Tobias Scheer
Chapter for the Continuum Companion to Phonology
Revised **and unabridged** version, November 17th, 2009

Generative phonology and its evolution

1. Introduction

This chapter is designed to provide an overview of the evolution of generative phonology since its inception in the 60s. It is of course hopeless to try to cover all aspects of the subject matter in a book chapter. There is a growing body of specialized and well-documented literature on the history of (generative) linguistics and phonology that inquires on individual pieces of the general picture and sometimes covers larger periods.[1] Given this situation, below I take the opportunity to concentrate on a small number of what I consider to be central issues and salient steps of the evolution. I assume that the reader is familiar with the general background and the broad properties of the theories mentioned: rather than introduced, these will be located with respect to the different evolutional strands, and with respect to each other. The following pages thus propose cherry-picking on the grounds of a personal choice, rather than historiography (that ambitions at exhaustivity). There are thus a number of theories, issues and sub-fields that the reader will not come across, or which will only be mentioned in passing. Among these are social variation (generative phonology and William Labov), dialectal variation, markedness, diachronics, learnability, laboratory phonology, experimental phonology, subsegmental structure (melodic primes) or questions regarding "substance", i.e. the relationship with phonetics in general and the evolution of the typically generative notion of *systematic phonetics* in particular.

Rather than being historiographic, the ambition thus follows the traditional motivation of historians: to learn from the past in order to understand the present, and to shape the future. Namely, knowing about the past is a good vaccination against multiple reinventions of the wheel. We will see that real progress of ideas notwithstanding (autosegmental representations, interactionism, parallel computation), generative phonology is quite immature a science in this respect: major debates of the past fade away into amnesia a generation later, and the field moves in circles without noticing the footsteps that it walks in, sometimes putting a lot of energy and rhetoric into denying the circular movement. The attitude towards the balance of structure and process is a case in point that is examined below.

The kind of history that is told is thus not neutral or impartial: it is goal-oriented and functional in the sense that it aims at isolating important issues, key questions, central ideas, circular movements and also real scientific progress. Admitting that one does not look at facts like a robot, that one does not try to marshal oneself down to strict neutrality, is a description of reality, rather than a confession: there is no such thing as unoriented and purely factual historiography; those who pretend that history can be told from a strictly neutral vantage point merely try to achieve a rhetorical advantage that knights their own partial view on the facts with the promise of objectivity.[2]

The question is not whether history-telling is partial and oriented or not; the only thing that it is worth bothering is the degree of partiality and orientedness. Also, the best history is not necessarily the one that is written with the least degree of partiality: Michelet's history of the French Revolution is anything but impartial, but still invaluable today. History is not a

---

[1] See for example Newmeyer (1986, 1996), Anderson (1985), Encrevé (1997), Goldsmith (2005), Goldsmith & Laks (2000, forth), Koerner (2002c, 2004a), $Hulst, 2004 #3636£ (2004), Laks (2001) and the five contributions to Auroux et al. (2006) regarding generative phonology by J. Durand, B. Laks, J. Goldsmith and M. Noske, J. Jensen and J. Lowenstamm.

[2] See Koerner (2002a:154f) for a different position that promotes a kind of enlightened positivism ("broad positivism" in Koerner's words), which is close to pure historiography (i.e. tries to show "what really happened"), declines any ambition to explain the present and to act on the future, but admits that there is no absolute objectivity, and that somebody must make sense of the "facts".

concentration of unrelated and uninterpreted facts; it is only history when it makes sense, and sense can only be made by the historian from hindsight.

Calling on the historical experience of adult, mature or "successful" sciences, Noam Chomsky unswervingly repeats since the 60s that science is after insight, rather than after methodological correctness, impartiality or other formal and secondary virtues.

Finally, a word is in order to make explicit that the pages below only do "internal" history, that is history based on published (or unpublished, but written) scholarly work. "External" history is not solicited: who was friends with whom, who broke up with whom because of an unpleasant review, who was the teacher of whom and programmed his pupils to think this or that way, who was married with whom, who studied where, who has which character sketch in the opinion of whom, who wrote which letter or e-mail to whom and so on is a kind of information that may well contribute to the understanding of the overall picture, but which must be discussed elsewhere.

2. The philosophical foundations of generative grammar

2.1. Cognitive Science, modules

Generative grammar is an actor and a result of the so-called cognitive revolution of the 50s and 60s which established a new episteme in a number of classical domains of knowledge, and gave rise to a new field, Cognitive Science. Based on work of the 30s and 40s by Alan Turing, John von Neumann, Norbert Wiener, Warren McCulloch and others, as well as on the (early) 19th century thinking known as phrenology (Franz Josef Gall, 1758-1828), the cognitive revolution considered the mind/brain[3] as a set of specialized and interconnected input-output devices that carry out computation according to a pre-specified programme. Computation involves a set of ordered instructions (serialism), and its input are units that belong to a specialized (domain-specific) vocabulary (which cannot be interpreted by other computational systems) and is retrieved from a long-term storage device. Computation itself is carried out in what psychologists call short-term or workbench memory.

The computational devices at hand were later called modules, and the theory of the modularity of the mind/brain was formally articulated by Fodor (1983). The cognitive revolution was a joint venture of philosophers, psychologists, anthropologists, linguists, neuro-scientists and people working on artificial intelligence, to name but the areas where its impact was most salient (Gardner 1985 provides an informed survey). Applications in the real world include early cybernetics and robotics, and the most spectacular result today are modern computers, which work on the basis of modularity and serialism.

The designated enemy of the cognitive revolution was behaviourism, which dominated American psychology since the early 20th century. The basic tenet of this theory is that mental/brainal activity cannot be accessed or measured: something is certainly going on, but humans do not have access to relevant data. Since science is about measuring real data, it is non-scientific to speculate on what happens in the mind/brain, which is therefore treated as a black box. The only thing that can be scientifically measured is the input to this black box (stimulus) and its output, i.e. human behaviour (response).

The cognitive revolution mocks the behaviourists' prohibition to inquire on how the mind/brain works: excluding a natural object from the consideration of human curiosity is the worst that one can do to scientific endeavour. Therefore everybody who subscribes to the

---

[3] Here and below, this vocabulary item is only used for the sake of exposition: it refers to either the mind or the brain in cases where the distinction is not relevant, but where the reduction to either would be wrong. Mind/brain is used in Chomsky's (recent) writings, where it is specifically related to what he calls the language organ. This meaning is *not* what is intended here (it would be anachronistic in the context of the 50s-60s anyway).

premises of the cognitive revolution accepts that his object of study is physically located in the brain. Also, his theories are liable for complying with a general theory of the mind/brain: whatever the basic principles that are found to organize mental/brainal activity, they will have to be obeyed by all theories of individual cognitive functions across disciplines. For example, a linguistic theory that requires unlimited storage capacities can be rejected out of hand qua cognitive realism: the brain has certainly an incredibly large storage capacity, which however is finite.

Cognitive Science slowly emerged in the 60s and 70s. Behind its confrontation with behaviourism is a deeply rooted cultural break that divides philosophy since ancient times: empiricism/ monism vs. mentalism/ rationalism/ dualism. The representative of the former world view in Antiquity is held to be Aristotle, and its modern defenders are typically anglo-saxon: like behaviourists, John Locke (1632-1704) and David Hume (1711-1776) were arguing against "speculative philosophers" (such as René Descartes, 1596-1650) who are not basing their statements on data and the material world but talk about things that are out of reach for serious scientific inquiry. According to Locke, humans are like a blank sheet when they are born, and nothing can exist in the mind that was not created by empirical sensory-motor experience. This is the classical materialist position of enlightenment that opposes metaphysical thinking as much as Descartes' ideas: both are based on non-material, non-substantive grounds, if in a different way.

Positivism, another incarnation of empiricism, holds that the only valid methodology in science is inductive (bottom-up): data need to be accumulated and at some point of density produce insight without any element intervening that is not based on them. In other words, experience, observation and sensory input is the only source of knowledge: data-independent hypothesis do not play any role in scientific endeavour. Therefore, positivism opposed systematically "speculations" about things that we cannot see: in the second half of the 19th century Ernst Mach (1838-1916) for example rejected atoms as pure speculation because they were too small to be observed. Leading in this conception of science was the so-called Vienna Circle that developed logical positivism in the 20s and 30s of the 20th century around Rudolf Carnap (Auguste Comte, 1798-1857, is usually credited as the father of positivism).

2.2. Cartesian Linguistics: rationalism, anti-empiricism

The alternative world view is represented by Plato in Antiquity, and by philosophers such as René Descartes, Immanuel Kant (1724-1804), Wilhelm von Humboldt (1767-1835) and Karl Popper (1902-1994) in modern times. Descartes considered that the mind and the body are different entities (cogito ergo sum), an insight that expresses the Western dualistic view of the human being. Based on the recognition of the mind as an independent entity, rationalism holds that reason, rather than mere observation and sensory input, is the relevant source of knowledge and defines what is true.

