Chopin, Mazurka, Op. 50 No. 1: Tonal generation

Chopin Mazurka, Op. 50 #1: mm. 13-16
 1. Harmonic Background 2. Score

1

2

C

Harmonic background (Part 1) Roman numerals: ii^6 - $\text{V}^4 - 3$ | i | I | ii^6 - $\text{V}^4 - 3$ | I

Figured bass notation: $\text{i} - 6$ | ii^6 | $\text{V}^4 - 3$ | i | I | ii^6 | $\text{V}^4 - 3$ | I

Part 2 includes dynamics like *p* and *mf*, and articulation like accents (^) and slurs.

Chopin Mazurka, Op. 50 #1: mm. 1-16 - Fundamental Structure

mm. 3 7 11 16

Harmonic background Roman numerals: I | I | I | V | I

Figured bass notation: I | I | I | V | I

Accents (^) are placed above the notes in measures 3, 7, 11, 16, and 17.

What the coupling really enables in the mazurka passage though, is for mm. 5-8 to begin in the *same* way, i.e. in the same register, that mm. 1-4 did, despite the register transfer that moved the structural melodic line up an octave in the first four measures. This is significant, because for us to interpret the G major mazurka theme of mm. 1-16 as a compound sentence, mm. 5-8 in it have to be interpretable as a *repeat* of

the compound basic idea of the first four measures. Which means that the first two measures of mm. 5-8 have to repeat the basic idea of mm. 1-2, and the last two measures of mm. 5-8 have to repeat the contrasting idea of mm. 3-4. And the coupling of registers between mm. 1-4 and 5-8 ensures that mm. 5-6 *can* repeat the basic idea of mm. 1-2 – the unusual basic idea that is, harmonized by V⁷ rather than tonic harmony. If you compare the first two measures of image A with the first two measures of image B you will find this exactly to be the case. So, even if the interpretation of mm. 1-4 as a compound basic idea is problematic, at least the larger interpretation of the mazurka theme as a compound sentence seems to be working, insofar as mm. 5-8 seem to be repeating the material of mm. 1-4, as would happen in a compound sentence.

But we speak too soon, because the last two measures of mm. 5-8 do *not* seem to repeat the contrasting idea of mm. 3-4, as you can clearly see by comparing the last two measures of image A with the last two measures of image B. In fact, the structure of these last two measures in mm. 5-8, i.e. mm. 7-8, is quite unusual in and of itself. These measures begin by establishing the *Kopfton*, just as the contrasting idea in mm. 3-4 does, but rather than prolong the tonic that supports the *Kopfton* by means of the voice exchanges described in Example 2.3-21, and as happens in mm. 3-4, mm. 7-8 prolong this tonic with a neighboring *subdominant* harmony instead. The neighboring nature of this IV harmony is revealed in the way its soprano pitch C5 forms an upper neighbor to the B4 *Kopfton* that precedes it – hence its “half note with flag” notation, which is the Schenkerian convention for representing pitches that neighbor other, structural pitches in the *Urlinie*, like the *Kopfton*. And not only does this neighboring harmony confirm the fact that mm. 7-8 do not repeat mm. 3-4, it provides a rather unusual ‘plagal’ ending for such a 2-bar idea too.

So, it is quite clear that mm. 5-8 as a whole do not repeat the compound basic idea of mm. 1-4, particularly because of what happens in mm. 7-8. And this on top of the already problematic interpretation of mm. 1-4 as a compound basic idea to begin with. In this light, one might want to abandon the formal interpretation of the G major mazurka theme as a compound sentence altogether. But the problem with this is that the only viable alternative to this interpretation, which is to interpret the

theme as a 16-bar compound period, does not seem to work either. If we were to interpret the theme as a compound period, then mm. 5-8 of this structure should not contain the same harmonic-melodic material as the first four measures. In other words, it should be the ‘contrasting idea’ to the ‘basic idea’ of the first four measures – this opposition being a defining characteristic of the first, compound antecedent half of a compound period. Now this does seem to be the case for the G major mazurka theme, because of the significant difference between the content of its mm. 3-4 versus its mm. 7-8. But in a compound period, the next four measures, mm. 9-12 *should* contain the same harmonic-melodic material as the first four measures, since these measures are supposed to be, effectively, the repeat of the ‘basic idea’ of the first four measures – which is a defining characteristic of the second, compound *consequent* half of a compound period. In the mazurka theme, mm. 9-12 *do* seem to be quite similar to mm. 1-4, which is why Example 2.3-22 describes their harmonic structure by means of the same image, i.e. image A.

But as I mentioned earlier, there is a crucial difference between these measures too, which has to do with the final chord in measure 12, shown in parentheses at the end of the score in image A, which is absent from the music of mm. 1-4. This difference is crucial because that final chord in measure 12 is an applied vii^{o7}/ii chord, which resolves to the prolonged supertonic harmony in measure 13 – which implies that measure 12, and by extension mm. 9-12 as a whole, do not function as a self-contained unit, but have the function of a *continuation*, which pushes the music forward to the next measure. This makes mm. 9-12 very different from mm. 1-4, and therefore not a repeat of those measures, especially since the formal meaning of mm. 1-4 is clearly that of a ‘beginning’, whereas mm. 9-12, as a continuation, have the formal meaning of a ‘middle’. Moreover, a continuation function is exactly what mm. 9-12 are supposed to have if the larger 16-bar theme is a compound sentence – which is why the initial interpretation of the mazurka theme as a compound sentence is still the most viable form-functional interpretation of the theme.¹⁴ But as we saw above, this does not seem to work too well either.

¹⁴ If we were to ignore the applied chord at the end of measure 12, and therefore treat mm. 9-12 in the mazurka theme as being a genuine repeat of mm. 1-4, even then a compound periodic interpretation of the theme would not work. This is because mm. 5-8 in a compound period normally have a continuation function (Caplin (1998): 65), but mm. 5-8 in the mazurka theme are too similar to mm. 1-4 for us to interpret them as anything but a repeat of mm. 1-

In light of the above, one might want to abandon any conventional formal interpretation of the mazurka theme altogether, and just read it as made up of four, independent 4-bar phrases. But this is an unsatisfying, ‘escapist’ solution, for the theme does have a well-defined structure, easily interpretable as having at least a formal beginning and ending. We already saw how mm. 1-4 establish a tonic-supported *Kopfton*, in a manner that has conventional associations as a beginning. Because of the coupling of registers between mm. 1-4 and mm. 5-8, and the structural similarity of mm. 1-4 and 9-12, the *Kopfton* is therefore re-affirmed all the way to measure 12 in the mazurka too, elaborated momentarily by the subdominant-supported upper neighbor C5 in measure 8. And then the *Kopfton* clearly descends to the tonic pitch in the *Urlinie* in mm. 13-16 to formally close the melody – this being supported by a perfect authentic cadential chord progression that formally closes the rest of the passage as well.

To understand this last point, let us examine image C in Example 2.3-22, which depicts the harmonic background and score for the last four measures, i.e. mm. 13-16, of the mazurka theme. As the harmonic background illustrates, mm. 13-14 here prolong supertonic harmony, which was introduced by the aforementioned applied $\text{vii}^{\text{o7}}/\text{ii}$ chord in the previous measure. This supertonic harmony supports a soprano C5 on the downbeat of measure 13 that has been previously described as a neighbor to the B4 *Kopfton* of the passage, and hence its “half note with flag” notation. The C5 initiates a descending line in the soprano of mm. 13-14, viz. C5 – B4 – B4 – A4, marked as quarter notes with upward stems. Since it is supported by supertonic harmony, this descending line cannot be part of the *Urlinie*, and is really a motion into an inner voice. Only on the downbeat of the following measure, i.e. measure 15, do we see the C5 descend to the B4 *Kopfton* in the *Urlinie* itself, supported by root position tonic harmony, after which the *Kopfton*’s descent to the tonic G4 at the end of measure 16, supported by the tonic-confirming I – $\text{ii}^{\text{6}} – \text{V}^{\text{6/4} – 7/3} – \text{I}$ cadential progression, clearly signals the completion of the *Urlinie* and the close of the passage.

4. Unless we take the unusual subdominant ending of mm. 5-8 as implying its difference from mm. 1-4, and therefore its continuation function – in which case we would have to take the applied chord at the end of measure 12 as implying a continuation function for mm. 9-12 as well, which militates against our very first premise above about ignoring this applied chord, and which therefore prevents a compound periodic interpretation of the mazurka theme.

The argument of the preceding paragraphs can be summarized by the image at the bottom of Example 2.3-22, which describes an *Ursatz* for the entire G major mazurka theme. Here we see how the *Kopftón* is established and prolonged, as detailed in images A and B of the example, and how it finally descends to tonic over V – I harmony to close the piece, as detailed in image C of the example. From this it seems pretty clear that the grammatical structure of this passage yields itself to being formally (i.e. semantically) interpretable as having a beginning and an ending, and possibly a middle, where the *Kopftón* is prolonged by means of the neighboring C5 pitch. The only problem is that this does not seem to be interpretable in form-functional terms, i.e. as a compound sentence or period – at least not definitively so. So, one could *interpret* the mazurka passage as a compound sentence or period, but this is not necessarily how the passage's grammatical structure makes it *interpretable*.

One might object that the problem here is not the un- or under-interpretability of the mazurka theme, but the requirement that it be interpretable in form-functional terms. This is because the theory of formal functions, which was developed by William Caplin, is meant to be applied only to the music of Haydn, Mozart, and Beethoven. But this objection seems neither here nor there. We have been able to understand several mazurka passages in this chapter using form-functional notions, so there is no reason why we should not be able to understand this mazurka using such notions too. Moreover, until an alternative theory of formal functions is developed specifically for the music of Chopin, to say that current form-functional ideas cannot be used to interpret his mazurkas just amounts to begging the question of how Chopin's music is interpretable. And all of this just goes to show the complicated relationship between musical grammar, interpretation, and interpretability in Chopin's music, as illustrated by this G major mazurka theme, and by music in general.

It might be fitting to end this chapter by revisiting that very piece with which we started this chapter's discussions about grammatical generation, interpretation and interpretability, viz. Chopin's C major mazurka, Op. 67 No. 3. We looked at this piece closely earlier in the chapter, but we did not discuss its B

section, i.e. mm. 33-40, then – so we will come full circle by exploring that section now. Therefore, I will end this chapter with a brief discussion of that very section in the C major Op. 67 No. 3 mazurka.

As is typical of B sections in rounded binary forms, the B section of the C major mazurka is an 8-measure long “contrasting middle” (Caplin (1998): 75-81). Contrasting middles normally prolong dominant harmony, and so does this passage. However, Chopin prolongs the dominant of the dominant for most of this passage, resolving this to dominant harmony only at the very end of the passage. To understand this, consider Example 2.3-23, the last example of this chapter, which illustrates how the C major mazurka’s B section is generated. Level A in the example shows us how the main harmonic structure of the passage is a retroactive prolongation of G major by its dominant. Now this prolongation can be interpreted either as a prolongation of I by V in G major, or as a prolongation of V by a V/V in C major (the home key of the piece). Either interpretation is valid, but the question is which interpretation accords with how the grammatical structure of the passage is interpretable? The passage is interpretable as being in G major if it is understood as a closed structure, i.e. as having at least a beginning and an ending. However, the reading in C major would probably be how the passage is interpretable, if it is understood as an open structure. So, the real issue here is whether we should see this mazurka’s B section as a grammatically self-contained structure, or one that should be understood in the context of the entire piece’s structure. Normally, this involves deciding whether the B section is a complete grammatical ‘sentence’, or only part of a larger grammatical ‘sentence’ that is the whole piece. And as I discussed in section 1.2.2 of chapter 1.2, we do not seem to have a satisfactory answer to this question.

However, in the case of this B section, the fact that it is actually part of a larger ‘sentence’ is quite clear from a crucial fact about its grammatical structure, which is why level A in the example analyzes this passage as being in C, and not G, major. This crucial fact is that the final G chord in the passage is actual a dissonant G^7 chord – meaning that it is really a V^7 of C major. This means further that the B section, which prolongs G major, is actually prolonging V/V in C major – and should therefore be interpreted as an open structure in C major (normally followed by a closing passage in C major). Another

Example 2.3-23. Chopin, Mazurka in C major, Op. 67 No. 3: B-section generation, mm. 33-40

A

C major: V/V V⁷

B

V/V⁸ - 7 V⁷

C

V/V³ V⁷

8 - 7 - 6 - #7 - 8 - 7
5 - 5 - 6 - b6 - 5
3 - 3 - 4 - 4 - V³

D

V/V V⁷

8 - 7

mm. 33 36 39

E

8 - 7

interesting fact in this regard, is that in our discussion of the A section of this piece, the compound period that this A section is comprised of was interpreted as a formally closed structure too. This means that the entire mazurka is made up of two closed structures – two ‘sentences’ so to speak – which perfectly accords with its formal categorization as a rounded *binary* form.

Moving on to level B of Example 2.3-23, we see that the initial D major harmony of the passage is transformed into a V^7 of G major, by means of the 8-7 motion in the soprano over the bass. This clarifies this initial sonority as being a V of G major, and not a I in D major – although this possibility seems to have arisen in a number of Chopin mazurkas now, given his preference for opening phrases with non-tonic harmony. The transformation of V of G major to V^7 of G major is further elaborated in level C of the image, by means of voice-leading motions in all three upper voices of the V chord. This is reminiscent of the way the initial tonic harmony of the A section of this piece was elaborated exclusively through voice-leading motions in the upper voices too, which we discussed in Example 2.3-2, also in level C of that example. Subsequently, in level D, we see how this voice-leading expansion of the D major harmony is repeated twice to flesh out the 8-bar structure of the B section, before the harmony is transformed into a V^7 of G major through 8-7 motion over the bass.

Finally, level E shows us the surface of the B section, where we see how the voice-leading motions that expanded the initial D major harmony of the passage are themselves expanded by passing and neighboring tones, which realize a number of soprano-tenor voice exchanges as well. All of which means that the G major harmony that is being prolonged in this entire passage is not itself attained until the very end of the passage, when its prolonging V^7 harmony finally resolves to it in measure 40. This is what reinforces the feeling of this passage as being in G major. But as we saw above, when G major is attained it is manifested as a V^7 of C harmony, which ultimately refutes a G major reading of this passage.