Another central question for Descartes was whether humans are just some kind of highly sophisticated machine. His answer was no; sure human action is guided by all kinds of external influences and empirical impressions, but these alone cannot explain human behaviour: empirical data and sensory input cannot make sense and remain unparsable in absence of a validating and structuring device that exists independently of them. In science, insight is the result of intertwined (hence dualistic) bottom-up observation and top-down hypotheses, whereby the latter can have any origin, including one that has no sensory, empirical or experimental basis (Einstein said that "the more primitive the status of a science is, the more readily can the scientist live under the illusion that he is a pure empiricist").

Noam Chomsky and generative linguistics root in the Cognitive Science programme of the 50s and the rationalist/ dualist tradition. Nothing could be called generative that lies beyond these benchmarks.

This is true despite the fact that in his earliest writings Chomsky was influenced by Bloomfield's anti-mentalism (on which more below in section 3.2) and, on the philosophical side, by Willard Quine and Nelson Goodman, two continuators of the Vienna Circle. In his Logical Structure of Linguistic Theory, Chomsky defends Bloomfield's anti-mentalism with explicit reference to Quine:

> "It is true that many philosophers have given up, as a general requirement for significance, the kind f reductionism that our restatement of Bloomfield's program has as its goal […]. [I]t seems to me that it [the construction of a new notion of significance] will rule out mentalism for what were essentially Bloomfield's reasons, i.e., its obscurity and inherent untestability. Thus Quine rejects reductionism, suggests an alternative, and rejects mentalism" $Chomsky, 1955-56 #3244£ (1955-56:86).

It is true that the generative enterprise, if reduced to the word "generative", looks much like a computational venture with an empirical goal: a set is generated if we can provide a way to enumerate all its members ($Post, 1943 #3634£ 1943). LSLT was written in 1955-56 but never published until 1975 (except for chapter nine, which was outsourced as Chomsky's Ph.D dissertation). In the introduction to the 1975 publication, Chomsky insists on the rationalist roots of the generative enterprise, and on its fundamental opposition to empiricist thinking (Chomsky 1955-56: 12f).

There is reason to believe that what made Chomsky join the rationalist camp is the problem of language acquisition: this is what he builds on in the 1975 introduction to LSLT, and this is also the bone of contention of his review of $Skinner, 1957 #2247£'s book "Verbal Behaviour" (Chomsky 1959). It is this review and its radical confrontation with empiricist behaviourism that that is usually taken to be the initial spark of the generative enterprise. This is probably true as far as its philosophical foundations are concerned: the rational and dualist (as opposed to empiricist and monist) character of the generative project is made explicit in Chomsky's philosophical writings since the 60s and up to the present day in unremitting verbal effort (one important point of crystallization being $Chomsky, 1966 #1449£ 1966, and also $Chomsky, 1972 #2851£ 1972). The most visible consequences in linguistic theory are Universal Grammar, the inverted T model (on which more below) and the dualist distinction between competence and performance.

While Skinner held that language was but a specific type of human behaviour that is learned like one learns to play the piano or to drive a car, Chomsky argued that language is different in kind: it is a natural object, rather than an artefact. Therefore its acquisition mechanism is partly hard-wired and pre-specified upon birth in the same way as other aspects of human development whose genetic transmission is beyond doubt: the way young humans acquire their native tongue patterns with the way they come to grips with eating and walking: they do need repetition and experience, but the success is always complete (there is no point in talking about individuals that are good or bad at eating, walking or speaking), achieved without effort (nobody has ever complained about the task of learning to eat, to walk or to speak) and entirely independent of the motivation and intelligence of the individual. Also, humans cannot decide not to acquire a language (or to forget it for that matter), just as they cannot decide not to learn to walk or to eat. All these properties are the reverse of what is found when humans learn to play the piano, to drive a car or to try to get command over some other artefactual system created by man: they are more or less good at it, they sweat a lot during the learning process, the results depend on their motivation and eventually on their intelligence, and if they decide not to learn they won't.

Chomsky concludes that the cognitive processes which allow humans to learn artefactual systems and to acquire language are not the same. Just as the faculty to eat and to walk, the faculty of language, i.e. to acquire a native tongue, is genetically encoded (later on this biological conception of language will be known under the header of the language organ (see $Pinker, 1994 #1878£ 1994 and $Anderson, 2002 #3012£ 2002 for an overview).

The idea that the language faculty is genetically encoded also follows from the modular conception of Cognitive Science: if like other lower cognitive functions (such as audition and vision) language is a module in the sense of $Fodor, 1983 #2774£ (1983), it must be genetically programmed like all other modules.

On a number of occasions below, the discussion will draw on these foundational properties of the generative enterprise, namely in the face of persistent empiricist raids on generative phonology in the past decade (usage-based, statistical learning etc., e.g. Carr 2000, Bybee 2001), and also of the fact that mainstream generative phonology, Optimality Theory, is slowly leaving the generative realm without this arousing any significant debate.

3. Generative and structuralist phonology: so very different after all?

3.1. Tabula rasa following the death of the phoneme

In the collective memory of generativists, the key concept of structuralism, the phoneme, was successfully and irrecoverably refuted by Morris Halle's conference at the 1957 LSA meeting (the written version appeared in Halle 1959:22ff). The argument was about the uniform behaviour of sounds irrespectively of their phonemic status. In Russian, voiceless consonants are voiced before voiced obstruents irrespectively of whether the resulting voiced item is a phoneme or not. That is, [t͡s, t͡ʃ, x] have no voiced partner in the phonemic inventory of Russian, but undergo voicing exactly in the same way as those phonemes that participate in the correlation of voicing. Therefore, the argument went, phonology does not care for the phonemic/distinctive status of the objects that are manipulated. Writing separate statements for paired and unpaired sounds as would have been required by structuralist methodology would miss the generalization and impede insight.

Stephen Anderson (2000) has traced down this event, the argument in itself and its impact on the genesis of generative self-understanding until the present day (collective memory relies on the transmission of myths). He shows that while the argument was not exactly irrefutable even by structuralist standards of the late 50s, it was nevertheless the initial spark of the tabula rasa policy that early generative thinking promoted in the 60, and which incarnated as SPE (Chomsky & Halle 1968) at the end of the decade. That is, generativists sought to install the collective impression that generative phonology was fundamentally different from anything that man had produced before, that structuralist thinking was wrong and refuted en bloc, and that the cultural break initiated by Halle's LSA lecture opened the way for a new science. It should be added that this tabula rasa perspective is phonology-specific to some extent since the structuralist heritage in syntax was less incisive.

At this point considerations of social nature are usually made: through the 60s, first-generation generativists, which had all been brought up in the (mostly European) structuralist and diachronic tradition, trained young students without transmitting any structuralist knowledge and purposefully sent them to structuralist conferences where they would play the part of the agent provocateur who aggressively opposes the old thinking. This is what it took in order to get into the system, to occupy academic positions and so on. By the early 70s, the structuralist resistance was by and large annihilated, and the new generative paradigm ruled. This kind of social explanation is also instrumental in the debate regarding extrinsic rule ordering that is reported below (e.g. Goldsmith 2008:56f).

As was indicated in the introduction, I do not dwell on external history, but I think that a paragraph can mention what established historiography (e.g. Encrevé 1997, Koerner 2002a: 169ff) contributes to the issue (both regarding facts and the generative story-telling). Let us review below a number of issues that have been claimed to divide structuralist and generative thinking, and which may be instrumental in assessing the revolutionary picture.

3.2. Cognitive realism

A property that is typically quoted in this context is cognitive realism. It is certainly true that some structuralists considered the objects and structures that they were construing as abstract entities which describe the state of the language from the observer's point of view, but have no relationship to any physical implementation: the phonemic system of a language tells us something about the language independently of whether language is located in the brain, in the knee or on Mars. Hjelmslev's glossematics certainly defends this position.[4]

It is also true that anti-mentalism along the lines of behaviourist thinking (we cannot talk about things that are not accessible to direct investigation and measurement) was common: linguists do linguistics on the grounds of linguistic data, neurologists do neurology on the basis of neurological data, and there is no point in mixing up both; different disciplines look at the same object from different perspectives, and none is more important or insightful than any other.

American Structuralism was led onto this track by Leonard Bloomfield, who aimed to establish linguistics as a science that is freed from the philological burden and lives up to methodological standards of other disciplines (and namely of behaviourism in psychology). In an article entitled "Language or ideas?", Bloomfield (1936) argues in explicit reference to the aforementioned Vienna Circle for what he calls physicalism (and what today would be called reductionism or eliminationism, see section 8.4).

> "Linguistics as actually practised employs only such terms as are translatable into the language of physical and biological science; in this linguistics differs from nearly all other discussion of human affairs. Within the next generations mankind will learn that only such terms are usable in any science. The terminology in which at present we try to speak of human affairs – the terminology of 'consciousness', 'mind', 'perception', 'ideas', and so on – in sum, the terminology of mentalism and animism – will be discarded, much as we have discarded Ptolemaic astronomy, and will be replaced in minor part by physiological terms and in major part by terms of linguistics." Bloomfield (1936:89)
>
> "If we are right, then the term 'idea' is simply a traditional obscure synonym for 'speech-form', and it will appear that what we now call 'mental' events are in part private and unimportant events of physiology and in part social events (responses which in their turn act as stimuli upon other persons or upon the responder himself), namely acts of speech" Bloomfield (1936:95)

This illustrates Bloomfield's "mechanist" approach to language (his own term) and makes the empiricist/behaviourist roots of American structuralism explicit (see also Goldsmith & Laks forth, Anderson 1985:281f, Koerner 2002b).