Of course, there are passages in music that are not easily interpretable, as we saw in some of the mazurka examples discussed in this chapter. And this is not surprising given that how a musical passage is interpretable can be a genuinely ambiguous matter at times – just as it is in language, as we saw in our

discussion of Example 1.2-38 in the second chapter of this dissertation. But the important point to remember, and which has been the focus of this chapter – and more generally of this dissertation – is that this depends on the grammatical structure of a passage. When we take an internalist attitude to the study of music and language, as is inherent in the Minimalist approach being advocated by this dissertation, then we can see just how and why something is interpretable, which is different from saying that just because something can be *interpreted*, it must therefore be inherently ambiguous. Musical interpretation and musical interpretability are different things. But the only way to realize this is by undertaking an internalist study of how musical structures are grammatically generated – in other words, by taking a Minimalist approach to the study of generative musical grammar.

Epilogue: An Unanswered Question?

Though much of this dissertation was involved with the age-old attempt to compare music with language, in a sense its aim was really to understand music itself – to answer Leonard Bernstein’s Unanswered Question discussed in chapter 1.1. But looking at music through a linguistic lens, especially a Minimalist one, has hopefully answered at least parts of this question now – while also reaffirming the connection music does have to language in the process.

So, as this project draws to a close, one could perhaps hazard an answer to the above question, in which music’s economical computational form, its free and creative nature, and its organic internalist connection to language will all play a part. However, the project this dissertation undertook is a novel, and nascent one, so there is no doubt that many parts of the Unanswered Question still remain unanswered. Much remains to be done to develop a full-fledged generative theory of music, and especially a joint Minimalist Program for music *and* language. And ultimately, all of this will have to engage with questions about how our musical minds evolved, how they interact with the brain and other parts of our bodies, and how the various cultural institutions that music gives rise to relate to this psychological foundation.

For the time being, I would like to describe just two possible directions this project might take in the immediate future, to answer two questions that are potentially more tangible than the ‘big’ ones mentioned in the last paragraph.

The first direction this project might take is to develop a more comprehensive theory of the musical lexicon than the one proposed here. There are only two entities that are conceptually necessary within a Minimalist approach to grammar, (1) the lexicon, and (2) the LF and PF levels of representation. Only these two entities are conceptually necessary because this is all that a Merge-based grammatical architecture needs to meet the external conditions imposed on it by extra-grammatical systems – Merge

takes items from the lexicon, and merges them into two-member sets, and repeats this, both externally and internally (i.e. transformationally) until an S-structure is derived, which is nothing but an LF object when its phonetic information is abstracted away, and a PF object when its semantic information is abstracted away. But for this reason, a Minimalist model of *musical* grammar needs a much richer description of the lexicon than the one proposed in this dissertation, or else Merge will not have anything to merge to begin with. I proposed chords as being the main components of the musical lexicon in chapter 1.2, but much more work needs to be done to verify this proposal. Another possibility is to see if one can essentially ‘get rid’ of the lexicon altogether within a Minimalist framework, since that would make it unnecessary even to justify chords as lexical items. A proposal worth exploring in this regard is Morris Halle and Alec Marantz’s “distributed morphology” framework (Halle and Marantz (1993)), in which lexical items do not exist as theoretical primitives. Instead both complex lexical items *and* complex phrasal structures (i.e. sentences) are derived, according to this framework, via the operations of the computational system. In this sense, there is no artificial division between words and phrases/sentences in this framework, which means that one does not have to justify an independent lexicon as a conceptually necessary primitive in this framework either. So, a distributed morphological approach to music might possibly allow one to do away with the need to define a musical lexicon altogether, chordal or otherwise.

In addition to the lexicon, the LF and PF levels of representation are the only other entities that are conceptually necessary for a Minimalist approach to generative grammar. So, further work in a Minimalist approach to generative musical grammar in the immediate future might focus on developing richer descriptions of the musical equivalents of these levels of representation.

In chapter 2.1, I discussed how a richer description of a musical LF level of structure might be achieved by including ideas from William Caplin’s theory of formal functions in a grammatical theory of musical meaning. The situation is a bit more complicated, however, if one wants to develop a richer description of a musical PF level of structure. In chapter 2.2, my discussion of a musical PF level of structure focused only on the rhythmic aspects of such a structure. But in language, PF objects have more

to do with the *sound* structure of language, i.e. speech sounds – or in the case of sign language, bodily gestures. For example, linguists have long observed that speech sounds can be categorized according to certain universal features, known as “distinctive features”, which are units of speech sound associated with certain articulatory or acoustic cues, such as lip rounding, voicing, and nasality. The distinctive feature theory was first proposed by Roman Jakobson, Gunnar Fant and Morris Halle, who detected twelve distinctive features in the world’s languages, and was later improved by Halle and Noam Chomsky, who tried to describe a universal set of distinctive features, and a grammatical procedure that maps the outputs of grammar to these features (Jakobson, Fant and Halle (1951), Chomsky and Halle (1968)). These ideas can therefore be seen to predict the later Minimalist idea that grammar generates S-structures that map to phonology via certain PF objects in language – these PF objects now being understandable, to put it crudely, as strings of distinctive features, which are strung together as the final output of the workings of Merge.

Importantly, distinctive features are not sounds themselves, but abstract representations of speech sounds and gestures (i.e. acoustic and articulatory cues) associated with them. The reason for this lies in the fact that linguistic phonology is quite structured and regular, which it has to be for grammar to map to it, whereas actual speech sounds show a great deal of variety – there being no one-to-one mapping, for instance, between specific distinctive features and specific acoustic and articulatory cues, meaning that one cannot develop the kind of universal theory of phonology using acoustic and articulatory cues (i.e. using actual sounds and gestures) that one can with the more abstract and patterned distinctive features that represent them (Halle (1983): 94). This implies that when grammar maps to phonology, the distinctive feature-based PF level of representation allows for *phonetic interpretability* – but whether this structure is actually interpreted, i.e. whether it is actually realized in the form of speech sounds or gestures is a performance matter that is subject to the vagaries of the sensorimotor system.

Now, given the importance of sound for music, the sound structure of music is something that one should be able to make numerous statements about. For a Minimalist approach to music, one should ideally be

able to think of a structured musical phonology, made up of abstract representations of *musical* sounds, to which musical grammar maps via the workings of Merge. But the structure of musical sounds rarely enters into generative descriptions of musical grammar. Generative music theory does deal with chords and pitches in an abstract way, but not in terms of a set of distinctive features, which could constitute a structured phonology of music. Therefore, this is one avenue for future research that is both crucial and opportune.

One possibility in this regard is to develop a theory of timbre within a generative theory of music. Timbre is defined as “that attribute of auditory sensation, in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar” (Moore (2012): 430). In simpler terms, timbre is the characteristic quality that a sound has, based on various acoustic properties such as its spectral envelope, as a result of which it can help us distinguish one musical sound from another. But timbral contrasts are also unsystematic, and vary from language to language, as a result of which Jakobson, Fant and Halle dropped them from the list of acoustic cues on which they based their theory of distinctive features (although timbre did play a part in the earlier phonological theories of Nikolai Trubetzkoy, especially in describing the phonology of vowels). Even then, timbre might play a more significant role in music than it does in linguistic phonology (e.g. see Patel (2007): 9-11). It has yet to play a significant role in generative theories of musical structure, but these theories have also generally neglected certain musical idioms in which timbre is known to play a significant role. For example, timbre plays a role in helping us distinguish the percussive sounds made by drumming instruments – and drum music has been under-theorized in the generative approach. So, maybe a generative exploration of a drumming idiom, e.g. one of the West African “talking drum” traditions, might reveal an important role for timbre within a generative theory of musical structure, and this might facilitate the development of a structured musical phonology too.

One final avenue of future research into a structured musical phonology could be one that focuses on the categorical nature of perception. Categorical perception refers to the phenomenon in which a type of stimulus, say speech sounds, whose exemplars seem to exist on a continuum of some kind, are actually

perceived in terms of groups or categories. For example, speech sounds associated with the distinctive feature of voicing (called “plosives”) seem to exist on a continuum according to a quantity known as “voice-onset time” or VOT. VOT is often defined as the length of time between the release of air by the lips (i.e. “articulatory” release) and the beginning of vocal fold vibration. So, some plosives have short VOTs, meaning that the vocal folds begin to vibrate very soon after articulatory release during the pronunciation of these plosives, whereas others have high VOTs, meaning that there is a delay between articulatory release and the subsequent onset of vocal fold vibration. (Sometimes the vocal folds can begin to vibrate *before* articulatory release too, leading to a negative VOT value.) All of this implies that speech sounds associated with voicing can exist on a continuum marked by increasingly positive VOT values.

However, we do not *perceive* plosives on a continuum. Instead, we tend to perceive all plosives with a ‘low’ VOT as being essentially the same, and all plosives with a ‘high’ VOT as being essentially the same too – i.e. we perceive voicing categorically. Low VOT plosives are referred to as “voiced” (in English, this is represented by the phone [b], since all these plosives sound like a “b”, irrespective of their actual VOT), and high VOT plosives are referred to as “voiceless” (represented by the phone [p], since all these plosives sound like a “p”, again irrespective of their actual VOT) – with the dividing line between the two categories occurring, in English, at a VOT of approximately 20 ms. It is this binary categorization of speech sounds, e.g. into voiced/voiceless, that influenced linguists like Jakobson and Halle to think of distinctive features in binary terms too, so that the abstract representation of speech sounds (which is what distinctive features are) is done in terms of + or – feature values.

In musical terms, all of this relates to another acoustic phenomenon in music (like timbre) that has been under-theorized in generative approaches to music – but which might be of relevance for a structured theory of musical phonology. This is the phenomenon of *intonation*. Intonation plays a big role in actual music making. Many musical practices emphasize singing/playing in tune, or least tuning one’s instruments correctly prior to performance. At the very least, playing in tune is required for proper

musical expression because playing significantly out of tune would distort a musical passage, and prevent listeners from comprehending it.

However, the definition of what is ‘in tune’ varies from idiom to idiom, and from performance tradition to performance tradition. This is because musical sounds themselves exist in a continuum, which allows different musicians in different idioms to make sounds that exist along a continuum, and which therefore allows them to define what intonation is correct at different places along this continuum in different idioms.

Despite the continuous nature of intonation, the *perception* of intonation, however, seems to be categorical. A good example of this can be found in the use of microtonal pitches in North Indian Classical music. The significance of microtonal pitches (called “*sruti*”) in this idiom has always been a source of controversy, some scholars denying their importance, others advocating them. Much of this has to do with the fact that microtones are a source of great melodic and expressive variety in performance, which makes *rāgas* that contain them vary considerably too – which, in turn, makes it hard to theorize about *rāga* structure systematically, i.e. when their microtonal pitch aspects are included in such theorizing. But despite the variety of microtones within and across *rāgas*, which undoubtedly arises from the continuum of sounds from which the pitches of a *rāga* are derived, these microtones are normally perceived categorically, as members of abstract scale degree categories, such as “Re” or “Fa” – which is why microtonal pitches in a *rāga* are named with the standard *solfege* symbols of North Indian music (*Sa*, *Re*, *Ga*, *Ma*, *Pa*, *Dha*, and *Ni*). In other words, all microtonal pitches in *rāgas* are perceived as variants of one of the seven scale steps in a *rāga*, in the way that plosives are all perceived as variants of either the [b] or [p] phones.

Moreover, the ability to perceive sounds categorically also seems to be a sign of native competence in an idiom, just as it is in language. For example, people who do not have native competence in North Indian Classical music are often unable to distinguish between a scale degree flat-2 pitch and one that is microtonally flatter than it, and it is not uncommon for someone to feel that the music of another idiom is out of tune because of these microtonal variations in the pitch structure of the idiom – which

means that such a listener is unable to perceive such microtonal pitches categorically, i.e. as really being instances of a standard scale step such as scale degree flat-2. This is similar to the situation in language, where non-native speakers of a language often have difficulty in perceiving the speech sounds of that language categorically, in the way native speakers of the language would. A good case in point is the inability of many non-native speakers of English to distinguish between the [l] and [r] sounds, as being categorically distinct, and members of the separate phoneme categories /l/ and /r/.

The above suggests that categorical perception might have a role to play in a phonological theory of music, and in an enriched discussion of how grammar interfaces with phonology in music. So, one could argue, for example, that microtones do play a role in musical phrase structure, but in the generation of S-structures, which only depends on relationships between grammatical entities, and not sounds. When a musical PF level of representation arises from this generative procedure, i.e. when a ‘sound image’ of a phrase is generated, the PF objects that this is made up of could then include microtonal pitches too, thus allowing them to be articulated in actual performance – as long as these microtonal pitches are understood as members of certain pitch categories, which is presumably what musical PF objects are. This might be what allows different *rāgas* in Indian music to have idiosyncratic microtonal structures, and what allows different performers performing the same *rāga* to sing or play individual scale steps in a *rāga* in different, idiosyncratic ways. All of which would make a phonological theory of music, specifically a theory of musical PF, indispensable for describing the role of microtones in music, and for describing the role of intonation in music more generally.

All in all, there is much more work to be done in a Minimalist approach to musical grammar, and there are several directions this project might take in the near future – some of which I have discussed in this Epilogue. It is my hope that this dissertation has at least demonstrated the feasibility and value of this project, and that more research will be undertaken within this new paradigm in music-theoretic and linguistic scholarship as a result – in order to help us further understand music and language, and how they relate, as fundamental aspects of human nature.

Bibliography

- Abbate, Carolyn, and Roger Parker (2012) *A History of Opera*. New York: W. W. Norton.
- Abert, Hermann (2007) *W.A. Mozart*. (Ed. Cliff Eisen, Trans. Stewart Spencer from *Wolfgang Amadeus Mozart: eine Biographie* (1923-24)). New Haven: Yale University Press.
- Abney, Steven (1987) "The English Noun Phrase in Its Sentential Aspect." (PhD Dissertation, MIT).
- Abraham, Gerald (1939) *Chopin's Musical Style*. London: Oxford University Press.
- Adorno, Theodor W. (1998) "Music and Language: A Fragment." In *Quasi Una Fantasia: Essays on Modern Music*. (Trans. & Ed. Rodney Livingstone from *Quasi una Fantasia* (1963)). London: Verso, pp. 1-8.
- Agawu, V. Kofi (1989) "Schenkerian Notation in Concept and Practice." *Music Analysis* 8/3, pp. 275-301.
- Agawu, V. Kofi (1990) "Variation Procedures in Northern Ewe Song." *Ethnomusicology* 34/2, pp. 221-243.
- Agawu, V. Kofi (1991) *Playing with Signs: A Semiotic Interpretation of Classic Music*. Princeton: Princeton University Press.
- Agawu, V. Kofi (1992a) "Theory and Practice in the Analysis of the Nineteenth-Century Lied." *Music Analysis* 11/1, pp. 3-36.
- Agawu, V. Kofi (1992b) "Schubert's Sexuality: A Prescription for Analysis?" *19th-Century Music* 17/1, pp. 79-82.
- Agawu, V. Kofi (1994) "Ambiguity in Tonal Music: A Preliminary Study." In *Theory, Analysis and Meaning in Music*. (Ed.) Anthony Pople. Cambridge: Cambridge University Press, pp. 86-107.
- Agawu, V. Kofi (1995) *African Rhythm: A Northern Ewe Perspective*. Cambridge: Cambridge University Press.
- Agawu, V. Kofi (2000-01) "Embracing the Non-West." *Integral* 14/15, pp. 63-66.
- Agawu, V. Kofi (2003) *Representing African Music: Postcolonial Notes, Queries, Positions*. New York: Routledge.
- Agawu, V. Kofi (2004) "How We Got out of Analysis, and How to Get Back in Again." *Music Analysis* 23/2-3, pp. 267-286.
- Agawu, V. Kofi (2006) "Structural Analysis or Cultural Analysis? Competing Perspectives on the 'Standard Pattern' of West African Rhythm." *Journal of the American Musicological Society* 59/1, pp. 1-46.