Important figures of European structuralism were also anti-mentalist. Trubetzkoy (1939:12) writes that "the term 'psycho-phonetics' that Baudouin de Courtenay proposes must be rejected, since phonetics […] have much greater affinity with psychological phenomena than phonology, whose objects are defined above the level of the individual, i.e. have social value" (my own translation). Martinet (1960:8f, 1968:96f) also refuses the mentalist

---

[4] Of interest in this context is a review of Harris' Methods in Structural Linguistics where $Householder, 1952 #3637£ (1952) defines two positions regarding linguistic structure according to whether structure pre-exists analysis or is its result: "God's truth" linguists try to describe a structure that exists in languages independently of the linguist, while hocus-pocus linguists try to impose a structure on amorphous data in an attempt to organize them so that they make sense.

perspective along the same "social" line: he counts language among "human institutions" alongside with the family for example (family and language exist for social reasons).

Structuralist anti-(or non-)mentalism was an important target of the generative literature in the 60s (e.g. Katz 1964, see Koerner 2002a:188 on this issue): cognitive realism was opposed to purely taxonomic activity. In this context, the term *taxonomic* with a depreciative connotation in direction of structuralist practice was coined by Chomsky (1964:11); it stigmatizes the fact that the ultimate goal of structuralist inquiry is the mere establishment of inventories, of lists of items, rather than the understanding of how language is actually processed by the human mind/brain. One should be aware of the fact that taxonomic activity was at the heart of many scientific advances in the 19th century (e.g. the classification of species in biology that led to Darwin's insight, the periodic table in chemistry) and therefore, if anything, had a good reputation in science. Its transformation into a battering ram against structuralism in generative quarters thus needs to be understood in the context of the monist vs. dualist dispute: Chomsky mourns the *merely* taxonomic character of structuralist work. That is, he does not deny the necessity of taxonomic inventories – but just like in biology, chemistry and elsewhere, he argues, this cannot be all that needs to be found out: once we know what the (phonemic or lexical) units are, we need to determine the properties of the (generative) system that manipulates them.

This all being said, it is also true, however, that there is a mentalist tradition in structuralist thinking. Apart from Ferdinand de Saussure himself, the most quoted item here is certainly Sapir's (1933) article entitled "The psychological reality of the phoneme". During his fieldwork, Sapir had made the recurrent observation that illiterate speakers systematically insist on the fact that two sounds which are distinct phonemes are distinct, while two sounds in complementary distribution are the same thing. Jones (1967:212ff) offers an informed survey of the three conceptions of the phoneme that were entertained in structuralist thinking: physical (the phoneme is a group of sounds), functional (the phoneme is a unit that distinguishes meaning) and mentalist (the phoneme is a mental image). The founder of the mentalist tradition is actually the founder of phonemic theory, Baudouin de Courtenay (whose "psychological" conception is opposed by Trubetzkoy 1939:37f).

On another front, Postal (1968:269ff) describes "the mentalistic character of sound change" in structuralist work. Finally, Hockett (1948a: 269f) writes that "the analytical process thus parallels what goes on in the nervous system of a language learner, particularly, perhaps, that of a child learning his first language. [...] The child's 'analysis' consists on the other hand, of a mass of varying synaptic potentials in his central nervous system" (see also Hockett 1948b). Laks (2003, 2005) presents a very complete set of facts regarding this question.

Given this record, it must be concluded that cognitive realism is unable to draw a red line between structuralist and generative thinking, even if it is true that while all generativists subscribe to it, structuralists are divided. A related issue is the empiricist - rationalist distinction. While Sapir and Hjelmslev certainly represent the latter tradition, structuralism as a whole is regularly interpreted as a typical instantiation of empiricist and positivist thinking (e.g. Goldsmith & Laks forth). Bloomfield's position as expressed in the quote above supports this interpretation. The fieldwork-based tradition of American structuralism favoured indeed a purely bottom-up perspective where data are accumulated and slowly transformed into scientific statements. The so-called discovery procedure that is discussed next is an expression of this orientation.

3.3. Level Independence and modularity: whether or not and how morpho-syntactic information is used in phonology

Another touchstone of generative thinking that is often quoted (e.g. Durand 2006:2266) concerns the structuralist prohibition to use morpho-syntactic information in phonology. This prohibition is known as Level Independence and roots in the descriptive activity of structuralist fieldwork (Aronoff 1980 provides a good overview): it was held that the aforementioned discovery procedure had to guarantee a strictly bottom-up methodology whereby phonetic units (phones) are analyzed in order to build phonological units, whose analysis in turn give rise to morpho-phonemes, which finally lump together in clauses. Since units of the next higher level only exist thanks to the analysis of lower levels, the analysis at any given level cannot be influenced by items of higher levels that do not yet exist. By contrast on the generative side, the production scenario is top-down: syntax and morphology precede phonology; once their work is completed, morpho-syntactic information enters phonology in terms of labelled brackets and hash-marks #, to which phonology may freely refer.

It is certainly true that the structuralist prohibition to use morpho-syntactic information in phonology was real. It is also true, however, that structuralists were smart enough linguists to understand that phonology may be conditioned by morpho-syntax; therefore the law was systematically obviated by so-called juncture phonemes which appeared as "#" or "+" in the phonemic string at morpheme- and word boundaries (Scheer forth a discusses this issue at greater length). Level Independence thus enforced a translation of obviously morpho-syntactic information into truly phonological vocabulary, which at that time were phonemes. As a consequence, the phonemic status of juncture phonemes enforced free distribution: the occurrence of an independent phoneme must not be restricted by any contextual conditions. Hence juncture phonemes could not appear just at morpheme- or word boundaries: they also had to be found morpheme-internally.

Therefore Moulton (1947) for example sets up phonemic transcriptions such as /+pa+pir+/ and /+la+terne+/ for German *Papier* and *Laterne*, respectively. He first argues that "minimal pairs" such as *ich antworte: Terrasse* [ʔɪçˌʔantvoʁtə**t**^{h}ɛ'ʁasə] "I answer 'terrace'" vs. *ich antwortete: Rasse* [ʔɪçˌʔantvoʁtə**t**ɛ'ʁasə] "I answered 'race'" establish aspiration as a phoneme according to regular phonemic analysis, but that this can be circumvented if the phonemic level can make reference to the beginning of the word. In this case, voiceless stops are aspirated word-initially. Fortunately enough for him, aspiration also occurs morpheme-internally before stressed vowels. *Papier* [p^{h}a'p^{h}iiʁ] and *Laterne* [la't^{h}ɛʁnə] therefore identify as /+pa+pir+/ and /+la+terne+/. That is, all aspirated stops are preceded by a "+", whose distribution is thus unconstrained: it "accidentally" appears at all word boundaries.

Admitting carriers of morpho-syntactic information in the middle of morphemes, though, has disastrous consequences since they are for free and can be abused (actually were amply abused, e.g. Hill 1954) for predicting any alternation and its reverse. Generative phonology has done away with morpheme-internal boundary symbols (which existed only as the end point of a logical domino effect whose origin was Level Independence) precisely because it revoked the ban on using morpho-syntactic information in phonology. The difference between structuralist and generative practice, then, is one of overtness, rather than of principle: generativists merely legalized (and formalized) current structuralist practice.

This is also Goldsmith's (2008:49ff) conclusion, who points out that there were strict defenders of Level Independence in structuralism (Martin Joos, Charles Hockett) alongside with more liberal voices (Zellig Harris, Kenneth Pike) who allowed occasional "look-ahead" into morphology (e.g. Pike 1947). In other words, generativists stopped pretending that morpho-syntactic information needs to be freely distributed across the string because it is a phoneme. This, in turn, is a consequence of the fact that there were no phonemes in generative phonology anymore: the basic unit now was the segment, which was not associated

to any distributional requirement (in SPE, boundaries were said to be segments of a special kind, i.e. which bear the feature [–segment]).

But there is more to this debate: Level Independence is the structuralist way to express modularity. It sets into stone that morpho-syntax and phonology are two distinct ontological entities that do not speak the same language and can exchange information only through some kind of translation whose output is truly phonological vocabulary – hence the transformation of morpho-syntactic information into juncture phonemes (see Scheer forth a). Regarding this key question of grammatical architecture, structuralist and generative thinking thus go hand in hand and are opposed en bloc, as we will see in section 8.6 below, to linguistic applications of (pieces of) empiricist connectionism that either deny the existence of modular contours, or water them down (OT). Structuralist and generative thinking also converges with respect to dualism and its opposition to empiricist monism: Saussurian Langue vs. Parole is the structuralist equivalent of Chomskyan competence vs. performance (or I- vs. E-language).

Another debatable difference between structuralist and generative thinking is Universal Grammar and the genetic endowment of the language faculty. Space restrictions preclude going into any further detail (the question is addressed for example in Anderson 1985:280f).