Agawu, V. Kofi (2008a) "Topic Theory: Achievement, Critique, Prospects." In *Passagen/IMS Congress Zürich 2007: Fünf Hauptvorträge*. (Ed.) Laurenz Lütteken & Hans-Joachim Hinrichsen. Zurich: Bärenreiter, pp.

Agawu, V. Kofi (2008b) *Music as Discourse: Semiotic Adventures in Romantic Music*. New York: Oxford University Press.

Agawu, V. Kofi (forthcoming) "The Many Meanings of Meaning in Music." In *On Meaning*. (Ed.) Jose V. Ciprut. Cambridge: MIT Press, pp. unknown.

Agmon, Eytan (1997) "The Bridges That Never Were: Schenker on the Contrapuntal Origin of the Triad and Seventh Chord." *Music Theory Online* 3/1.

Albright, Daniel (1999) *Untwisting the Serpent: Modernism in Music, Literature, and Other Arts*. Chicago: University of Chicago Press.

Aldwell, Edward, and Carl Schacter (2011) *Harmony and Voice Leading (4th Ed.)*. New York: Schirmer.

Allanbrook, Wye J. (1992) "Two Threads through the Labyrinth: Topic and Process in the First Movements of K. 332 and K. 333." In *Convention in Eighteenth- and Nineteenth-Century Music: Essays in Honor of Leonard G. Ratner*. (Ed.) Wye J. Allanbrook, Janet M. Levy & William P. Mahrt. Stuyvesant: Pendragon Press, pp. 125-171.

Allanbrook, Wye J. (2008) "Mozart's K 331, First Movement: Once More, with Feeling." In *Communication in Eighteenth-Century Music*. (Ed.) Danuta Mirka & Kofi Agawu. Cambridge: Cambridge University Press, pp. 255-282.

Alpern, Wayne (1999) "Music Theory as a Mode of Law: The Case of Heinrich Schenker, Esq." *Cardozo Law Review* 20/5-6, pp. 1459-1511.

Arom, Simha (1991) *African Polyphony and Polyrhythm*. (Foreward by Gyorgy Ligeti, Trans. Martin Thom, Barbara Tuckett & Raymond Boyd from *Polyphonies et Polyrhythmies Instrumentales d'Afrique Centrale* (1985)). Cambridge: Cambridge University Press.

Arom, Simha (2001) "Prolegomena to a Biomusicology." In *The Origins of Music*. (Ed.) Nils L. Wallin, Björn Merker & Steven Brown. Cambridge: MIT Press, pp. 27-30.

Babbitt, Milton B. (1952) "Review: *Structural Hearing* by Felix Salzer." *Journal of the American Musicological Society* 5/3, pp. 260-265.

Babbitt, Milton B. (1965) "The Structure and Function of Musical Theory: I." *College Music Symposium* 5, pp. 49-60.

Babbitt, Milton B., and Ev Grimes (1986) "Conversations with American Composers: Ev Grimes Interviews Milton Babbitt." *Music Educators Journal* 72/9, pp. 52-53.

Bach, Emmon, and Barbara H. Partee (1980) "Anaphora and Semantic Structure." In *Papers from the Parasession on Pronouns and Anaphora*. (Ed.) Jody Kreiman & Almerindo E. Ojeda. Chicago: Chicago Linguistic Society, pp. 1-28.

- Baker, Mark C. (2001) *The Atoms of Language: The Mind's Hidden Rules of Grammar*. New York: Basic Books.
- Baraka, Amiri, and LeRoi Jones (1963) *Blues People: Negro Music in White America*. New York: William Morrow.
- Bashwiner, David (2009) "What is Musical Syntax? An Evolutionary Perspective." (Paper presented at the Twentieth Annual Conference of Music Theory Midwest, Minneapolis).
- Beach, David W. (1985) "The Current State of Schenkerian Research." *Acta Musicologica* 57/2, pp. 275-307.
- Berger, Alexander A., Ryan May, Jordan B. Renner, Neal Viradia, and Laurence E. Dahners (2011) "Surprising Evidence of Pelvic Growth (Widening) after Skeletal Maturity." *Journal of Orthopaedic Research* 29/11, pp. 1719-1723.
- Berkowitz, Aaron L. (2010) *The Improvising Mind: Cognition and Creativity in the Musical Moment*. New York: Oxford University Press.
- Bernstein, Leonard (1976) *The Unanswered Question: Six Talks at Harvard*. Cambridge: Harvard University Press.
- Berwick, Robert C., and Noam Chomsky (2011) "The Biolinguistic Program: The Current State of Its Development." In *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*. (Ed.) Anna Maria Di Sciullo & Cedric Boeckx. New York: Oxford University Press, pp. 19-41.
- Bever, Thomas G. (1970) "The Cognitive Basis for Linguistic Structures." In *Cognition and Language Development*. (Ed.) John R. Hayes. New York: Wiley, pp. 279-362.
- Bever, Thomas G., and Donald T. Langendoen (1971) "A Dynamic Model of the Evolution of Language." *Linguistic Inquiry* 2/4, pp. 433-463.
- Bhatkhande, Vishnu N. (1909-32) *Hindustānī Sangīta Paddhatī*. Hathras: Sangeet Karyalaya.
- Bhatkhande, Vishnu N. (1913-37) *Kramik Pustak Mālikā*. Hathras: Sangeet Karyalaya.
- Bickerton, Derek (2001) "Can Biomusicology Learn from Language Evolution Studies?" In *The Origins of Music*. (Ed.) Nils L. Wallin, Björn Merker & Steven Brown. Cambridge: MIT Press, pp. 153-164.
- Bigand, Emmanuel (1990) "Abstraction of Two Forms of Underlying Structure in a Tonal Melody." *Psychology of Music* 18/1, pp. 45-59.
- Bigand, Emmanuel, Richard Parncutt, and Fred Lerdahl (1996) "Perception of Musical Tension in Short Chord Sequences: The Influence of Harmonic Function, Sensory Dissonance, Horizontal Motion, and Musical Training." *Perception and Psychophysics* 58/1, pp. 124-141.
- Bigand, Emmanuel, Pierre Perruchet, and Maud Boyer (1998) "Implicit Learning of an Artificial Grammar of Musical Timbres." *Current Psychology of Cognition* 17/3, pp. 577-600.
- Bigand, Emmanuel, and Richard Parncutt (1999) "Perceiving Musical Tension in Long Chord Sequences." *Psychological Research* 62, pp. 237-254.

- Bigand, Emmanuel, Barbara Tillmann, Bénédicte Poulin-Charronnat, and D. Manderlier (2005) "Repetition Priming: Is Music Special?" *Quarterly Journal of Experimental Psychology* 58A/8, pp. 1347-1375.
- Bigand, Emmanuel, and Bénédicte Poulin-Charronnat (2006) "Are We "Experienced Listeners"? A Review of the Musical Capacities That Do Not Depend on Formal Musical Training." *Cognition* 100, pp. 100-130.
- Birken, Marcia, and Anne C. Coon (2008) *Discovering Patterns in Mathematics and Poetry*. Amsterdam: Rodopi.
- Blacking, John (1973) *How Musical is Man?* Seattle: University of Washington Press.
- Blacking, John (1990) "Transcultural Communication and the Biological Foundations of Music." In *La Musica Come Linguaggio Universale: Genesi e Storia di Un'idea* (Ed.) Raffaele Pozzi. Florence: Olschki, pp. 179-188.
- Blacking, John (1992) "The Biology of Music Making." In *Ethnomusicology: An Introduction*. (Ed.) Helen Myers. New York: W. W. Norton, pp. 301-314.
- Blasius, Leslie D. (1996) *Schenker's Argument and the Claims of Music Theory*. New York: Cambridge University Press.
- Blasius, Leslie D. (1997) *The Music Theory of Godfrey Winham*. Princeton: Princeton University Press.
- Boeckx, Cedric (2006) *Linguistic Minimalism: Origins, Concepts, Methods, and Aims*. New York: Oxford University Press.
- Boeckx, Cedric, Norbert Hornstein, and Jairo Nunes (2010) *Control as Movement*. New York: Cambridge University Press.
- Boer, Joep, Suvarnalata Rao, Wim van der Meer, and Jane Harvey (1999) *The Raga Guide: A Survey of 74 Hindustani Ragas*. Wyastone: Nimbus Records.
- Bohlman, Philip V. (1992) "Ethnomusicology's Challenge to the Canon: The Canon's Challenge to Ethnomusicology." In *Disciplining Music: Musicology and Its Canons*. (Ed.) Katherine Bergeron & Philip V. Bohlman. Chicago: University of Chicago Press, pp. 116-136.
- Boilés, Charles L. (1967) "Tepehua Thought-Song: A Case of Semantic Signaling." *Ethnomusicology* 11/3, pp. 267-292.
- Borchmeyer, Dieter (2003) *Drama and the World of Richard Wagner*. Princeton: Princeton University Press.
- Boretz, Benjamin A. (1970) "The Construction of Musical Syntax (I)." *Perspectives of New Music* 9/1, pp. 23-42.
- Boretz, Benjamin A. (1971) "Musical Syntax (II)." *Perspectives of New Music* 10/1, pp. 232-270.
- Boretz, Benjamin A. (1972) "Meta-Variations, Part 4: Analytical Fallout (I)." *Perspectives of New Music* 11/1, pp. 146-223.

- Boretz, Benjamin A. (1973) "Meta-Variations, Part 4: Analytical Fallout (II)." *Perspectives of New Music* 11/2, pp. 156-203.
- Boyden, Alan (1943) "Homology and Analogy: A Century after the Definitions Of "Homologue" and "Analogue" Of Richard Owen." *Quarterly Review of Biology* 18/3, pp. 228-241.
- Bresnan, Joan (1972) "Theory of Complementation in English Syntax." (PhD Dissertation, MIT).
- Brett, Philip (1994) "Musicality, Essentialism and the Closet." In *Queering the Pitch: The New Gay and Lesbian Musicology*. (Ed.) Philip Brett, Elizabeth Wood & Gary Thomas. New York: Routledge, pp. 9-26.
- Brockman, John (1995) *The Third Culture: Beyond the Scientific Revolution*. New York: Touchstone.
- Brown, Matthew G. (2005) *Explaining Tonality: Schenkerian Theory and Beyond*. Rochester: University of Rochester Press.
- Brown, Matthew G., and Douglas J. Dempster (1989) "The Scientific Image of Music Theory." *Journal of Music Theory* 33/1, pp. 65-106.
- Brown, Steven (2001a) "The 'Musilanguage' Model of Music Evolution." In *The Origins of Music*. (Ed.) Nils L. Wallin, Björn Merker & Steven Brown. Cambridge: MIT Press, pp. 271-300.
- Brown, Steven (2001b) "Are Music and Language Homologues?" *Annals of the New York Academy of Sciences* 930, pp. 372-374.
- Brown, Steven, Björn Merker, and Wallin, Nils L. (2001) "An Introduction to Evolutionary Musicology." In *The Origins of Music*. (Ed.) Nils L. Wallin, Björn Merker & Steven Brown. Cambridge: MIT Press, pp. 3-24.
- Brown, Steven, Michael J. Martinez, and Lawrence M. Parsons (2006) "Music and Language Side by Side in the Brain: A PET Study of the Generation of Melodies and Sentences." *European Journal of Neuroscience* 23, pp. 2791-2803.
- Burkhart, Charles (1978) "Schenker's Motivic Parallelisms." *Journal of Music Theory* 22/2, pp. 145-175.
- Burling, Robbins (1965) "How to Choose a Burmese Numerical Classifier." In *Context and Meaning in Cultural Anthropology*. (Ed.) Melford E. Spiro. London: Collier-Macmillan, pp. 243-264.
- Burstein, L. Poundie (2010) "Schenkerian Analysis and Occam's Razor." (Paper presented at the Sixth International Conference on Music Theory, Tallinn, Estonia).
- Cadwallader, Allen, and David Gagne (1988) *Analysis of Tonal Music: A Schenkerian Approach*. New York: Oxford University Press.
- Caplin, William E. (1998) *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart and Beethoven*. New York: Oxford University Press.
- Caplin, William E. (2005) "On the Relation of Musical Topoi to Formal Function." *18th-Century Music* 2/1, pp. 113-124.
- Carnap, Rudolf (1952) "Meaning Postulates." *Philosophical Studies* 3/5, pp. 65-73.