3.4. Extrinsically ordered rules

Among all candidates for telling structuralist and generative phonology apart, though, extrinsic rule ordering is the most debated one. This way of relating underlying and surface representations is the application of computation to linguistics as conceived of in Cognitive Science. Serialism was also practised in syntax in the form of ordered transformations through the 60s and 70s until the Pisa lectures (Chomsky 1981).

Structuralists recognized an underlying and a surface form, but were not really explicit on, or very interested in what was happening between them. As long as two phones could be predicted from the context and hence a single phoneme could be set up, it did not matter a lot what the underlying object exactly looked like (diacritics were commonplace, e.g. /E/ for complementarily distributed [e]/[ɛ]), or what exactly happened when the phoneme was transformed into the surface object.

A central feature of generative phonology, extrinsic rule ordering, thus really seems to set generative endeavour apart from structuralist phonology. This is the interpretation that could be found in the regular generative record, where Chomsky's (1951) Masters thesis was held to be the birth of extrinsic rule ordering – until Bloomfield's (1939) Menomini Morphophonemics were rediscovered in generative circles. The central mechanism laid out in this article were also extrinsically ordered statements that transformed morphophonemes into phonemes/phones. Namely since Bromberger & Halle's (1989:65ff) "note on recent history", a body of literature has developed that debates the issue, and especially the question whether Chomsky, who does not mention Bloomfield's article in his Masters thesis, could really have been unaware of its existence (Huddleston 1972, Encrevé 1997, Koerner 2003, 2004b).

Bromberger & Halle argue that Bloomfield's article was so untypical for its time and overthrew established wisdom in such a way that it was not taken seriously and enjoyed an accordingly weak posterity in the 40s and 50s. This is also what Chomsky (1986:13, note 3) says: "Bloomfield (1939) […] was radically different in character from the work of the period and inconsistent with his own theories of language, and remained virtually without influence or even awareness despite Bloomfield's great prestige." Goldsmith (2008) has studied this issue in great detail. He presents an article by Wells (1949), published in Language, which is about "automatic alternations", i.e. which are non-allophonic. Goldsmith shows that Wells' study of alternations that involve two distinct phonemes is perfectly generative in the sense that it sets up a system where underlying ("hypothetical" in Wells' terminology) forms are

related to the surface through intermediate ("fictive") forms, which are created by extrinsically ordered rules.

Goldsmith therefore concludes that extrinsically ordered rules can certainly not be used in order to oppose structuralist and generative thinking. Rather than the revolutionary legend that was entertained in generative quarters, the written record documents a continuity whereby central properties of generative phonology root in ideas that were marginal, but certainly present in structuralist thinking. Goldsmith (2008:57) therefore writes that "we should not characterize the work of linguists such as Wells, Harris and Hockett as the last gasp of a dying structuralism, but as a body of scholarship out of which generative phonology was a natural development."

3.5. Static system vs. dynamic derivation, systemic vs. environment-based pressure

This notwithstanding, of all candidate criteria for the distinction between structuralist and generative phonology, extrinsic rule ordering seems to me the one that is most insightful. For it expresses the difference that both theories carry in their name: structuralism is static, while generativism is dynamic. Structuralists aim at building structure, whereas the goal of generativists is to *derive* all and only those strings that are well-formed. The presence of dynamic seeds in structuralist thinking (Goldsmith 2008 documents the timid discussion of the issue namely by Harris 1951, who was supportive but left it at a marginal question) does not withstand the global characterization of the two major linguistic theories of the 20th century in these terms. The transition, then, may well be described as a shift in prominence of the dynamic aspect: a quantum jump confers a new quality to the accumulation of events that remain effectless until they reach a certain density.

The static vs. dynamic character of the theories is also instrumental when one turns tables and does not only look at what generativism has that structuralism did not have (this was our perspective thus far). An interesting question is also to ask what structuralism had, but which was not taken over in generative phonology. The answer is quite striking: no less than the structuralist key insight, i.e. the relevance of the system, and of systemic pressure. The idea that was born in linguistics but has irradiated far beyond this discipline (e.g. sociology, Lévi-Strauss 1958, psychoanalysis, Lacan 1966) is the following: a sound is not a linguistic object per se – it is promoted to this status only by the (paradigmatic, i.e. vertical) relationship that it entertains with other sounds of the system in which it lives. In other words, there is no such thing as a phonological object in isolation: every phonological unit relates to other units so to form a system – or a structure.

As a consequence, phonological units cannot be considered in isolation, neither as an object with individual phonetic properties, nor as an actor in a phonological process. Two phonetically identical items may react on the same external pressure in completely different ways according to the place that they occupy in their respective systems. Work that explains (diachronic) phonological processes by systemic pressure is based on Martinet's (1955) push- and drag-chains (see also Kuryłowicz 1948).

Systemic pressure is completely absent in SPE where segments are exposed to one single (syntagmatic) force: their environment. As a consequence of the universal ambition of generative grammar, then, cross-linguistic studies of, say palatalization in 300 languages are made which compare things that are not comparable. Quite strikingly, SPE or the early generative literature for that matter does not argue against the relevance of the system for phonology – it just eclipses it without discussion. This is one of the reasons that have led to the aforementioned tabula rasa interpretation of the generative takeover in the 60s: the central insight of the past three decades was simply ignored without discussion or refutation.

Contrast and systemic pressure were absent from generative thinking until the early 80s when Lexical Phonology introduced the notion of structure preservation ($Kiparsky, 1982 #1824£ 1982:45): "[at level one, a] phonological rule may not apply to create some segment which is nondistinctive – that is, not a phoneme in the language" in $Borowsky, 1989 #3635£'s (1989:148) wording. Another instrument of the 80s also appealed to contrast: a specific version of underspecification, contrastive underspecification, proposed that only features which contribute contrast are present underlyingly, the others being filled in by default rules (Steriade 1987, Archangeli 1988).

Structure preservation and contrastive underspecification did not have much posterity, though. It is only very recently that generativists have recognized the mistake to ignore the obvious existence of a relationship between sound inventories and phonological processing. As far as I can see, the attempt at incorporating this aspect into generative theory started with so-called Licensing Constraints in Government Phonology (Charette & Göksel 1994, 1996, Kaye 2001) and more recently is pursued by work in Toronto around Elan Dresher (Hall 2007, $Dresher, 2009 #3572£ 2009), and by Nick Clements (2000, 2001). Contrast is also a prominent issue in OT (Steriade 2007 provides a well-informed overview of the literature). An informed survey of the history of contrast since Sapir can be found in Dresher (2009:37ff).

It took generative phonology quite some time to understand that it was in error to throw the system over board, and the recent evolution towards cumulativity of scholarly insight can be taken as a real progress: lexical items are exposed to two forces, one paradigmatic and static (the system), the other syntagmatic and dynamic (the environment). The structuralist school has focused on the former, the generative enterprise on the latter – but phonology is made of both.

3.6. Conclusion: continuity, rather than rupture

In conclusion, the revolutionary story-telling that is found in generative quarters since the 60s is not supported by evidence: no property of generative phonology can be identified that was absent in structuralist thinking. However, this does not necessarily withstand the generative interpretation of the transition: it is typically true that what was to become a central tenet of generative thinking was present in the structuralist literature, but confined to a niche existence. The variety of positions that were taken by structuralists is so large that any given idea that is sent into the structualist valley will have a good chance to provoke an echo, however weak. Unlike generative linguistics, structuralism was an amalgam of a variety of approaches with just a few common orientations (such as the systemic notion and the phoneme), rather than a doctrinal edifice.

The impression of early generativists that they were struggling against a structuralist establishment which was hostile to their ideas is therefore hardly surprising, and may be actually by and large correct. It is idle to charge generativists with ignorance regarding Bloomfield (1939), Wells (1949) or Harris (1951): what they did is to effect a quite radical change by defining a new centre of gravity for phonological inquiry, even if theirs is not necessarily the paternity of the ideas.

What the discussion boils down to is the tabula rasa-based story telling that is practised in generative quarters, which needs to be corrected: Encrevé (1997), Koerner (2002a) and Goldsmith (2008) are instrumental in this sense. By contrast, the eviction of the systemic idea from generative phonology without refutation or even discussion is to be considered as a conscious obstruction of cumulativity that the field is only repairing today.

4. Normal science: SPE

4.1. The generative architecture: modularity and translation

As indicated by the section title, the pages below follow Kuhn's (1962) analytic tool for the evolution of scientific theories:[5] first generation generative phonology was established during the 60s and incarnated as SPE in 1968. It was then practised for a decade or so with the status of what Kuhn calls normal science, i.e. following the canonical rules established without calling them into question even though irritating or frankly conflicting evidence was available and well-known. In the case of SPE, its Achilles' heel was made explicit by Chomsky and Halle themselves in the famous ninth chapter of SPE, on which more below.

Since normal science was practised and largely spread before the actual publication of SPE (of which individual chapters were circulated in the 60s), both the revolutionary and the revisionist movement that it aroused could be launched in 1968, or shortly thereafter. The revolutionary attack that set out to overthrow the system as such was led by the Natural Phonologies (which originate in David Stampe's 1972 Ph.D), while the revisionist perspective was opened by Paul Kiparsky's (1968-73) manuscript *How abstract is phonology?*

The communist and socialist enterprises are the subject matter of section 5 below. Before we turn to them, let us look at some key properties of SPE, and namely at those which have set the generative standards and are still in place today against all odds and subsequent evolutions. SPE makes explicit the architecture of generative grammar that was heralded in Aspects (Chomsky 1965). This architecture is a direct expression of the frame that Cognitive Science defines for cognitive activity (see section 2): language as such is a module, but its internal structure is modular as well.[6] The following quote makes explicit that generative phonology is conceived of as a cognitive module, i.e. a computational unit that modifies an input according to a pre-defined set of instructions that applies "mechanically", i.e. without taking into account any external factors (computation is autistic or, in CogSci terminology, encapsulated).

> "The rules of the grammar operate in a mechanical fashion; one may think of them as instructions that might be given to a mindless robot, incapable of exercising any judgment or imagination in their application. Any ambiguity or inexplicitness in the statement of rules must in principle be eliminated, since the receiver of the instructions is assumed to be incapable of using intelligence to fill in gaps or to correct errors." Chomsky & Halle (1968:60)

The modular architecture of the generative grammar is embodied in the so-called inverted T model that is depicted under (1) below. While the inverted T was first introduced in Aspects (Chomsky 1965:15ff), the modalities of communication between morpho-syntax and phonology were defined in a book on phonology, SPE, rather than in work on syntax. This bias also runs through further evolution: theories of the interface between morpho-syntax and phonology are phonological theories made by phonologists on the grounds of phonological data (with one semi-exception, Distributed Morphology).[7]

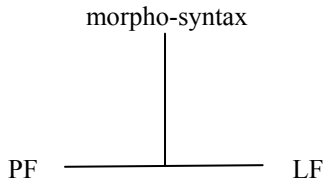
---

[5] The revolutionary metaphor is commonplace in linguistic historiography and, as Koerner (2002a) points out, part of generative story-telling. Koerner dwells on the adequacy of this reference at greater length, also in regard of Kuhn's model.

[6] A fair debate is about whether we are talking about sub-modules, or just about an aggregation of independent modules whose collaboration happens to produce language. The same issue arises in other areas of the cognitive system: we know today that vision for example is made of different sub-systems that compute colour, shape, face recognition and movement independently.
Another aspect of the debate regarding language-internal computational units is the fact that Chomsky (1981) has called the sub-systems of GB modules (case theory, theta theory etc.). Space restrictions preclude further discussion (but see Scheer forth a).

[7] Syntacticians usually do not go beyond statements such as "and then PF takes over". In the recent minimalist

(1) the inverted T model:
all concatenation before all interpretation



In the perspective of the inverted T, (morpho-)syntax is the central concatenative system which has the privilege of concatenation: its output is sent to two interpretational modules, phonology (PF) and semantics (LF), which assign a form and a meaning to the morpho-syntactic structure.[8] Morpho-syntax and the two interpretative modules are thus procedurally ordered so that words and sentences are pieced together before being shipped off to interpretation at PF and LF. That is, the function of phonology in the (production-oriented) generative perspective is to translate morpho-syntactic structure into a code that may be used by the sensory-motor system.

"Phon" in the word phonology is thus an (understandable) misunderstanding: natural language is modality-independent. That is, linguistic structure may run through any channel as long as humans are able to produce and perceive contrast. Two modalities are known to be used by humans in order to ship messages in natural language: vocal and signed. The former is the default, but the latter is used if the former is not available (there are documented cases where deaf children develop a sign language ex nihilo, i.e. in absence of any linguistic stimulus, e.g. $Senghas, 2004 #2829£ et al. 2004). A more accurate definition of phonology and its function is thus the translation back and forth between a physical signal and morpho-syntactic structure: phonology is a translational device.

In addition, SPE adds an extra proviso (which does not follow from the inverted T structure): all concatenation must be done before all interpretation. While the inverted T model stands unchallenged up to the present day, the full completion of concatenative activity before any interpretation can begin will be a bone of contention in further development. The competing view is so-called interactionism according to which concatenation and interpretation are interspersed. We will see below that this is the key idea of Lexical Phonology, and today also the backbone of minimalist syntax.

A direct consequence of this modular architecture is the necessity for *translation*: every module works on its own proprietary vocabulary (this is how modules are defined in CogSci: domain-specificity, e.g. Hirschfeld & Gelman 1994). Morpho-syntax for example reacts on things like number, gender, animacy and so forth, which is what the morpho-syntactic part of lexical entries is made of. By contrast, labiality, occlusion and the like are entirely transparent to morpho-syntax. On the other hand, phonology works on these, but is unable to parse number, gender and the like.

Therefore SPE and all subsequent generative theories establish a translational mechanism that transforms morpho-syntactic structure into objects that can be read by the phonological computation. In SPE morpho-syntactic structure is translated into hash-marks # by a mapping algorithm (all major categories as well as projections thereof are preceded and followed by a #), later on into prosodic constituency (Prosodic Phonology, Selkirk 1981 et

---

environment, they actually use PF as a dustbin for things that they want to get rid of in order to have a "clean" syntax (e.g. deletion of words or even entire phrases "at PF"). The "dirty" phonology, then, is not anything that they are concerned with (see Scheer forth a).

[8] Whether syntax and morphology are the same or two distinct computational systems is a debated issue that does not matter for the discussion.

passim). Being a consequence of modularity (which, recall, is shared with structuralism), translation is thus a stable property of structuralist and generative practice – the only thing that changes is the output, which is defined in terms of the phonological units that were used in the respective theories: juncture *phonemes* in structuralist times, hash-marks which were defined as a special kind of segment in SPE (#s were [-segment] segments, while "real" segments were [+segment]), and finally autosegmental constituency in the early 80s when all areas of phonology were autosegmentalized.

On the other hand, it is also true that SPE violates the translation requirement since apart from hash-marks, it also makes brackets and labels available in the phonology, that is cyclic structure (on which more shortly) and genuine morpho-syntactic vocabulary such as verb, noun, adjective. For example, a word such as *theatricality* appears in phonology as $[[[\text{theatr}]_N \text{ ic} + \text{al}]_A \text{ i} + \text{ty}]_N$, and phonological computation can freely refer to the information that this or that portion of the string is a noun, a verb or an adjective. Hence typical SPE-style rules such as "k palatalizes before e, but only if this e is the dative singular". We will see below that reference to so-called morphological diacritics is one of the reasons why SPE came under fire in the 70s. The problem was only (partially) solved in the mid-80s when Prosodic Phonology established the so-called Indirect Reference Principle (e.g. Nespor & Vogel 1986) which prohibits reference to untranslated morpho-syntactic information in phonological processes.

4.2. Cyclic derivation

Another key property of generative thinking, cyclic derivation, is directly related to the modular architecture. The idea that phonological (and semantic) interpretation follows morpho-syntactic structure, i.e. proceeds from inside-out (from the most to the least embedded constituent) bears the stamp of a generative copyright. That is, a string with the morpho-syntactic structure [[[A] B] C] is computed in such a way that first A is interpreted alone; the result is then assessed together with B (i.e. AB), and finally the product of this computation is interpreted together with C, i.e. ABC. It is important to note that the computational system, i.e. the set of instructions that cause modifications, is invariable: whatever the string, it is interpreted by the same phonology. Also, there was no requirement regarding so-called strict cyclicity (a notion that we return to below) in SPE, and Chomsky & Halle (1968) are explicit on the fact that cyclic derivation concerns morphological as much as syntactic structure: it applies "to all surface structure whether internal or external to the word" (Chomsky & Halle 1968:27).

Cyclic derivation represents a genuinely new idea that was absent from linguistic thinking in Antiquity and the Middle Ages, and from the neogrammarian and structuralist paradigms in modern times. It first appeared in Chomsky et al. (1956:75), but was codified only in SPE, where it is called the Transformational Cycle. In the 70s, it was also known as the Phonological Cycle (Mascaró 1976), and today cyclic derivation (or cyclic spell-out) is known as derivation by phase, which is the spine of current minimalist syntax (Chomsky 2000 et passim).

4.3. Anderson's prism: structure vs. process

Finally, let us consider something that SPE does *not* have, and which will become central in the further evolution of the field: representations. It was already mentioned that structure (and hence systemic pressure), the heart of structuralist thinking, was eclipsed in SPE. On a scale whose extremes are computation and structure, SPE certainly ranges very close to the computational end. It is true that SPE is not only made of computation, if only because

computation necessarily computes something – objects –, which are thus different from computation. This is the dualistic philosophy of Cognitive Science: modules are input-output devices that take *objects* as input, transform them, and return their altered versions.