- Carnie, Andrew (2002) *Syntax: A Generative Introduction*. Oxford: Blackwell.
- Castellano, Mary A., Jamshed J. Bharucha, and Carol L. Krumhansl (1984) "Tonal Hierarchies in the Music of North India." *Journal of Experimental Psychology: General* 113/3, pp. 394-412.
- Chib, Satyendra K. Sen (2004) *Companion to North Indian Classical Music*. New Delhi: Munshiram Manoharlal Publishers.
- Chomsky, Noam (1957) *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam (1959) "A Review of B. F. Skinner's Verbal Behavior." *Language* 35, pp. 26-58.
- Chomsky, Noam (1965) *Aspects of a Theory of Syntax*. Cambridge: MIT Press.
- Chomsky, Noam (1966) *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York: Harper & Row.
- Chomsky, Noam (1970) "Remarks on Nominalization." In *Readings in English Transformational Grammar*. (Ed.) Roderick A. Jacobs & Peter S. Rosenbaum. Waltham: Ginn, pp. 184-221.
- Chomsky, Noam (1973) "Conditions on Transformations." In *A Festschrift for Morris Halle*. (Ed.) Stephen Anderson & Paul Kiparsky. New York: Holt, Rinehart and Winston, pp. 232-286.
- Chomsky, Noam (1975) "Questions of Form and Interpretation." In *The Scope of American Linguistics*. (Ed.) Robert Austerlitz. Lisse: Peter de Ridder Press, pp. 159-196.
- Chomsky, Noam (1980a) *Rules and Representations*. Oxford: Blackwell.
- Chomsky, Noam (1980b) "On Binding." *Linguistic Inquiry* 11, pp. 1-46.
- Chomsky, Noam (1981) *Lectures on Government and Binding: The Pisa Lectures*. Berlin: De Gruyter.
- Chomsky, Noam (1986a) *Barriers*. Cambridge: MIT Press.
- Chomsky, Noam (1986b) *Knowledge of Language: Its Nature, Origin, and Use*. Westport: Praeger Publishing.
- Chomsky, Noam (1988) *Language and Problems of Knowledge: The Managua Lectures*. Cambridge: MIT Press.
- Chomsky, Noam (1991) "Linguistics and Cognitive Science: Some Problems and Mysteries." In *The Chomskyan Turn*. (Ed.) Asa Kasher. Oxford: Blackwell, pp. 26-53.
- Chomsky, Noam (1995a) "Introduction." In *The Minimalist Program*. Cambridge: MIT Press, pp. 1-12.
- Chomsky, Noam (1995b) "A Minimalist Program for Linguistic Theory." In *The Minimalist Program*. Cambridge: MIT Press, pp. 167-218. (Originally published in *The View from Building 20: Essays in Honor of Sylvain Bromberger* (Ed.) Kenneth L. Hale & Samuel J. Keyser. Cambridge: MIT Press (1993): 1-52).
- Chomsky, Noam (1995c) "Categories and Transformations." In *The Minimalist Program*. Cambridge: MIT Press, pp. 219-394. (Based on "Bare Phrase Structure", originally published in *Government Binding*

- Theory and The Minimalist Program* (Ed.) Gert Webelhuth. Oxford: Oxford University Press (1995): 383-439).
- Chomsky, Noam (2000) *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.
- Chomsky, Noam (2002) *On Nature and Language*. Cambridge: Cambridge University Press.
- Chomsky, Noam (2004) "Beyond Explanatory Adequacy." In *Structures and Beyond. The Cartography of Syntactic Structure Vol 3*. (Ed.) Adriana Belletti. Oxford: Oxford University Press, pp. 104-131.
- Chomsky, Noam (2005) "Three Factors in Language Design." *Linguistic Inquiry* 36/1, pp. 1-22.
- Chomsky, Noam (2006) *Language and Mind (3rd Ed.)*. New York: Cambridge University Press.
- Chomsky, Noam (2007) "Approaching UG from Below." In *Interfaces + Recursion = Language?: Chomsky's Minimalism and the View from Syntax-Semantics*. (Ed.) Uli Sauerland & Hans-Martin Gartner. Berlin: De Gruyter, pp. 1-29.
- Chomsky, Noam, and Morris Halle (1968) *The Sound Pattern of English*. New York: Harper Row.
- Clark, Alexander, and Remi Eyraud (2007) "Polynomial Time Identification in the Limit of Substitutable Context-Free Languages." *Journal of Machine Learning Research* 8, pp. 1725-1745.
- Clark, Eve V. (2003) *First Language Acquisition*. Cambridge: Cambridge University Press.
- Clarke, Eric F. (2005) *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford: Oxford University Press.
- Cohn, Richard L. (1992a) "Schenker's Theory, Schenkerian Theory: Pure Unity or Constructive Conflict." *Indiana Theory Review* 13/1, pp. 1-19.
- Cohn, Richard L. (1992b) "The Autonomy of Motives in Schenkerian Accounts of Tonal Music." *Music Theory Spectrum* 14/2, pp. 150-170.
- Cohn, Richard L. (1996) "Maximally Smooth Cycles, Hexatonic Systems, and the Analysis of Late-Romantic Triadic Progressions." *Music Analysis* 15/1, pp. 9-40.
- Cohn, Richard L. (1997) "Neo-Riemannian Operations, Parsimonious Trichords, and Their 'Tonnetz' Representations." *Journal of Music Theory* 41/1, pp. 1-66.
- Cohn, Richard L. (1998) "Introduction to Neo-Riemannian Theory: A Survey and a Historical Perspective." *Journal of Music Theory* 42/2, pp. 167-180.
- Cohn, Richard L. (1999) "As Wonderful as Star Clusters: Instruments for Gazing at Tonality in Schubert." *19th-Century Music* 22/3, pp. 213-232.
- Cohn, Richard L. (2004) "Uncanny Resemblances: Tonal Signification in the Freudian Age." *Journal of the American Musicological Society* 57/2, pp. 285-323.
- Cohn, Richard L. (2012) *Audacious Euphony: Chromatic Harmony and the Triad's Second Nature*. New York: Oxford University Press.

- Cohn, Richard L., and Douglas J. Dempster (1992) "Hierarchical Unity, Plural Unities: Toward a Reconciliation." In *Disciplining Music: Musicology and Its Canons*. (Ed.) Katherine Bergeron & Philip V. Bohlman. Chicago: University of Chicago Press, pp. 156-181.
- Collins, Christopher, and Edward Stabler (2012) "A Formalization of Minimalist Syntax." (Unpublished paper).
- Cone, Edward T. (1968) *Musical Form and Musical Performance*. New York: W. W. Norton.
- Cook, Nicholas (1989a) "Schenker's Theory of Music as Ethics." *Journal of Musicology* 7/4, pp. 415-439.
- Cook, Nicholas (1989b) "The Future of Theory " *Indiana Theory Review* 10, pp. 69-73.
- Cook, Nicholas (2007) *The Schenker Project: Culture, Race, and Music Theory in Fin-De-Siècle Vienna*. New York: Oxford University Press.
- Cooke, Deryck (1959) *The Language of Music*. Oxford: Oxford University Press.
- Cooper, Robin (1977) "Abstract Structure and the Indian Raga System." *Ethnomusicology* 21/1, pp. 1-32.
- Crain, Stephen, and Mineharu Nakayama (1986) "Structure Dependence in Grammar Formation." *Language* 63, pp. 522-543.
- Dahlhaus, Carl (1989) "Nationalism and Folk Music." In *Four Studies in the Music of the Later Nineteenth Century*. (Trans. Mary Whittall from *Zwischen Romantik und Moderne: Vier Studien zur Musikgeschichte des späteren 19. Jahrhunderts* (1974)). Berkeley & Los Angeles: University of California Press, pp. 90-94.
- Darwin, Charles (1871a) *The Descent of Man: And Selection in Relation to Sex, Vol 1*. London: John Murray.
- Darwin, Charles (1871b) *The Descent of Man: And Selection in Relation to Sex, Vol 2*. London: John Murray.
- DeBellis, Mark (2010) "Schenkerian Theory's Empiricist Image." *Music Theory Spectrum* 32/2, pp. 111-120.
- DeClercq, Trevor, and David Temperley (2011) "A Corpus Analysis of Rock Harmony." *Popular Music* 30/1, pp. 47-70.
- Dennett, Daniel C. (1995) *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon and Schuster.
- Derrida, Jacques (1987) *The Postcard: From Socrates to Freud and Beyond*. (Trans. & Ed. Alan Bass from *La carte postale: De Socrate à Freud et au-delà* (1980)). Chicago: University of Chicago Press.
- Derrida, Jacques (1998) *Of Grammatology*. (Trans. & Ed. Gayatri Chakravorty Spivak from *De la Grammatologie* (1967)). Baltimore: Johns Hopkins University Press.
- Descartes, Rene (1988) *Discourse on the Method*. In *Descartes: Selected Philosophical Writings*. (Trans. Robert Stoothoff from *Discours de la Méthode* (1637)). Cambridge: Cambridge University Press. pp. 20-56

- Deutsch, Diana (1982) *Psychology of Music*. New York: Academic Press.
- Dews, Peter K. (1994) "Adorno, Post-Structuralism and the Critique of Identity." In *Mapping Ideology*. (Ed.) Slavoj Žižek. London: Verso, pp. 46-65.
- Dowling, Walter J. and Dane L. Harwood (1986) *Music Cognition*. New York: Academic Press.
- Drake, Carolyn, and Daisy Bertrand (2003) "The Quest for Universals in Temporal Processing in Music." In *The Cognitive Neuroscience of Music*. (Ed.) Isabelle Peretz & Robert J. Zatorre. New York: Oxford University Press, pp. 21-31. (Originally published in *Annals of the New York Academy of Sciences* 930 (2001): 17-27).
- Dubiel, Joseph (1990) "When You Are a Beethoven: Kinds of Rules in Schenker's Counterpoint." *Journal of Music Theory* 34/2, pp. 291-340.
- Dubiel, Joseph (1999) "Composer, Theorist, Composer/ Theorist." In *Rethinking Music*. (Ed.) Nicholas Cook & Mark Everist. Oxford: Oxford University Press, pp. 262-286.
- Dummett, Michael A. E. (1996) "Language and Communication." In *The Seas of Language*. Oxford: Oxford University Press, pp. 166-187. (Originally published in Alexander L. George (Ed.), *Reflections on Chomsky*. Oxford: Basil Blackwell Press, 1989, 192-212).
- Elman, Jeffrey L. (1993) "Learning and Development in Neural Networks: The Importance of Starting Small." *Cognition* 48/1, pp. 71-99.
- Elman, Jeffrey L., Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim R. Plunkett (1996) *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge: MIT Press.
- Epstein, David (1979) *Beyond Orpheus: Studies in Musical Structure*. Cambridge: MIT Press.
- Evans, Gareth (1982) *The Varieties of Reference*. (Ed. John McDowell) Oxford: Oxford University Press.
- Everett, Daniel L. (2009) "Pirahã Culture and Grammar: A Response to Some Criticisms." *Language* 85/2, pp. 405-442.
- Everett, Walter T. (2001) *The Beatles as Musicians. Volumes 1 & 2*. New York: Oxford University Press.
- Falk, Dean (1990) "Brain Evolution in *Homo*: The "Radiator" Theory." *Behavioral and Brain Sciences* 13, pp. 333-381.
- Federhofer, Hellmut (1982) "Fux's "Gradus ad Parnassum" As Viewed by Heinrich Schenker." (Trans. Alfred Mann) *Music Theory Spectrum* 4, pp. 66-75.
- Fedorenko, Evelina, Michael K. Behr, and Nancy Kanwisher (2011) "Functional Specificity for High-Level Linguistic Processing in the Human Brain." *Proceedings of the National Academy of Sciences* 108/39 pp. 16428-16433.
- Feld, Steven (1974) "Linguistic Models in Ethnomusicology." *Ethnomusicology* 18/2, pp. 197-217.
- Feld, Steven (1982) *Sound and Sentiment: Birds, Weeping, Poetics and Song in Kaluli Expression*. Philadelphia: University of Pennsylvania Press.

- Feld, Steven, and Aaron A. Fox (1994) "Music and Language." *Annual Review of Anthropology* 23, pp. 25-53.
- Fillmore, Charles J. (1968) "The Case for Case." In *Universals in Linguistic Theory*. (Ed.) Emmon Bach & Robert T. Harms. New York: Holt, Rinehart, and Winston, pp. 1-88.
- Fitch, William T. (2006) "The Biology and Evolution of Music: A Comparative Perspective." *Cognition* 100/1, pp. 173-215.
- Fitch, William T., Marc D. Hauser, and Noam Chomsky (2005) "The Evolution of the Language Faculty: Clarifications and Implications." *Cognition* 97/2, pp. 179-210.
- Fodor, Jerry A., and Zenon W. Pylyshyn (1988) "Connectionism and Cognitive Architecture: A Critical Analysis." *Cognition* 28/1-2, pp. 3-71.
- Forte, Allen (1973) *The Structure of Atonal Music*. New Haven: Yale University Press.
- Forte, Allen, and Steven Gilbert (1982) *Introduction to Schenkerian Analysis*. New York: W. W. Norton.
- Frazier, Lyn, Lori Taft, Tom Roeper, Charles Clifton, and Kate Ehrlich (1984) "Parallel Structure: A Source of Facilitation in Sentence Comprehension " *Memory and Cognition* 12/5, pp. 421-430.
- Fromkin, Victoria A., Robert Rodman, and Nina M. Hyams (2009) *An Introduction to Language (9th Ed.)*. Boston: Wadsworth.
- Fux, Johann Joseph (1943) *Steps to Parnassus: The Study of Counterpoint*. (Partial Trans. Alfred Mann from the first part of Part 2 of *Gradus ad Parnassum* (1725)). New York: W. W. Norton.
- Ganger, Jennifer, and Michael R. Brent (2004) "Reexamining the Vocabulary Spurt." *Developmental Psychology* 40/4, pp. 621-632.
- Gauldin, Robert (1995) *A Practical Approach to Sixteenth-Century Counterpoint*. Long Grove: Waveland Press.
- Gazdar, Gerald J. M., Ewan H. Klein, Geoffrey K. Pullum, and Ivan A. Sag (1985) *Generalized Phrase Structure Grammar*. Oxford: Basil Blackwell.
- Gentner, Timothy Q., Kimberly M. Fenn, Daniel Margoliash, and Howard C. Nusbaum (2006) "Recursive Syntactic Pattern Learning by Songbirds." *Nature* 440/7088, pp. 1204-1207.
- Girard, Aaron R. (2007) "Music Theory in the American Academy." (PhD Dissertation, Harvard).
- Gjerdingen, Robert O. (1986) *A Classic Turn of Phrase: Music and the Psychology of Convention*. Philadelphia: University of Pennsylvania Press.
- Gjerdingen, Robert O. (1989) "Meter as a Mode of Attending: A Network Simulation of Attentional Rhythmicity in Music." *Integral* 3, pp. 67-91.
- Gjerdingen, Robert O. (1990) "Categorization of Musical Patterns by Self-Organizing Neuronlike Networks." *Music Perception* 7/4, pp. 339-370.