What are thus the objects that are manipulated by computation in SPE? Feature matrices that are made of binary features. While serialism (i.e. extrinsically ordered rules) is closely associated with generative phonology (albeit not its invention, as we have seen in section 3.4), Chomsky & Halle have never claimed the paternity of binary features: these were introduced by Roman Jakobson (Jakobson 1939, Jakobson et al. 1952) in a structuralist environment. SPE has merely replaced their original acoustic definition by articulatory values. In this environment, thus, extrinsically ordered rules change the values of articulatorily defined binary features, which are the only structure that SPE recognizes. Or rather, to be precise, SPE recognizes yet another type of item beyond binary features: segments, which are an aggregation of features. A segment is defined as a feature matrix whereby all segments are made of the same features and the same number of features.

The non-computational part of SPE, its structure, is thus made of features and feature matrices. An interesting question that is not usually asked in the (contemporary or later) literature is whether it makes sense to talk of this kind of structure in terms of representations. As we will see below, the advent of autosegmental representations has also introduced the notion of *well-formedness*, which is central for phonological thinking in the 80s. It is therefore reasonable, in any case useful, to establish a strong bond between representations and well-formedness: only representations can be well- or ill-formed, and something can only be a representation if it can be ill-formed. This, however, is clearly not the case of features and feature matrices: there are some combinations that either cannot or do not occur, but never because of an intrinsic impossibility that is defined by grammar.

For example, a vowel cannot be [+high] and [+low] at the same time – but this is for physiological reasons. And some language may well not possess this or that combination of feature values of the universal set of features – actually most of the astronomical number of logically possible combinations do not occur. But this again is for reasons that have got nothing to do with grammar: inventory selection and overgeneration will drive SPE into trouble as we will see below. Finally, core properties of features (such as their binary nature: they have either a positive or a negative value, no third choice is available) and of feature matrices (e.g. no feature may occur twice in a given matrix) are set in stone before the derivation begins: no lexical item can contravene, and no computational action can create monster features without value or monster matrices. By contrast, autosegmental representations may become ill-formed in the course of a derivation. They therefore introduce a new quality into phonological thinking, which in retrospect turns out to be a milestone in the history of phonology.

Let us now place this discussion in a larger context. The balance between structure and process is the prism that Stephen Anderson (1985) has sent light through in order to understand the history of phonology in the 20th century. Anderson has detected a regular see-saw movement between theories that stand far on one side of the spectrum, and others that approach the opposite extreme. Also, both properties are in a reverse proportional relationship: when one goes up, the other goes down.

A correct prediction suggests that the balance between representations and computation is a valid instrument for the understanding of what the swarming come and go of terminology, theories, concepts and schools is all about: writing at the representational peak of the 80s, Anderson extrapolates that phonology stands at the dawn of a new computational, hence anti-representational round. Here is the last sentence of his book (p.350).

> "If current attention to the possibilities of novel sorts of representations leads to a climate in which the importance of explicit formulation of rule-governed regularities disappears from view, the depth of our

> knowledge of phonology will in all likelihood be poorer for it. We hope that this book has demonstrated that neither a theory of rules nor a theory of representations constitutes a theory of phonology by itself."

Little did he know how right he was, i.e. how far on the extreme computational end OT would take phonology a couple of years later. Given its predictive success, the structure vs. process prism will be used below as a measure of the evolution of the field. Anticipating on the facts to be reported, we may note that the generative micro-cosmos (as compared to Anderson's larger span) has achieved a complete loop from an almost exclusively computational theory over the heavily representation-oriented autosegmental 80s back to the exclusively computational environment of OT (where literally no structure is left since even structure is supposed to "emerge", i.e. to be the result of computation). It was mentioned before that going in circles does not really witness a mature field.

5. Revolution and revision

5.1. What is wrong with SPE: creating a revolutionary situation

5.1.1. Overgeneration

The fundamental problem of SPE is that it can describe all phonological processes that exist, as well as all others. For a theory that set out to build a system (a grammar) that is able to generate all structures that are attested (or well-formed), and none that is not attested (or ill-formed), this situation amounts to a declaration of bankruptcy. Chomsky & Halle (1968) had lucidly spotted the problem in the ninth chapter of SPE, but the remedy that they were suggesting, a theory of markedness, never really emerged.

Overgeneration concerns both structure and computation, and is regularly pointed out in the post-SPE literature (e.g. $Goyvaerts, 1981 #540£ 1981, $Kaye, 1989 #1791£ 1989:58ff). It was already mentioned that given free combinability of positive and negative values, the fixed (and universal) set of binary features (say, twenty) creates a bewildering number of logically possible segments, of which of course only a vanishing fraction exists (either in any given language or cross-lingusitically, even when physiological impossibilities are counted out).

On the computational side, the universal rule format A → B / K does not impose any restriction on what kind of object can instantiate A, B and K. Anything may be turned into any other thing (no restriction on the relationship between A and B): n → ŋ / __k is just as plausible as n → f / __k or r → k / __ŋ. Also, anything can provoke any change (no restriction on the relationship between A → B and K): n → ŋ / __k is just as good as n → ŋ / __p. Recall that this kind of arbitrariness is shared with structuralism.

It is obvious for most phonologists[9] that natural language does not work like that. Certain processes are recurrent (e.g. dental nasals become velar before a velar obstruent, dental nasals become labial before a labial obstruent), others are rare, and some do not occur (e.g. an [s] that is turned into [b], see Ewen & Hulst 2001:3ff on this). It is therefore wrong for sure to leave both variables A and B without any restriction.

It is also self-evident for most phonologists that any context may not provoke any effect: there are precise causalities,[10] and it is not reasonable to assume that a given object may trigger any process and its reverse. Hence, the relationship between K and A is not free

---

[9] Though not for all: today substance-free phonology (Hale & Reiss 2008, Blaho 2008) holds that substance, i.e. melodic properties of segments, are irrelevant for phonological computation: substance is the subject matter of phonetics (if phonology talks about substance, it is liable for substance abuse). Hence anything and its reverse can be turned into anything and its reverse in any context and its reverse.

[10] Even if it is true that a number of processes escape this rule: this is a central issue of the abstractness debate that is discussed in sections 5.2 and 5.3 below.

but obeys precise causal patterns: given A, K may only be chosen among a restricted subset of possible objects. And for any given trigger K, only a restricted subset of objects that instantiate A will react (dental nasals may become velar before a velar, but not before a labial stop).

Finally, wild overgeneration is also promoted by so-called morpho-phonology: it was already mentioned that phonology could freely refer to morpho-syntactic labels (e.g. "velars palatalize before front vowels, but only if these represent the dative singular") since SPE granted free access to labelled brackets that were inherited from morpho-syntax.

5.1.2. Related items must have a common underlying form

Another area where overgeneration produced strange blooms is the relationship between etymologically, paradigmatically or semantically related items. In SPE and early generative phonology, it was admitted without discussion that items for which the linguist can determine a relatedness of this kind must have a common underlying form from which the surface variation is derived by phonological rules. Otherwise, it was argued, a generalisation is missed.

For the sake of illustration, it is useful to choose the wildest case on record: the work by Theodore Lightner is certainly not representative for the average post-SPE phonologist. But the fact that theory allowed Lightner to do what he did without objecting is an unmistakable indication that something is wrong: (phonological) theory must somehow define what a possible derivational relationship is.

Lightner (1978:18f,1981,1985) holds that the following pairs are derived from a common underlying form: *eye* and *ocular*, *thunder* and *detonation*, *dental* and *tooth*, *rebel* and *bellicose*, *cardiac* and *heart*, *three* and *third*, *gynaecology* and *queen*, *sweet* and *hedonism* and so on. Since the alternations h-k (***h****eart* - ***c****ardiac*), d-θ (*thir****d*** - *four****th***) and s-h (***s****weet* - ***h****edonism*) suppose Grimm's Law, Verner's Law and the Ancient Greek s > h shift, Lightner concludes that these processes are performed by the grammar of present-day English natives.

What these examples show is that the more two items that are supposed to be related by a phonological derivation are distant (etymologically, paradigmatically, semantically), the more demanding they are for underlying representations and for phonological computation. That is, the underlying representation that needs to be set up so that *sweet* and *hedonism* can be derived from it is highly *abstract*, i.e. distant from the surface form (of at least one of the items). Also, the phonological rules that are needed in order to carry out the transformation are numerous and restate diachronic processes that have occurred several hundred, sometimes several thousand years ago. Therefore a typical feature of SPE-style analyses is to propose synchronically underlying forms that are close to or identical with diachronically distant stages of the language (in the case of SPE Old English, sometimes Common Germanic), and to set up rules that mimic diachronic evolution. It does not take a lot to understand that present-day natives have not acquired Old English lexical forms and the rules that transformed them over centuries.

The origin, development and (non-)justification of the tacit rule according to which a large (phonological) computational component and a small number of lexical entries make a better world than the reverse proportion, or anything in between for that matter, is discussed in a large body of literature (e.g. Foley 1977, Kenstowicz & Kisseberth 1977, 1979; Anderson 1985:331f provides an informed overview).