- Gjerdingen, Robert O. (2007) *Music in the Galant Style: Being an Essay on Various Schemata Characteristic of Eighteenth-Century Music for Courtly Chambers, Chapels, and Theaters, Including Tasteful Passages of Music Drawn from Most Excellent Chapel Masters in the Employ of Noble and Noteworthy Personages, Said Music All Collected for the Reader's Delectation on the World Wide Web*. New York: Oxford University Press.
- Goethe, Johann Wolfgang von (1891) *Goethes Werke: Herausgegeben Im Auftrage Der Großherzogin Sophie Von Sachsen: II. Abtheilung: Goethes Naturwissenschaftliche Schriften: 6. Band: Zur Morphologie: I. Theil*. (Ed. Bernhard Suphan & Rudolf Steiner) Weimar: Hermann Böhlau.
- Gould, Stephen J. (1991) "Exaptation: A Crucial Tool for Evolutionary Psychology." *Journal of Social Issues* 47/3, pp. 43-65.
- Gould, Stephen J., and Elizabeth S. Vrba (1982) "Exaptation: A Missing Term in the Science of Form." *Paleobiology* 8/1, pp. 4-15.
- Graffi, Giorgio (2001) *200 Years of Syntax: A Critical Survey*. Amsterdam: John Benjamins Publishing.
- Halle, Morris (1983) "On Distinctive Features and Their Articulatory Implementation." *Natural Language & Linguistic Theory* 1/1, pp. 91-105
- Halle, Morris, and Alec Marantz (1993) "Distributed Morphology and the Pieces of Inflection." In *The View from Building 20*. (Ed.) Kenneth Hale & S. Jay Keyser. Cambridge: MIT Press, pp. 111-176.
- Hanslick, Eduard (1986) *On the Musically Beautiful: A Contribution to the Revisal of Musical Aesthetics*. (Trans. Geoffrey Payzant from *Vom Musikalisch-Schönen* (1854)). Indianapolis: Hackett.
- Hasty, Christopher F. (1997) *Meter as Rhythm*. New York: Oxford University Press.
- Hatten, Robert S. (2004) *Interpreting Musical Gestures, Topics, and Tropes: Mozart, Beethoven, Schubert*. Bloomington: Indiana University Press.
- Hatten, Robert S. (2009) "Interpreting Beethoven's *Tempest* Sonata through Topics, Gestures and Agency." In *Beethoven's Tempest Sonata: Perspectives of Analysis and Performance*. (Ed.) Pieter Bergé, William E. Caplin & Jeroen d'Hoe. Leuven: Peeters, pp. 163-180.
- Hauser, Marc D., Elissa L. Newport, and Richard N. Aslin (2001) "Segmentation of the Speech Stream in a Non-Human Primate: Statistical Learning in Cotton-Top Tamarins." *Cognition* 78, pp. B53-B64.
- Hauser, Marc D., Noam Chomsky, and William T. Fitch (2002) "The Faculty of Language: What is It, Who Has It, and How Did It Evolve?" *Science* 298/5598, pp. 1569-1579.
- Hauser, Marc D., and Josh McDermott (2003) "The Evolution of the Music Faculty: A Comparative Perspective." *Nature Neuroscience* 6/7, pp. 663-668.
- Hinzen, Wolfram (2006) *Mind Design and Minimal Syntax*. Oxford: Oxford University Press.
- Hinzen, Wolfram (2011) "Emergence of a Systemic Semantics through Minimal and Underspecified Codes." In *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*. (Ed.) Anna Maria Di Sciullo & Cedric Boeckx. New York: Oxford University Press, pp. 417-439.

Hofstadter, Douglas R. (1979) *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books.

Hornstein, Norbert (1995) *Logical Form: From GB to Minimalism*. Oxford: Blackwell.

Huang, Cheng-Teh James (2006) "Move Wh in a Language without Wh-Movement." In *Syntax: Critical Concepts in Linguistics*. (Ed.) Robert Freidin & Howard Lasnik. New York: Routledge, pp. 397-440. (Originally published in *The Linguistic Review* 1 (1981-82): 369-416).

Hubel, David H., and Torsten N. Wiesel (1970) "The Period of Susceptibility to the Physiological Effects of Unilateral Eye Closure in Kittens." *Journal of Physiology* 206/2, pp. 419-436.

Humboldt, Wilhelm von (1999) *On Language. On the Diversity of Human Language Construction and Its Influence on the Mental Development of the Human Species*. (Trans. Peter Heath & Ed. Michael Losonsky from *Über die Verschiedenheit des menschlichen Sprachbaus und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts* (1836)). Cambridge: Cambridge University Press.

Huron, David B. (2001) "Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles." *Music Perception* 19/1, pp. 1-64.

Huron, David B. (2003) "Is Music an Evolutionary Adaptation?" In *The Cognitive Neuroscience of Music*. (Ed.) Isabelle Peretz & Robert J. Zatorre. New York: Oxford University Press, pp. 57-78. (Originally published in *Annals of the New York Academy of Sciences* 930 (2001): 43-61).

Huron, David B. (2006) *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge: MIT Press.

Hutchinson, John R., and Mariano Garcia (2002) "Tyrannosaurus Was Not a Fast Runner." *Nature* 415/6875, pp. 1018-1021.

Jackendoff, Ray S. (1977) *X-Bar Syntax: A Theory of Phrase Structure*. Cambridge: MIT Press.

Jackendoff, Ray S. (1999) "Parallel Constraint-Based Generative Theories of Language " *Trends in Cognitive Sciences* 3/10, pp. 393-400

Jackendoff, Ray S. (2002) *Foundations of Language: Brain, Meaning, Grammar, Evolution*. New York: Oxford University Press.

Jackendoff, Ray S. (2009) "Parallels and Nonparallels between Language and Music." *Music Perception* 26/1, pp. 195-204.

Jackendoff, Ray S., and Fred Lerdahl (1980) "Discovery Procedures vs Rules of Musical Grammar in a Generative Music Theory." *Perspectives of New Music* 18/1-2, pp. 503-510.

Jackendoff, Ray S., and Fred Lerdahl (2006) "The Capacity for Music: What Is It, and What's Special About It?" *Cognition* 100/1, pp. 33-72.

Jairazbhoy, Nazir A. (1972) "Factors Underlying Important Notes in North Indian Music." *Ethnomusicology* 16/1, pp. 63-81.

Jairazbhoy, Nazir A. (1995) *The Rags of North Indian Music: Their Structure and Evolution*. Bombay: Popular Prakashan.

- Jakobson, Roman, C. Gunnar M. Fant, and Morris Halle (1951) *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*. Cambridge.: MIT Press.
- Jenkins, Lyle (2000) *Biolinguistics: Exploring the Biology of Language*. Cambridge: Cambridge University Press.
- Jeppesen, Knud (1992) *Counterpoint: The Polyphonic Vocal Styles of the Sixteenth Century*. New York: Dover.
- Kandel, Eric R., James H. Schwartz, and Thomas M. Jessell (1991) *Principles of Neural Science, 3rd Ed.* New York: Elsevier.
- Kassler, Jamie C. (1983) "Heinrich Schenker's Epistemology and Philosophy of Music: An Essay on the Relations between Evolutionary Theory and Music Theory." In *The Wider Domain of Evolutionary Thought*. (Ed.) David R. Oldroyd & Ian Langham. Dordrecht: Reidel, pp. 221-260.
- Kassler, Michael (1967) "A Trinity of Essays: Toward a Theory That is the Twelve-Note-Class System. Toward Development of a Constructive Tonality Theory Based on Writings by Heinrich Schenker. Toward a Simple Programming Language for Musical Information Retrieval." (PhD Dissertation, Princeton).
- Kassler, Michael (1977) "Explication of the Middleground of Schenker's Theory of Tonality." *Miscellanea Musicologica* 9, pp. 72-81.
- Katz, Jerrold J. (1980) "Chomsky on Meaning." *Language* 56/1, pp. 1-41.
- Katz, Jonah, and David M. Pesetsky (2011) "The Identity Thesis for Language and Music." (Unpublished paper).
- Kayne, Richard S. (1994) *The Antisymmetry of Syntax*. Cambridge: MIT Press.
- Keiler, Allan R. (1977) "The Syntax of Prolongation (Part I)." *In Theory Only* 3/5, pp. 3-27.
- Keiler, Allan R. (1978a) "The Empiricist Illusion: Narmour's Beyond Schenkerism." *Perspectives of New Music* 17/1, pp. 161-195.
- Keiler, Allan R. (1978b) "Bernstein's *The Unanswered Question*" and the Problem of Musical Competence." *Musical Quarterly* 64/2, pp. 195-222.
- Keiler, Allan R. (1979-80) "Reply: Discovery Procedures vs Rules of Musical Grammar in a Generative Music Theory." *Perspectives of New Music* 18/1-2, pp. 511-516.
- Keiler, Allan R. (1981) "Two Views of Musical Semiotics." In *The Sign in Music and Literature*. (Ed.) Wendy Steiner. Austin: University of Texas Press, pp. 138-168.
- Keiler, Allan R. (1983-84) "On Some Properties of Schenker's Pitch Derivations." *Music Perception* 1/2, pp. 200-228.
- Keiler, Allan R. (1989) "The Origins of Schenker's Theory: How Man is Musical." *Journal of Music Theory* 33/2, pp. 273-298.

- Keiler, Allan R. (1996) "Melody and Motive in Schenker's Earliest Writings." In *Critica Musica: Essays in Honor of Paul Brainard*. (Ed.) John Knowles. Amsterdam: Gordon & Breach, pp. 169-191.
- Kerman, Joseph (1980) "How We Got into Analysis, and How to Get Out." *Critical Inquiry* 7/2, pp. 311-331.
- Kielian-Gilbert, Marianne C. (2003) "Interpreting Schenkerian Prolongation." *Music Analysis* 22/1-2, pp. 51-104.
- Kitagawa, Yoshihisa (1986) "Subjects in Japanese and English." (PhD Dissertation, University of Massachusetts, Amherst).
- Klein, Michael (2005) *Intertextuality in Western Art Music*. Bloomington: Indiana University Press.
- Klumpenhouwer, Henry (2006) "In Order to Stay Asleep as Observers: The Nature and Origins of Anti-Cartesianism in Lewin's Generalized Musical Intervals and Transformations." *Music Theory Spectrum* 28/2, pp. 277-289.
- Koelsch, Stefan (2010) "Towards a Neural Basis of Music-Evoked Emotions." *Trends in Cognitive Sciences* 14/3, pp. 131-137.
- Koelsch, Stefan, Thomas C. Gunter, D. Yves V. Cramon, Stefan Zysset, Gabriele Lohmann, and Angela D. Friederici (2002) "Bach Speaks: A Cortical 'Language-Network' Serves the Processing of Music." *NeuroImage* 17/2, pp. 956-966.
- Komar, Arthur J. (1971) *Theory of Suspensions: A Study of Metrical and Pitch Relations in Tonal Music*. Princeton: Princeton University Press.
- Koopman, Hilda (1984) *The Syntax of Verbs: From Verb Movement Rules in the Kru Languages to Universal Grammar*. Dordrecht: Foris Publications.
- Koopman, Hilda, and Dominique Sportiche (1991) "The Position of Subjects." *Lingua* 85/1, pp. 211-258.
- Kornai, András (2008) *Mathematical Linguistics*. London: Springer.
- Korsyn, Kevin (1988) "Schenker and Kantian Epistemology." *Theoria* 3, pp. 44-50.
- Korsyn, Kevin (1993) "Schenker's Organicism Reexamined." *Integral* 7, pp. 82-118.
- Korsyn, Kevin (2010) "Schenker's Vienna: Nicholas Cook on Culture, Race and Music Theory in *Fin-De-Siècle* Austria." *Music Analysis* 28/1, pp. 153-179.
- Kostka, Stefan, Dorothy Payne, and Byron Almen (2013) *Tonal Harmony (7th Ed.)*. New York: McGraw-Hill.
- Kramer, Lawrence (2001) *Musical Meaning: Towards a Critical History*. Berkeley & Los Angeles: University of California Press.
- Kripke, Saul A. (1980) *Naming and Necessity*. Cambridge: Harvard University Press.
- Krumhansl, Carol L. (1979) "The Psychological Representation of Musical Pitch in a Tonal Context." *Cognitive Psychology* 11/3, pp. 346-374.

- Krumhansl, Carol L. (1997) "An Exploratory Study of Musical Emotions and Psychophysiology." *Canadian Journal of Psychology* 51/4, pp. 336-352.
- Krumhansl, Carol L., and Roger N. Shepard (1979) "Quantification of the Hierarchy of Tonal Functions within a Diatonic Context." *Journal of Experimental Psychology: Human Perception and Performance* 5, pp. 579-594.
- Kuhl, Patricia K., and James D. Miller (1975) "Speech Perception by the Chinchilla: Voiced-Voiceless Distinction in Alveolar Plosive Consonants." *Science* 190/4209, pp. 69-72.
- Kuhl, Patricia K., and James D. Miller (1978) "Speech Perception by the Chinchilla: Identification Functions for Synthetic VOT Stimuli." *Journal of the Acoustical Society of America* 63/3, pp. 905-916.
- Kunej, Drago, and Ivan Turk (2001) "New Perspectives on the Beginnings of Music: Archeological and Musicological Analysis of a Middle Paleolithic Bone Flute." In *The Origins of Music*. (Ed.) Nils L. Wallin, Björn Merker & Steven Brown. Cambridge: MIT Press, pp. 235-268.
- Lakoff, George (1990) *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press.
- Langmore, Naomi E. (1998) "Functions of Duet and Solo Songs of Female Birds." *Trends in Ecology and Evolution* 13/4, pp. 136-140.
- Langmore, Naomi E. (2000) "Why Female Birds Sing." In *Animal Signals: Signalling and Signal Design in Animal Communication*. (Ed.) Yngve Espmark, Trond Amundsen, & Gunilla Rosenqvist. Trondheim: Tapir Academic Press, pp. 317-327.
- Lansky, Paul (2009) "Reflections on Spent Time." (Keynote address to the International Computer Music Conference, Montreal, Canada, August 18, 2009).
- Lappin, Shalom, Robert Levine, and David E. Johnson (2000) "The Structure of Unscientific Revolutions." *Natural Language & Linguistic Theory* 18/3, pp. 665-771.
- Larson, Steve (1997-98) "Musical Forces and Melodic Patterns." *Theory and Practice* 22-23, pp. 55-71.
- Larson, Steve (2004) "Musical Forces and Melodic Expectations: Comparing Computer Models and Experimental Results." *Music Perception* 21/4, pp. 457-498.
- Larson, Steve (2010) "Path and Purpose in a Raga." (Paper presented at the First International Conference on Analytical Approaches to World Music, Amherst).
- Larson, Steve (2012) *Musical Forces: Motion, Metaphor, and Meaning in Music*. Bloomington: Indiana University Press.
- Larson, Steve, and Leigh VanHandel (2005) "Measuring Musical Forces." *Music Perception* 23/2, pp. 119-136.
- Lasnik, Howard (2002) "The Minimalist Program in Syntax." *Trends in Cognitive Sciences* 6, pp. 432-437.
- Legate, Julie A., and Charles D. Yang (2002) "Empirical Re-Assessment of Stimulus Poverty Arguments." *The Linguistic Review* 19, pp. 151-162.