The alternatives to a phonological derivation that is based on a common underlying form are known and have always been practised: related items such as *electri[k]* and *electri[s]-ity* can either represent two distinct lexical entries (*electricity* is stored as a single lexical entry, which means that its pronunciation does not involve any morphological

concatenative or phonological activity), or allomorphy (there are two stems recorded in the lexicon, *electri[k]* and *electri[s]*, which are selected by a morphological, rather than a phonological context: the morpheme *-ity* selects the latter).[11]

The two issues mentioned in section 5.1 defined the research agenda in the 70s. To cut a long story short, everybody tried to reduce the expressive power of the grammar – its abstractness. By which means exactly this should be done was the dominant question. In any event, all phonological theories that individuated in the early and mid 80s have to some extent learned the lesson that many alternations which early generativists believed were produced by online phonological computation do not represent a synchronically active process. That is, alternating forms may as well be recorded as independent lexical items, or related by allomorphy. The set of alternations that phonological theory is called to account for, then, is far smaller than what SPE thought it was: phonology has to shrink in order to be viable.

5.2. Revolution: Natural (Generative) Phonology

The major challenger of SPE in the 70s were the two Natural Phonologies, one generative, the other not. The initial spark of the latter is David Stampe's Ph.D (Stampe 1972), which has been developed into Natural Phonology (NP, Donegan & Stampe 1978, 1979, Dressler 1974, 1984, Hurch & Rhodes (eds.) 1996, Dziubalska-Kołaczyk 2002). The former, Natural Generative Phonology (NGP), was founded by Theo Vennemann (1971, 1976 [1971][12]) and developed namely with his student Joan Hooper (today Bybee: Vennemann 1972a,b, 1974a,b, Hooper 1975, 1976).

Both Natural Phonologies share a basic set of assumptions and principles, but as indicated by their name one is more revolutionary than the other: while Natural Generative Phonology accepts the basic generative architecture, Natural Phonology has left generative grounds, namely on the count of functionalism, which is endorsed.[13]

The Natural Phonologies promote a radical means of reducing the expressive power of the grammar: they cut down the set of alternations that represent phonological activity by some 80% or 90% in comparison to SPE (my estimate). In order to do that, a structuralist notion is revived that was abandoned by SPE: the morpho-phonological level. That is, N(G)P does not deny that an alternation such as *electri[k]* - *electri[s]-ity* is real – only does it represent morpho-phonological, rather than phonological computation. In other words, we face allomorphy (or suppletion in traditional vocabulary).

In NGP, the two criteria that divide alternations into one or the other type are the True Generalization Condition (Hooper 1976:13ff) and the No-Ordering Condition (Hooper 1976:18ss). According to the former, only phonetically accessible information can be used in the formulation of phonological rules, while the latter prohibits rule ordering. On this backdrop, alternations that do not suffer any exception in the entire language and exclusively

---

[11] The process whereby k is turned into s (before *-ity*) is called velar softening. It is emblematic for the discussion of derivational relatedness and has therefore aroused a substantial body of literature that includes Chomsky & Halle (1968:219ff,426f), Kiparsky (1982a:40f), Kaye (1995:312,328), Halle (2005) and McMahon (2007). Hayes (1995a) and Green (2007:172ff) provide a documented overview. On the structuralist side, Harris (1951:221f) argues in favour of allomorphy.

[12] Vennamann (1976 [1971]) is quoted as a 1971 manuscript by Hooper (1975) and Zwicky (1974). The development of NP and NGP was parallel, rather than based on a common ancestor.

[13] See Newmeyer 1998 for an overview of what he takes to be the two basic approaches to linguistics, formalist and functionalist thinking. There is no place to further discuss the differences between the two Natural Phonologies. Laks (2006) offers more material on this issue (se also Hooper 1975:544ff) and also discusses their further evolution: while NGP has no modern offspring (Joan Hooper, who today publishes under her maiden name Joan Bybee, is engaged into a usage-based, i.e. empiricist approach, and thus has left generative grounds altogether), NP continues to be active (see the references quoted).

appeal to phonetically retrievable information are called natural and granted phonological status (P-rules). By contrast, alternations that are not surface-true and/or whose statement requires non-phonetic, that is morpho-syntactic information, are rejected into the pool of morpho-phonemic rules (MP-rules). This contrast is the NGP-adaptation of Stampe's (1972) original distinction between processes (P-rules) and rules (MP-rules).

On Stampe's view, the former are natural, innate, productive, unsuppressible and effect minimal structural changes (hence they apply in loanword adaptation and interfere when non-native languages are spoken), while the latter are conventional, learned, dispensable (they do not interfere in loanwords and when foreign languages are spoken) and may be responsible for structural changes that involve more phonetic distance. In this environment, Stampe's perspective on language acquisition is that what needs to be done in order to acquire a language is to suppress those innate processes that are not effective in the language at hand. The split between processes/P-rules and rules/MP-rules reinstalls structuralist Level Independence in its rights, this time however without circumventing it in practice.

This restrictive definition of phonology attacks two of the three sources of overgeneration in SPE: reference to morpho-syntactic information (which is prohibited for P-rules) and rule ordering (which is also ruled out in NGP). The third kind of overgeneration is the arbitrary relationship between the context of a rule and the change that it causes. Here it is hoped that the restriction to phonetically accessible information will automatically introduce sound causality.

This setup was accompanied (or even partly provoked) by a diachronic reasoning regarding the life-cycle of alternations. That is, what we see today in alternations such as *electri[k] - electri[s]-ity* are vestiges of a once-phonologically controlled alternation that has aged (e.g. Vennemann 1972a). Alternations are born as phonetic regularities before moving into grammar where they are first phonological but at some point start to be riddled with morphological conditions, followed by lexical factors, and finally are levelled out or eliminated from the language by some other means. Therefore, asking the question how much of what we see is controlled by phonology is if not identical, at least concomitant with the question how much diachronics are in synchronic sound patterns.

In sum, then, Natural Phonology is a radical answer to the overgeneration problem, and the label revolutionary is certainly adequate, be it only in recognition of the fact that extrinsic rule ordering, the prime candidate for the generative identity, is outlawed (at least in NGP). Natural Phonology is thus anti-abstract in the sense that it allows for minimal derivational depth (i.e. distance between underlying and surface forms). Finally, phonology looks quite phonetic in its mirror: only phonetically defined items can be taken into account by phonological rules (P-rules). It needs to be noted that the three ingredients mentioned – prohibition of morpho-syntactic information in phonology, anti-abstractness and the reduction of phonology to phonetic and surface-true statements – do not suppose each other. Gussmann (2007) for example endorses the former, but rejects the two latter.

5.3. Revision: Kiparsky

5.3.1. How abstract is phonology?

The revisionist research programme was laid out in Kiparsky (1968-73) and carried out mostly by Paul Kiparsky himself through the 70s.[14] It attacks only one source of overgeneration, the one that is mentioned in the title of Kiparsky's seminal article: abstractness, i.e. the "derivational distance" between underlying and surface forms. The two other sources of overgeneration, reference to morpho-syntactic information and the arbitrary relationship between the context of a rule and the change that is caused, are left untouched.

Abstractness is attacked along two lines: Kiparsky restricts possible underlying forms, and he marshals the complexity of the derivation by two means: the restriction of computation to so-called derived environments and the requirement for rules to use material that is freshly introduced on the last cycle. The two restrictions on computation are united under the roof of Kiparsky's version of the Strict Cycle Condition (SCC).

The Alternation Condition (Kiparsky 1968-73, 1973a) defines what a possible underlying representation is: in case a morpheme shows no alternation on the surface, it must not be any different in its underlying form. This results in a ban against so-called absolute neutralisation, i.e. items in underlying forms that never appear on the surface (on which more shortly). On the other hand, Kiparsky (1973a) restricts the application of rules in such a way that a certain rule class may only target derived environments. This was called the Revised Alternation Condition: obligatory neutralization rules apply only in derived environments. An environment is derived iff it has been produced by the concatenation of two morphemes, or by the application of a phonological rule. This embodies the idea that phonological processes do not apply to monomorphemic strings, i.e. when the trigger and the target belong to the same morpheme.

Trisyllabic Shortening (or Laxening) may illustrate the ban on absolute neutralization and the request for derived environments. The process produces alternations whereby a long vowel or diphthong of bisyllabic items appears as a short vowel when a suffix is added: *div[aj]ne - div[ɪ]nity*, *op[ej]que - op[æ]city* etc.[15] Monomorphemic items such as *n[aj]ghtingale (nightingale)* and *[aj]vory (ivory)*, however, systematically resist the process, even though they satisfy the trisyllabic condition.

SPE reacts in a non-systematic way that misses the obvious morphological generalisation and sets up abstract underlying forms: the application of the rule is eluded simply by destroying either the target or the triggering context of each individual lexical item. Instead of /aj/, *nightingale* is said to have an underlying /i/, i.e. /nixtVngael/ (Chomsky & Halle 1968:234); and instead of /i/, the last segment of *ivory* is made a glide, i.e. /ivorj/ (Chomsky & Halle 1968:181). Independent rules that are ordered after Trisyllabic Shortening then take /ix/ to [aj] (via /i/), and vocalize the final glide of *ivory*.

The Revised Alternation Condition kills two birds with one stone: 1) it dispenses with the absolute neutralization of /i/ in /nixtVngael/ and /j/ in /ivorj/ (which never appear on the surface) because it 2) offers a different reason for the non-application of the rule, i.e. the request for the triggering environment to be morphologically complex.