- LeGuin, Elisabeth (2006) *Boccherini's Body: An Essay in Carnal Musicology*. Berkeley & Los Angeles: University of California Press.
- Lendvai, Erno (1971) *Béla Bartók: An Analysis of His Music*. (Intro. Alan Bush) London: Kahn & Averill Publishers.
- Lerdahl, Fred (1992) "Cognitive Constraints on Compositional Systems." *Contemporary Music Review* 6/2, pp. 97-121.
- Lerdahl, Fred (2001) *Tonal Pitch Space*. New York: Oxford University Press.
- Lerdahl, Fred (2003a) "The Sounds of Poetry Viewed as Music." In *The Cognitive Neuroscience of Music*. (Ed.) Isabelle Peretz & Robert J. Zatorre. New York: Oxford University Press, pp. 413-429. (Originally published in *Annals of the New York Academy of Sciences* 930 (2001): 337-354).
- Lerdahl, Fred (2003b) "Two Ways in Which Music Relates to the World." *Music Theory Spectrum* 25/2, pp. 367-373.
- Lerdahl, Fred (2009) "Genesis and Architecture of the GTTM Project." *Music Perception* 26/3, pp. 187-194.
- Lerdahl, Fred, and Ray S. Jackendoff (1977) "Toward a Formal Theory of Tonal Music." *Journal of Music Theory* 21/1, pp. 111-171.
- Lerdahl, Fred, and Ray S. Jackendoff (1983) *A Generative Theory of Tonal Music*. Cambridge: MIT Press.
- Lerdahl, Fred, and Ray S. Jackendoff (1983-84) "An Overview of Hierarchical Structure in Music." *Music Perception* 1/2, pp. 229-252.
- Lerdahl, Fred, and Ray S. Jackendoff (1985) "A Reply to Peel and Slawson's Review Of "A Generative Theory of Tonal Music"." *Journal of Music Theory* 29/1, pp. 145-160.
- Lerdahl, Fred, and Carol L. Krumhansl (2007) "Modeling Tonal Tension." *Music Perception* 24/4, pp. 329-366.
- Lerner, Neil W. and Joseph N. Straus (2006) "Introduction: Theorizing Disability in Music." In *Sounding Off: Theorizing Disability in Music*. (Ed.) Neil W. Lerner & Joseph N. Straus. New York: Routledge, pp. 1-10.
- Levin, David J. (1999) *Richard Wagner, Fritz Lang, and the Nibelungen: The Dramaturgy of Disavowal*. Princeton: Princeton University Press.
- Levitin, Daniel J., and Vinod Menon (2003) "Musical Structure is Processed in 'Language' Areas of the Brain: A Possible Role for Brodmann Area 47 in Temporal Coherence." *NeuroImage* 20, pp. 2142-2152.
- Levy, Janet M. (1987) "Covert and Casual Values in Recent Writings About Music." *Journal of Musicology* 5/1, pp. 3-27.
- Lewin, David B. (1986) "Music Theory, Phenomenology, and Modes of Perception." *Music Perception* 3/4, pp. 327-392.

- Lewin, David B. (1987) *Generalized Musical Intervals and Transformations*. New Haven: Yale University Press.
- Lidov, David (1997) "Our Time with the Druids: What (and How) We Can Recuperate from Our Obsession with Segmental Hierarchies and Other "Tree Structures"." *Contemporary Music Review* 16/4, pp. 1-28.
- Livio, Mario (2008) *The Golden Ratio: The Story of Phi, the World's Most Astonishing Number*. New York: Broadway Books.
- London, Justin (2004) *Hearing in Time: Psychological Aspects of Musical Meter*. New York: Oxford University Press.
- London, Justin (2012) "Schemas, Not Syntax: A Reply to Patel." In *Language and Music as Cognitive Systems*. (Ed.) Patrick Rebuschat, Martin Rohrmeier, John Hawkins, & Ian Cross. New York: Oxford University Press, pp. 242-247.
- Mâche, Francois-Bernard (2001) "The Necessity of and Problems with a Universal Musicology." In *The Origins of Music*. (Ed.) Nils L. Wallin, Björn Merker & Steven Brown. Cambridge: MIT Press, pp. 473-480.
- MacLarnon, Ann M., and Gwen P. Hewitt (1999) "The Evolution of Human Speech: The Role of Enhanced Breathing Control." *American Journal of Physical Anthropology* 109/3, pp. 341-363.
- Maess, Burkhard, Stefan Koelsch, Thomas C. Gunter, and Angela D. Friederici (2001) "Musical Syntax is Processed in Broca's Area: An MEG Study." *Nature Neuroscience* 4/5, pp. 540-545.
- Marcus, Gary F., Ursula Brinkmann, Harald Clahsen, Richard Wiese, and Steven Pinker (1995) "German Inflection: The Exception That Proves the Rule." *Cognitive Psychology* 29/3, pp. 189-256.
- Marr, David (2010) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. (Foreword Shimon Ullman, Afterword Tomaso Poggio; originally published in 1982) Cambridge: MIT Press.
- Marsden, Alan A. (2005) "Generative Structural Representation of Tonal Music." *Journal of New Music Research* 34/4, pp. 409-428.
- Mavromatis, Panayotis, and Matthew G. Brown (2004) "Parsing Context-Free Grammars for Music: A Computational Model of Schenkerian Analysis." In *Proceedings of the 8th International Conference on Music Perception and Cognition*. pp. 414-415.
- McClary, Susan (1991) *Feminine Endings: Music, Gender, & Sexuality*. Minneapolis: University of Minnesota Press.
- McCloskey, James (1997) "Subjecthood and Subject Positions." In *Elements of Grammar*. (Ed.) Liliane Haegeman. Dordrecht: Kluwer Academic, pp. 197-235.
- McKee, Eric (2004a) "Dance and the Music of Chopin: The Waltz." In *The Age of Chopin: Interdisciplinary Inquiries*. (Ed.) Halina Goldberg. Bloomington: Indiana University Press, pp. 106-161.

- McKee, Eric (2004b) "Extended Anacruses in Mozart's Instrumental Music." *Theory and Practice* 29, pp. 1-37.
- McMullen, Erin, and Jenny R. Saffran (2004) "Music and Language: A Developmental Comparison." *Music Perception* 21, pp. 289-311.
- Meeus, Nicolas (2000) "Toward a Post-Schoenbergian Grammar of Tonal and Pre-Tonal Harmonic Progressions." *Music Theory Online* 6/1.
- Merriam, Alan (1964) *The Anthropology of Music*. Chicago: Northwestern University Press.
- Meyer, Leonard B. (1956) *Emotion and Meaning in Music*. Chicago: University of Chicago Press.
- Meyer, Leonard B. (1973) *Explaining Music: Essays and Explorations*. Chicago: University of Chicago Press.
- Meyer, Leonard B. (1976) "Grammatical Simplicity and Relational Richness: The Trio of Mozart's G Minor Symphony." *Critical Inquiry* 2/4, pp. 693-761.
- Meyer, Leonard B. (1996) *Style and Music: Theory, History, and Ideology*. Chicago: University of Chicago Press.
- Meyer, Leonard B. (1998) "A Universe of Universals." *Journal of Musicology* 16/1, pp. 3-25.
- Milewski, Barbara (1999) "Chopin's Mazurkas and the Myth of the Folk." *19th-Century Music* 23/2, pp. 113-135.
- Mill, John Stuart (1974) *The Collected Works of John Stuart Mill, Volume VII: A System of Logic Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation (Books I-III)*. (Ed. John M. Robson, Intro. Robert F. McRae) Toronto: University of Toronto Press.
- Miller, George A. (2003) "The Cognitive Revolution: A Historical Perspective." *Trends in Cognitive Sciences* 7/3, pp. 141-144.
- Molino, Jean (2001) "Toward an Evolutionary Theory of Language and Music." In *The Origins of Music*. (Ed.) Nils L. Wallin, Björn Merker & Steven Brown. Cambridge: MIT Press, pp. 165-176.
- Monelle, Raymond (2006) *The Musical Topic: Hunt, Military and Pastoral*. Bloomington: Indiana University Press.
- Monk, Ray (1991) *Ludwig Wittgenstein: The Duty of Genius*. New York: Penguin.
- Moore, Brian C. J. (2012) *An Introduction to the Psychology of Hearing (6th Ed.)*. London: Academic Press.
- Morgan, Robert P. (1978) "The Theory and Analysis of Tonal Rhythm." *Musical Quarterly* 64/4, pp. 435-473.
- Morris, Carrie, Julie A. Woodward, Byron F. Deen, and Michael J. Richard (2007) "Analysis of the Anatomic Changes of the Aging Face Using Computer-Assisted Tomography." (Paper presented at the

38th Annual Fall Scientific Symposium, American Society of Ophthalmic Plastic and Reconstructive Surgery, New Orleans).

Morris, Robert D. (2001) "Variation and Process in South Indian Music: Some 'Kritis' and Their 'Sangatis'." *Music Theory Spectrum* 23/1, pp. 74-89.

Morris, Robert D. (2006) "Architectonic Composition in South Indian Classical Music: The 'Navaragamalika Varnam'." In *Analytical Studies in World Music*. (Ed.) Michael S. Tenzer. Oxford: Oxford University Press, pp. 303-331.

Morton, Eugene S. (1996) "A Comparison of Vocal Behavior among Tropical and Temperate Passerine Birds." In *Ecology and Evolution of Acoustic Communication in Birds*. (Ed.) Donald E. Kroodsma & Edward H. Miller. Ithaca: Cornell University Press, pp. 258-268.

Mukherji, Nirmalangshu (2010) *The Primacy of Grammar*. Cambridge: MIT Press.

Naimpalli, Sadanand (2008) *Theory and Practice of Tabla*. Mumbai: Popular Prakashan.

Narmour, Eugene (1977) *Beyond Schenkerism: The Need for Alternatives in Musical Analysis*. Chicago: University of Chicago Press.

Narmour, Eugene (1983-84) "Some Major Theoretical Problems Concerning the Concept of Hierarchy in the Analysis of Tonal Music." *Music Perception* 1/3, pp. 129-199.

Narmour, Eugene (1990) *The Analysis and Cognition of Basic Melodic Structures*. Chicago: University of Chicago Press.

Nattiez, Jean-Jacques (1982) "Varese's 'Density 21.5': A Study in Semiological Analysis." (Trans. Anna Barry from the original work from 1975) *Music Analysis* 1/3, pp. 243-340.

Nattiez, Jean-Jacques (1983a) "The *Rekkukara* of the Ainu (Japan) and the *Katajjaq* of the Inuit (Canada): A Comparison." *The World of Music* 25/2, pp. 33-44.

Nattiez, Jean-Jacques (1983b) "Some Aspects of Inuit Vocal Games." *Ethnomusicology* 27/3, pp. 457-475.

Nattiez, Jean-Jacques (1989) "Reflections on the Development of Semiology in Music." (Trans. Katharine Ellis from "Reflexions sur le développement de la sémiologie musicale" (1988)) *Music Analysis* 8/1-2, pp. 21-75.

Nattiez, Jean-Jacques (1993) *Wagner Androgyne: A Study in Interpretation*. (Trans. Stewart Spencer from *Wagner androgyne: essai sur l'interprétation* (1990)). Princeton: Princeton University Press.

Nattiez, Jean-Jacques (1999) "Inuit Throat-Games and Siberian Throat Singing: A Comparative, Historical, and Semiological Approach." *Ethnomusicology* 43/3, pp. 399-418.

Neale, Stephen R. A. (1990) *Descriptions*. Cambridge: MIT Press.

Nettl, Bruno (2001) "An Ethnomusicologist Contemplates Universals in Musical Sound and Musical Culture." In *The Origins of Music*. (Ed.) Nils L. Wallin, Björn Merker & Steven Brown. Cambridge: MIT Press, pp. 463-472.

- Nevins, Andrew, David Pesetsky, and Celine Rodrigues (2009a) "Pirahã Exceptionality: A Reassessment." *Language* 85/2, pp. 355-404.
- Nevins, Andrew, David Pesetsky, and Celine Rodrigues (2009b) "Evidence and Argumentation: A Reply to Everett (2009)." *Language* 85/3, pp. 671-681.
- Nowak, Martin A. (2006) *Evolutionary Dynamics: Exploring the Equations of Life*. Cambridge: Harvard University Press.
- Nowak, Martin A., Joshua B. Plotkin, and Vincent A. A. Jansen (2000) "The Evolution of Syntactic Communication." *Nature* 404/6777, pp. 495-498.
- Palmer, Caroline, and Carol L. Krumhansl (1987) "Independent Temporal and Pitch Structures in Determination of Musical Phrases." *Journal of Experimental Psychology: Human Perception and Performance* 13/1, pp. 116-126.
- Parncutt, Richard (1989) *Harmony: A Psychoacoustical Approach*. Berlin: Verlag.
- Partee, Barbara H. (1975) "Comments on C.J. Fillmore's and N. Chomsky's Papers." In *The Scope of American Linguistics*. (Ed.) Robert Austerlitz. Lisse: Peter de Ridder Press, pp. 197-209.
- Partee, Barbara H. (1996) "The Development of Formal Semantics in Linguistic Theory." In *The Handbook of Contemporary Semantic Theory*. (Ed.) Shalom Lappin. Oxford: Blackwell, pp. 11-38.
- Partee, Barbara H. (2004) *Compositionality in Formal Semantics: Selected Papers by Barbara H. Partee*. Oxford: Blackwell Publishing.
- Pastille, William A. (1984) "Heinrich Schenker: Anti-Organicist." *19th-Century Music* 8/1, pp. 29-36.
- Pastille, William A. (1985) "Ursatz: The Musical Philosophy of Heinrich Schenker." (PhD Dissertation, Cornell University).
- Pastille, William A. (1990) "Music and Morphology: Goethe's Influence on Schenker's Thought." In *Schenker Studies*. (Ed.) Hedi Siegel. Cambridge: Cambridge University Press, pp. 29-44.
- Patel, Aniruddh D. (1998) "Syntactic Processing in Language and Music: Different Cognitive Operations, Similar Neural Resources?" *Music Perception* 16/1, pp. 27-42.
- Patel, Aniruddh D. (2003) "Language, Music, Syntax and the Brain." *Nature Neuroscience* 6/7, pp. 674-681.
- Patel, Aniruddh D. (2007) *Music, Language, and the Brain*. New York: Oxford University Press.
- Patel, Aniruddh D., Edward Gibson, Jennifer Ratner, Mireille Besson, and Phillip J. Holcomb (1998) "Processing Syntactic Relations in Language and Music: An Event-Related Potential Study." *Journal of Cognitive Neuroscience* 10/6, pp. 717-733.
- Patel, Aniruddh D., and Joseph R. Daniele (2003) "An Empirical Comparison of Rhythm in Language and Music." *Cognition* 87/1, pp. B35-45.
- Patel, Aniruddh D., John R. Iversen, Marlies Wassenaar, and Peter Hagoort (2008) "Musical Syntactic Processing in Agrammatic Broca's Aphasia." *Aphasiology* 22, pp. 776-789.

- Peretz, Isabelle (2006) "The Nature of Music from a Biological Perspective." *Cognition* 100/1, pp. 1-32.
- Peretz, Isabelle, and Max Coltheart (2003) "Modularity of Music Processing." *Nature Neuroscience* 6/7, pp. 688-691.
- Perlmutter, David M. (1992) "Sonority and Syllable Structure in American Sign Language." *Linguistic Inquiry* 23/3, pp. 407-442.
- Pesetsky, David M. (1982) "Paths and Categories." (PhD Dissertation, MIT).
- Pesetsky, David M. (2007) "Music Syntax is Language Syntax." (Unpublished paper).
- Pesetsky, David M. (2013) "Phrasal Movement and Its Discontents: Diseases and Diagnoses." In *Diagnosing Syntax*. (Ed.) Lisa Lai-Shen Cheng & Norbert Corver. New York: Oxford University Press, pp. 123-157.
- Piattelli-Palmarini, Massimo (2000) "The Metric of Open-Mindedness." *Natural Language & Linguistic Theory* 18/4, pp. 859-862.
- Pinker, Steven (1994) *The Language Instinct*. New York: William Morrow & Company.
- Pinker, Steven (1998) *How the Mind Works*. New York: W. W. Norton.
- Pinker, Steven (2002) *The Blank Slate: The Modern Denial of Human Nature*. London: Penguin.
- Pinker, Steven, and Ray S. Jackendoff (2005) "The Faculty of Language: What's Special About It?" *Cognition* 95, pp. 201-236.
- Piston, Walter (1978) *Harmony (4th Ed.)*. (Ed.) Mark DeVoto New York: W. W. Norton.
- Plunkett, Kim R. (1995) "Connectionist Approaches to Language Acquisition." In *Handbook of Child Language*. (Ed.) Paul Fletcher & Brian MacWhinney. Oxford: Blackwell, pp. 36-72.
- Plunkett, Kim R. (1997) "Theories of Early Language Acquisition." *Trends in Cognitive Sciences* 1/4, pp. 146-153.
- Plunkett, Kim R. and Ramin C. Nakisa (1997) "A Connectionist Model of the Arabic Plural System." *Language and Cognitive Processes* 12/5-6, pp. 807-836.
- Polanyi, Livia (2003) "The Linguistic Structure of Discourse." In *Handbook of Discourse Analysis*. (Ed.) Deborah Schiffrin, Deborah Tannen & Heidi E. Hamilton. Oxford: Blackwell, pp. 265-281.
- Pollard, Carl, and Ivan A. Sag (1994) *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Powers, Harold S. (1976) "The Structure of Musical Meaning: A View from Banaras (a Metamodal Model for Milton)." *Perspectives of New Music* 14/2, pp. 307-334.
- Powers, Harold S. (1980a) "India, Subcontinent Of." In *The New Grove Dictionary of Music and Musicians, Vol. 9*. (Ed.) Stanley Sadie. London: MacMillan, pp. 69-91.
- Powers, Harold S. (1980b) "Language Models and Musical Analysis." *Ethnomusicology* 24/1, pp. 1-60.

- Powers, Harold S. (1992) "Reinterpretations of Tradition in Hindustani Music: Omkarnath Thakur Contra Vishnu Narayan Bhattachande." In *The Traditional Indian Theory and Practice of Music and Dance*. (Ed.) Jonathan B. Katz. Leiden: Brill, pp. 9-52.
- Powers, Harold S. (1959) "The Background of the South Indian Rāga-System." (PhD Dissertation, Princeton University).
- Pressing, Jeff (1983) "Cognitive Isomorphisms between Pitch and Rhythm in World Musics: West Africa, the Balkans and Western Tonality." *Studies in Music* 17, pp. 38-61.
- Pullum, Geoffrey K., and Gerald J. M. Gazdar (1982) "Natural Languages and Context-Free Languages." *Linguistics and Philosophy* 4/4, pp. 471-504.
- Quinn, Ian and Panayotis Mavromatis (2011) "Voice-Leading Prototypes and Harmonic Function in Two Chorale Corpora." In *Mathematics and Computation in Music*. (Ed.) Carlos Agon, Andreatta Moreno, Gérard Assayag, Emmanuel Amiot, Jean Bresson & John Mandereau. Berlin: Springer, pp. 230-240.
- Quiring, Rebecca, Uwe Walldort, Urs Kloter, and Walter J. Gehring (1994) "Homology of the Eyeless Gene of Drosophila to the Small Eye Gene in Mice and Aniridia in Humans." *Science* 265/5173, pp. 785-789.
- Raffman, Diana (1993) *Language, Music and Mind*. Cambridge: MIT Press.
- Rahn, Jay (1983) *A Theory for All Music: Problems and Solutions in the Analysis of Non-Western Forms*. Toronto: University of Toronto Press.
- Rameau, Jean-Philippe (1971) *Treatise on Harmony*. (Intro. & Trans. Philip Gossett from *Traité de l'harmonie réduite à ses principes naturels* (1722)). New York: Dover.
- Ramus, Franck, Marc D. Hauser, Cory Miller, Dylan Morris, and Jacques Mehler (2000) "Language Discrimination by Human Newborns and by Cotton-Top Tamarin Monkeys." *Science* 288/5464, pp. 349-351.
- Ratner, Leonard G. (1985) *Classic Music: Expression, Form and Style*. New York: Schirmer.
- Robson, David (2013) "Are There Really 50 Eskimo Words for Snow?" *New Scientist* 2896, pp. 72-73.
- Rodgers, Stephen (2006) "Mental Illness and Musical Metaphor in the First Movement of Berlioz's *Symphonie Fantastique*." In *Sounding Off: Theorizing Disability in Music*. (Ed.) Neil W. Lerner & Joseph N. Straus. New York: Routledge, pp. 235-256.
- Roeper, Thomas, and Angeliek van Hout (2006) "The Representation of Movement in -ability Nominalizations: Evidence for Covert Category Movement, Edge Phenomena, and Local LF." (Paper presented at the Conference on Nominalizations and Quantification, Saarbrücken).
- Rohrmeier, Martin (2008) "A Generative Grammar Approach to Diatonic Harmonic Structure." In *Proceedings of the 4th Sound and Music Computing Conference*. (Ed.) Haralambos Spyridis, Anastasia Georgaki, Georgios Kouroupetroglou & Christina Anagnostopoulou. Lefkada: pp. 97-100.
- Rohrmeier, Martin (2011) "Towards a Generative Syntax of Tonal Harmony." *Journal of Mathematics and Music* 5/1, pp. 35-53.

- Rohrmeier, Martin, and Ian Cross (2008) "Statistical Properties of Harmony in Bach's Chorales." In *Proceedings of the 10th International Conference on Music Perception and Cognition*. (Ed.) Ken'ichi Miyazaki, Mayumi Adachi, Yuzuru Hiraga, Yoshitaka Nakajima & Minoru Tsuzaki. pp. 619-627.
- Rosen, Charles (1995) *The Romantic Generation*. Cambridge: Harvard University Press.
- Ross, John R. (1967) "Constraints on Variables in Syntax." (PhD Dissertation, MIT).
- Rothgeb, John E. (1968) "Harmonizing the Unfigured Bass: A Computational Study." (PhD Dissertation, Yale).
- Rothgeb, John E. (1975) "Strict Counterpoint and Tonal Theory." *Journal of Music Theory* 19/2, pp. 260-284.
- Rothgeb, John E. (1980) "Simulating Musical Skills by Digital Computer." *Computer Music Journal* 4/2, pp. 36-40.
- Rothstein, William N. (1989) *Phrase Rhythm in Tonal Music*. New York: Schirmer.
- Rothstein, William N. (1990a) "Rhythmic Displacement and Rhythmic Normalization." In *Trends in Schenkerian Research*. (Ed.) Allen Cadwallader. New York: Schirmer, pp. 87-113.
- Rothstein, William N. (1990b) "The Americanization of Heinrich Schenker." In *Schenker Studies*. (Ed.) Hedi Siegel. Cambridge: Cambridge University Press, pp. 193-203. (Originally published in *In Theory Only* 9/1 (1986): 5-17).
- Rothstein, William N. (1991) "On Implied Tones." *Music Analysis* 10/3, pp. 289-328.
- Rothstein, William N. (1995) "Beethoven 'with and without Kunstgepräg': Metrical Ambiguity Reconsidered." *Beethoven Forum* 4, pp. 165-193.
- Rothstein, William N. (2002) "Conservatory Schenker vs. University Schenker." *Tijdschrift voor Muziektheorie* 7/3, pp. 239-241.
- Rothstein, William N. (2004) "Playing with Forms: Mozart's Rondo in D Major, K. 485." In *Engaging Music: Essays in Music Analysis*. (Ed.) Deborah Stein. New York: Oxford University Press, pp. 202-214.
- Rumelhart, David E., Paul Smolensky, James L. McClelland, and Geoffrey E. Hinton (1986) "Schemata and Sequential Thought Processes in PDP Models." In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Vol. 2: Psychological and Biological Models*. (Ed.) James L. McClelland, David E. Rumelhart & The PDP Research Group. Cambridge: MIT Press, pp. 7-57.
- Rumelhart, David E., and James L. McClelland (1986) "PDP Models and General Issues in Cognitive Science." In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Vol. 1: Foundations*. (Ed.) James L. McClelland, David E. Rumelhart & The PDP Research Group. Cambridge: MIT Press, pp. 110-146.
- Russ, Michael (1993) "On Schenkerism: A Closed Circle of Elite Listeners?" *Music Analysis* 12/2, pp. 166-285.
- Russell, Bertrand (1905) "On Denoting." *Mind, New Series* 14/56, pp. 479-493.

- Russell, Bertrand (1912) *The Problems of Philosophy*. New York: Henry Holt.
- Ruwet, Nicholas (1987) "Methods of Analysis in Musicology." (Intro. & Trans. Mark Everist from "Méthodes d'analyse en musicologie" (1966)) *Music Analysis* 6/1-2, pp. 11-36.
- Saffran, Jenny R. (2002) "Constraints on Statistical Language Learning." *Journal of Memory and Language* 47, pp. 172–196.
- Saffran, Jenny R. (2003) "Musical Learning and Language Development." *Annals of the New York Academy of Sciences* 999, pp. 397-401.
- Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport (1996) "Statistical Learning by 8-Month-Old Infants." *Science* 274/5294, pp. 1926-1928.
- Saffran, Jenny R., Elizabeth K. Johnson, Richard N. Aslin, and Elissa L. Newport (1999) "Statistical Learning of Tone Sequences by Human Infants and Adults." *Cognition* 70/1, pp. 27-52.
- Saffran, Jenny R., Seth D. Pollak, Rebecca L. Seibel, and Anna Shkolnik (2007) "Dog is a Dog is a Dog: Infant Rule Learning is Not Specific to Language." *Cognition* 105, pp. 669-680.
- Salzer, Felix (1962) *Structural Hearing: Tonal Coherence in Music*. New York: Dover.
- Salzer, Felix, and Carl Schachter (1969) *Counterpoint in Composition*. New York: McGraw-Hill.
- Samarotto, Frank (1999) "Strange Dimensions: Regularity and Irregularity in Deep Levels of Rhythmic Reduction." In *Schenker Studies* 2. (Ed.) Carl Schachter & Hedi Siegel. Cambridge: Cambridge University Press, pp. 222-238.
- Samarotto, Frank (2007) "Thick Formalism: Structure as Meaning in Brahms's Im Herbst, Op. 104 #5." (Paper presented at the Annual Meeting of the Society for Music Theory, Baltimore).
- Santorini, Beatrice, and Anthony Kroch (2007) "The Syntax of Natural Language: An Online Introduction Using the Trees Program " Electronic Book (available at <http://www.ling.upenn.edu/~beatrice/syntax-textbook/>)
- Schachner, Adena, Timothy F. Brady, Irene M. Pepperberg, and Marc D. Hauser (2009) "Spontaneous Motor Entrainment to Music in Multiple Vocal-Mimicking Species." *Current Biology* 19, pp. 831-836.
- Schachter, Carl (1990) "Either/Or." In *Schenker Studies*. (Ed.) Hedi Siegel. Cambridge: Cambridge University Press, pp. 165-179.
- Schachter, Carl (1999a) "Rhythm and Linear Analysis I: A Preliminary Study." In *Unfoldings: Essays in Schenkerian Theory and Analysis*. (Ed.) Joseph N. Straus. New York: Oxford University Press, pp. 17-53. (Originally published in *Music Forum* 4 (1976): 281-334).
- Schachter, Carl (1999b) "Rhythm and Linear Analysis II: Durational Reduction." In *Unfoldings: Essays in Schenkerian Theory and Analysis*. (Ed.) Joseph N. Straus. New York: Oxford University Press, pp. 54-78. (Originally published in *Music Forum* 5 (1980): 197-232).
- Schachter, Carl (1999c) "Rhythm and Linear Analysis III: Aspects of Meter." In *Unfoldings: Essays in Schenkerian Theory and Analysis*. (Ed.) Joseph N. Straus. New York: Oxford University Press, pp. 79-117. (Originally published in *Music Forum* 6 (1987): 1-59).

- Schachter, Carl (1999d) "Structure as Foreground: 'Das Drama Des Ursatzes'." In *Schenker Studies* 2. (Ed.) Carl Schachter & Hedi Siegel. Cambridge: Cambridge University Press, pp. 298-314.
- Schachter, Carl (2005) *Sound and Time: Analysis for the Performer of Tonal Music*. New York: Oxford University Press.
- Schenker, Heinrich (1969) *Five Graphic Analyses. With an Introduction and Glossary*. (Trans. Felix Salzer from *Funf Urlinie-Tafeln* (1932)). New York: Dover.
- Schenker, Heinrich (1973) *Harmony*. (Ed. Oswald Jonas, Trans. Elisabeth M. Borgese from *Neue musikalische Theorien und Phantasien Vol 1: Harmonielehre* (1906)). Cambridge: MIT Press.
- Schenker, Heinrich (1979) *Free Composition*. (Intro. Allen Forte, Trans. & Ed. Ernst Oster from *Neue musikalische Theorien und Phantasien Vol 3: Der Freie Satz* (1935)). New York: Longman.
- Schenker, Heinrich (1987) *Counterpoint*. 2 Vols. (Trans. John Rothgeb & Jürgen Thym from *Neue musikalische Theorien und Phantasien Vol 2: Kontrapunkt* (1910/1922)). New York: Schirmer.
- Schenker, Heinrich (1988) "The Spirit of Musical Technique." (Trans. William A. Pastille from "Der Geist der musikalischen Technik" (1895)) *Theoria* 3, pp. 86-104.
- Schenker, Heinrich (1994) *The Masterwork in Music: A Yearbook Vol. 1*. (Trans. & Ed. William Drabkin, Trans. Ian Bent, Richard A. Kramer, John Rothgeb & Hedi Siegel from *Das Meisterwerk in der Musik: Ein Jahrbuch Vol. 1* (1925)). Cambridge: Cambridge University Press.
- Schenker, Heinrich (1996) *The Masterwork in Music: A Yearbook Vol. 2*. (Trans. & Ed. William Drabkin, Trans. Ian Bent, John Rothgeb & Hedi Siegel from *Das Meisterwerk in der Musik: Ein Jahrbuch Vol. 2* (1926)). Cambridge: Cambridge University Press.
- Schenker, Heinrich (2004) *Der Tonwille: Pamphlets in Witness of the Immutable Laws of Music Vol. 1*. (Trans. & Ed. William Drabkin, Trans. Ian D. Bent, William Drabkin, Joseph Dubiel, Timothy L. Jackson, R. Joseph Lubben & Richard Snarrenberg from *Der Tonwille: Flugblätter zum Zeugnis unwandelbarer Gesetze der Tonkunst Vol. 1* (1921-23)). New York: Oxford University Press.
- Schenker, Heinrich (2005) *Der Tonwille: Pamphlets in Witness of the Immutable Laws of Music Vol. 2*. (Trans. & Ed. William Drabkin, Trans. Ian D. Bent, William Drabkin, Joseph Dubiel, R. Joseph Lubben, William Renwick & Richard Snarrenberg from *Der Tonwille: Flugblätter zum Zeugnis unwandelbarer Gesetze der Tonkunst Vol. 2* (1923-24)). New York: Oxford University Press.
- Schmalfeldt, Janet (1991) "Towards a Reconciliation of Schenkerian Concepts with Traditional and Recent Theories of Form." *Music Analysis* 10/3, pp. 233-287.
- Schmalfeldt, Janet (2011) *In the Process of Becoming: Analytical and Philosophical Perspectives on Form in Early Nineteenth-Century Music*. Oxford: Oxford University Press.
- Schoenberg, Arnold (1967) *Fundamentals of Musical Composition*. (Ed.) Gerald Strang, Intro. Leonard Stein. New York: St. Martins Press.
- Scholz, Barbara C., and Geoffrey K. Pullum (2006) "Irrational Nativist Exuberance." In *Contemporary Debates in Cognitive Science*. (Ed.) Robert J. Stainton. Malden: Blackwell, pp. 59-80.

- Scruton, Roger V. (1987) "Analytical Philosophy and the Meaning of Music." *Journal of Aesthetics and Art Criticism* 46, pp. 169-176.
- Scruton, Roger V. (1999) "Reply to Hamilton." *Philosophical Books* 40/3, pp. 155-159.
- Searle, John R. (1997) "Minds, Brains and Programs." In *Mind Design II*. (Ed.) John Haugeland. Cambridge: MIT Press, pp. 183-204. (Originally published in *Behavioral and Brain Sciences* 3/3 (1980): 417-457).
- Sloboda, John A. (1985) *The Musical Mind: The Cognitive Psychology of Music*. London: Oxford University Press.
- Sloboda, John A., and Patrik N. Juslin (2001) "Psychological Perspectives on Music and Emotion." In *Music and Emotion: Theory and Research*. (Ed.) Patrik N. Juslin & John A. Sloboda. New York: Oxford University Press, pp. 71-104.
- Smoliar, Stephen J. (1980) "A Computer Aid for Schenkerian Analysis." *Computer Music Journal* 4/2, pp. 41-59.
- Snell, James L. (1979) "Design for a Formal System for Deriving Tonal Music." (Master's Thesis, State University of New York, Binghamton).
- Snell, James L. (1983) "Computerized Hierarchical Generation of Tonal Compositions." In *Informatique Et Musique*. (Ed.) Hélene Charnassé. Ivry: Elmeratto, pp. 197.
- Solie, Ruth A. (1980) "The Living Work: Organicism and Musical Analysis." *19th-Century Music* 4/2, pp. 147-156.
- Solomon, Maynard (2004) "The Violin Sonata in G Major, Op. 96: Pastoral, Rhetoric, Structure." In *The Beethoven Violin Sonatas: History, Criticism, Performance*. (Ed.) Lewis Lockwood & Mark Kroll. Urbana: University of Illinois Press, pp. 110-128.
- Somfai, László (1996) *Béla Bartók: Composition, Concepts, and Autograph Sources*. Berkeley & Los Angeles: University of California Press
- Spitzer, Michael (2004) *Metaphor and Musical Thought*. Chicago: University of Chicago Press.
- Spitzer, Michael (2008) "A Metaphoric Model of Sonata Form: Two Expositions by Mozart." In *Communication in Eighteenth-Century Music*. (Ed.) Danuta Mirka & Kofi Agawu. Cambridge: Cambridge University Press, pp. 189-229.
- Stabler, Edward (2009) "Computational Models of Language Universals." In *Language Universals*. (Ed.) Morten H. Christiansen, Christopher Collins & Shimon Edelman. Oxford: Oxford University Press, pp. 200-223.
- Stabler, Edward (2011) "Computational Perspectives on Minimalism." In *Oxford Handbook of Linguistic Minimalism*. (Ed.) Cedric Boeckx. New York: Oxford University Press, pp. 617-642.
- Straus, Joseph N. (1987) "The Problem of Prolongation in Post-Tonal Music." *Journal of Music Theory* 31/1, pp. 1-21.

- Straus, Joseph N. (2006) "Inversional Balance and the 'Normal' Body in the Music of Arnold Schoenberg and Anton Webern." In *Sounding Off: Theorizing Disability in Music*. (Ed.) Neil W. Lerner & Joseph N. Straus. New York: Routledge, pp. 257-268.
- Strawson, Peter F. (2004) "On Referring." In *Logico-Linguistic Papers (2nd Ed.)*. Aldershot: Ashgate, pp. 1-20. (Originally published in *Mind, New Series* 59/235 (1950): 320-344).
- Sundberg, Johan and Björn Lindblom (1991) "Generative Theories for Describing Musical Structure." In *Representing Musical Structure*. (Ed.) Peter Howell, Robert West & Ian Cross. London and San Diego: Academic Press, pp. 245-272.
- Swartz, Anne (1983) "Folk Dance Elements in Chopin's Mazurkas." *Journal of Musicological Research* 4/3-4, pp. 417-425.
- Taruskin, Richard (2009) *The Oxford History of Western Music: Volume 5 (Late Twentieth Century)*. New York: Oxford University Press.
- Temperley, David (2001) *The Cognition of Basic Musical Structures*. Cambridge: MIT Press.
- Temperley, David (2003) "End-Accented Phrases: An Analytical Exploration." *Journal of Music Theory* 47/1, pp. 125-154.
- Temperley, David (2004a) "Bayesian Models of Musical Structure and Cognition." *Musicae Scientiae* 8/2, pp. 175-205.
- Temperley, David (2004b) "Communicative Pressure and the Evolution of Musical Styles." *Music Perception* 21/3, pp. 313-337.
- Temperley, David (2007) *Music and Probability*. Cambridge: MIT Press.
- Temperley, David (2009) "A Unified Probabilistic Model for Polyphonic Music Analysis." *Journal of New Music Research* 38/1, pp. 3-18.
- Temperley, David (2011) "Composition, Perception, and Schenkerian Theory." *Music Theory Spectrum* 33/2, pp. 146-168.
- Temperley, David, and Christopher Bartlette (2002) "Parallelism as a Factor in Metrical Analysis." *Music Perception* 20/2, pp. 117-149.
- Thakur, Omkarnath (1938-62) *Sangītānjalī*. Hathras: Sangeet Karyalaya.
- Thompson, D'Arcy W. (1917) *On Growth and Form*. Cambridge: Cambridge University Press.
- Tomlin, Russell S. (1986) *Basic Word Order: Functional Principles*. Beckenham: Croom Helm.
- Trainor, Laurel J. (2005) "Are There Critical Periods for Musical Development?" *Developmental Psychobiology* 46/3, pp. 262-278.
- Travis, John (1994) "Wiring the Nervous System." *Science* 266/5185 pp. 568-570.
- Travis, Lisa deMena (1984) "Parameters and Effects of Word Order Derivation." (PhD Dissertation, MIT).

- Trehub, Sandra E. (2001) "Human Processing Predispositions and Musical Universals." In *The Origins of Music*. (Ed.) Nils L. Wallin, Björn Merker & Steven Brown. Cambridge: MIT Press, pp. 427-448.
- Trehub, Sandra E. (2003) "The Developmental Origins of Musicality." *Nature Neuroscience* 6/7, pp. 669-673.
- Trehub, Sandra E., Anna M. Unyk, and Laurel J. Trainor (1993) "Adults Identify Infant-Directed Music across Cultures." *Infant Behavior and Development* 16/2, pp. 193-211.
- Trehub, Sandra E., and Erin E. Hannon (2006) "Infant Music Perception: Domain-General or Domain-Specific Mechanisms?" *Cognition* 100/1, pp. 73-99.
- Turing, Alan M. (1936) "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society Series 2* 42, pp. 230-265.
- Turing, Alan M. (1950) "Computing Machinery and Intelligence." *Mind, New Series* 59/236, pp. 433-460.
- Turing, Alan M. (1952) "The Chemical Basis of Morphogenesis." *Philosophical Transactions of the Royal Society B: Biological Sciences* 237/641, pp. 37-72.
- Tymoczko, Dmitri (2003) "Function Theories: A Statistical Approach." *Musurgia* 10/3-4, pp. 35-64.
- Tymoczko, Dmitri (2006) "The Geometry of Musical Chords." *Science* 313/5783, pp. 72-74.
- Tymoczko, Dmitri (2011) *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. New York: Oxford University Press.
- Uriagereka, Juan (1998) *Rhyme and Reason: An Introduction to Minimalist Syntax*. Cambridge: MIT Press.
- Uriagereka, Juan (2000) "On the Emptiness of 'Design' Polemics." *Natural Language & Linguistic Theory* 18/4, pp. 863-871.
- Van den Toorn, Pieter (1995) *Music, Politics and the Academy*. University of California Press:
- Van Dijk, Teun A. (1972) *Some Aspects of Text Grammars: A Study in Theoretical Poetics and Linguistics*. The Hague: Mouton.
- Van Dijk, Teun A. (2003) "Critical Discourse Analysis." In *Handbook of Discourse Analysis*. (Ed.) Deborah Schiffrin, Deborah Tannen & Heidi E. Hamilton. Oxford: Blackwell, pp. 352-371.
- van Heijningen, Caroline A. A., Jos de Vissera, Willem Zuidemaa, and Carel ten Catea (2009) "Simple Rules Can Explain Discrimination of Putative Recursive Syntactic Structures by a Songbird Species." *Proceedings of the National Academy of Sciences* 106/48 pp. 20538-20543.
- Wade, Bonnie C. (1979) *Music in India: The Classical Traditions*. Englewood Cliffs: Prentice-Hall.
- Webster, James (1991) "The Analysis of Mozart's Arias." In *Mozart Studies*. (Ed.) Cliff Eisen. Cambridge: Cambridge University Press, pp. 101-199.
- Westergaard, Peter (1975) *An Introduction to Tonal Theory*. New York: W. W. Norton.

- Whittall, Arnold (1987) "Analysis as Performance." In *Atti Del XIV Congresso Della Societa Internazionale di Musicologia*. (Ed.) Angelo Pompilio, Donatella Restani, Lorenzo Bianconi, & F. Alberto Gallo. Bologna: International Musicological Society, pp. 654-659.
- Wiesel, Torsten N., and David H. Hubel (1963) "Single-Cell Responses in Striate Cortex of Kittens Deprived of Vision in One Eye." *Journal of Neurophysiology* 26, pp. 1003-1017.
- Wigner, Eugene P. (1960) "The Unreasonable Effectiveness of Mathematics in the Natural Sciences." *Communications in Pure and Applied Mathematics* 13/1, pp. 1-14.
- Willner, Channan (1999) "Sequential Expansion and Handelian Phrase Rhythm." In *Schenker Studies 2*. (Ed.) Carl Schachter & Hedi Siegel. Cambridge: Cambridge University Press, pp. 192-221.
- Wilson, Jeffrey A., and Matthew T. Carrano (1999) "Titanosaurs and the Origin of "Wide-Gauge" Trackways: A Biomechanical and Systematic Perspective on Sauropod Locomotion." *Paleobiology* 25/2, pp. 252-267.
- Winograd, Terry (1968) "Linguistics and the Computer Analysis of Tonal Harmony." *Journal of Music Theory* 12/1, pp. 2-49.
- Wise, Christopher (2011) *Chomsky and Deconstruction: The Politics of Unconscious Knowledge*. New York: Macmillan.
- Wittgenstein, Ludwig (1958) *The Blue and Brown Books*. Oxford: Blackwell.
- Wittgenstein, Ludwig (1998) *Culture and Value*. (Trans. Peter Winch from *Vermischte Bemerkungen* (1977)). Oxford: Blackwell.
- Wright, Anthony A., Jacquelyne J. Rivera, Stewart H. Hulse, Melissa Shyan, and Julie J. Neiworth (2000) "Music Perception and Octave Generalization in Rhesus Monkeys." *Journal of Experimental Psychology: General* 129/3, pp. 291-307.
- Zadeh, Chloe (2012) "Formulas and the Building Blocks of *Thumri* Style: A Study In "Improvised" Music " *Analytical Approaches to World Music* 12/1.
- Zagona, Karen (1982) "Government and Proper Government of Verbal Projections." (PhD Dissertation, University of Washington, Seattle).
- Zarlino, Gioseffo (1968) *The Art of Counterpoint: Part Three of Le Institutioni Harmoniche, 1558*. (Partial Trans. Guy A. Marco & Claude V. Palisca from Book 3 of *Le institutioni harmoniche* (1558)). New Haven: Yale University Press.
- Zbikowski, Lawrence M. (2002) *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*. New York: Oxford University Press.
- Zbikowski, Lawrence M. (2008) "Dance Topoi, Sonic Analogues and Musical Grammar: Communicating with Music in the Eighteenth Century." In *Communication in Eighteenth-Century Music*. (Ed.) Danuta Mirka & Kofi Agawu. Cambridge: Cambridge University Press, pp. 283-309.
- Zhang, Juzhong, Garman Harbottle, Changsui Wang, and Zhaochen Kong (1999) "Oldest Playable Musical Instruments Found at Jiahu Early Neolithic Site in China." *Nature* 401/6751, pp. 366-368.

Zuckerkandl, Victor (1971) *The Sense of Music (Rev. Ed.)*. Princeton: Princeton University Press.