Another trouble when the underlying form of invariable morphemes may be distinct from their surface form are so-called free rides. In our example, these concern the converse surface situation, i.e. cases where the third bust last vowel of a monomorphemic item is short.

---

[14] Cole (1995) and Bermúdez-Otero (forth:§1.2) provide a good overview of this movement, which is also explained by Kiparsky himself in Kiparsky (1982b, 1993), and by Anderson (1981:530ff).

[15] Trisyllabic Shortening is probably not a synchronically active process: it faces quite a number of counterexamples such as *obese* [ɔwbiis] - *obese-ness* [ɔwbiisnɛs], *obes-ity* [ɔwbiisitɪ]. Hayes (1995) and Green (2007:172ff) provide an informed review of its status today.

Hence Kiparsky (1982b:148) points out that on the SPE analysis, the underlying form of *[æ]libi*, *c[æ]mera* or *P[æ]mela* cannot be determined: the third but last vowel could either be faithfully short and hence appear as such on the surface, or it could be underlyingly long, i.e. */aj/libi*, */kaj/mera*, */paj/mela*; in this case, the /aj/ will be turned into [æ] by Trisyllabic Shortening – a free ride, i.e. one without consequences. The ban on absolute neutralization also does away with this unwarranted indeterminacy.

Finally, the restriction of rule application to derived environments allows phonology to do away with some of the incriminated reference to morpho-syntactic information. For example, vowel shortening applies to *mean* [miin] – *meant* [mɛnt] but not to *paint* [pejnt], *pint* [pajnt], *mount* [mawnt] because the latter are morphologically simplex, and their long vowel is specified as such in the lexicon (rather than derived by rule): /miin+t/ vs. /pejnt/. Shortening, then, applies only to derived environments (Kiparsky 1985a:87).

On the other hand, Kiparsky's struggle for defining a reasonable line of division between computation and the lexicon also meant that certain parts of morpho-syntactically conditioned phonological processes need to be defended against the concretist ambition. By the late 70s indeed, the abstract SPE mainstream was opposed by so-called concrete analyses that followed the principle "the more concrete an analysis (i.e. the less distance between underlying and surface forms) the better it is" (e.g. Leben & Robinson 1977, Tranel 1981).

This line of thought had to struggle with the fact that all attempts at devising a formal measure of different degrees of abstractness failed: a so-called evaluation measure (or evaluation metric) could not be set up (Kiparsky 1974, Campbell 1981, Goyvaerts 1981). In this context, abstract analyses with no other limitation than the learnability of rules (cf. Skousen 1981) persisted (e.g. Dinnsen 1980, Gussmann 1980, Dresher 1981). Kenstowicz & Kisseberth (1977:1-62) offer extensive discussion of the issue. They argue in Kenstowicz & Kisseberth (1979:204ff) that in some cases, there is no alternative to the abstract option.

This is also about the position of Kiparsky, who favours grammar-internal principles such as the Revised Alternation Condition and the SCC (to be discussed next) and is not prepared to do anything just in order to pay tribute to concreteness. Lexical Phonology, the theory that emerged from the abstractness debate in the early 80s (on which more below), may be viewed as an attempt at maintaining as much morpho-phonology as possible in the computational device of phonology while cutting away the wildest outgrowths of unrestricted SPE (see Scheer forth a).

Finally, Kiparsky has merged the condition on derived environments with Chomsky's (1973) Strict Cycle Condition that was originally devised for syntax and later on applied to phonology (Kean 1974, Mascaró 1976). Chomsky's (1973:243) original formulation is as follows: "[n]o rule can apply to a domain dominated by a cyclic node A in such a way as to affect solely a proper subdomain of A dominated by a node B which is also a cyclic node." The effect is that rules are blocked whose structural description is met by a string which is made exclusively of material that belongs to previous cycles. That is, given $[[AB]_i\ C]_j$, a rule that is triggered by AB can apply at cycle i, but not at cycle j. Or, in other words, multiple application of rules is prohibited.

Chomsky's (and Kean's and Mascaró's) condition on the applicability of rules is entirely irrelevant for derived environment effects: it will not prevent rules from applying to monomorphemic strings since these have necessarily been introduced on the latest (the only) cycle. Thus Trisyllabic Shortening (*s[ej]ne - s[æ]n-ity*) will happily apply to *n[aj]tingale* and *[aj]vory* under Chomsky's SCC.

Nonetheless, Kiparsky (1982a,b) introduces his version of the SCC as if it were just a restatement of Mascaró's.[16]

---

[16] Cole (1995:72) discusses the two distinct origins of Kiparsky's SCC at greater length, and Scheer (forth a,b)

"With some simplification, his [Mascaró's] proposal was:

(47) Strict Cycle Condition (SSC):
a. Cyclic rules apply only to derived representations.
b. Def.: A representation φ is derived w.r.t. rule R in cycle j iff φ meets the structural analysis of R by virtue of a combination of morphemes introduced in cycle j or the application of a phonological rule in cycle j."

Kiparsky (1982b:153f)

Kiparsky's attempt to kill two birds ("use new material!" and derived environment effects) with one stone (his scrambled SCC) was considered an important achievement in the 80s, but then turned out to lead into a dead end: the derived environment condition being riddled with counter-examples, Kiparsky (1993) himself declares the bankruptcy of his version of the SCC ten years later.

This does not mean that the original Chomskyan baby in Kiparsky's blended SCC ought to be thrown out with the bathwater. The idea that "old" strings that have already experienced interpretation are unavailable for further computation was picked up by Kaye (1992, 1995) who developed the form under which it appears in current syntactic Phase Theory, where it is known as the Phase Impenetrability Condition (PIC) (Chomsky 2000 et passim).

5.3.2. The evaluation measure and common underlying forms

Following the footsteps of Kiparsky's revisionist programme which was only marginally concerned with the question of common underlying forms, a substantial body of literature tried to outlaw Lightner-type excesses without however jumping into the NGP conclusion that phonologically driven alternations must be surface-true and may not make reference to any morpho-syntactic information.

Objective and measurable criteria were sought that could be applied to any given alternation in order to decide whether it is the result of phonological computation on the basis of a common underlying form or not. Candidate criteria that were discussed tried to measure what is natural, simple, elegant, phonetically plausible, psychologically real or typologically invariant (e.g. Hellberg 1978, Koutsoudas 1980, Dinnsen 1980), but this was quite inconclusive.

To date the question remains open whether an item (such as *electric-ity*) that looks morphologically complex is really considered as such by the grammatical system. All modern theories have somehow swung into a midfield position: it is not good or bad per se to have a big or a small lexicon, or to do little or a lot of phonological computation. Arguments must be made on a case-by-case basis: while some alternating items represent two independent lexical entries for sure, the online computation of others is beyond doubt; the swampy midfield, however, is large enough for still much debate to come. Carvalho (2002:134ff) provides extensive discussion of this question (and Carvalho 2004 considers the role of analogy as a fourth player besides phonological rules, independent lexical entries and allomorphy).

The discussion regarding common underlying forms also runs under the header of (anti-)lexicalism (treating *electric-ity* as a single lexical entry is the lexicalist position). The parallel oscillating evolution in both phonology and syntax is quite remarkable: after a decidedly lexicalist period in the second half of the 70s and all through the 80s (triggered in

---

follows the development of all no look-back devices from Chomsky (1973) until the modern syntactic Phase Impenetrability Condition. Finally, it is to be noted that the marriage of Chomsky's SCC with derived environment effects was first proposed by Halle (1978).

syntax by Chomsky's 1970 remarks on nominalisation), the minimalist environment in syntax, and OT in phonology today have moved back to the anti-lexicalist heydays of the 60s (e.g. Newmeyer 2005, Williams 2007).

6. Lexical Phonology

6.1. Interaction with morphology: the lexicon, stratal architecture and interactionism

At the outset of Kiparsky (1982a,b), the author explains that Lexical Phonology (LP) is the synthesis of three strands: affix classes, affix ordering and the abstractness debate. Looking back over the 50 years of generative phonology, LP is probably the theory that was (and is) most influential besides SPE. OT is younger, and it remains to be seen whether it will show the same adaptive capacity as LP. The inherent interest of LP is demonstrated by the fact that it could sail through three quite different general environments with no major modification: rule-based post-SPE in the early 80s, the autosegmental 80s and today output-oriented OT (in the coat of DOT and Stratal OT, on which more below).

It is of course impossible in the frame of this chapter to do justice to all aspects and sub-theories of LP. Giegerich (1999:7ff), McMahon (2000:35), Rubach & Booij (2003) and Bermúdez-Otero (forth) offer well-informed historical overviews (see also Scheer forth a). The goal of this section can only be to introduce the basic idea, i.e. what LP stands for still today. As a matter of fact, people do not think of the abstractness debate when LP is mentioned. Rather, they think of a theory that has defined the general generative architecture for the interplay with morphology and syntax. Indeed, the synthesis of affix classes and affix ordering has produced a new idea, so-called interactionism, which lies at the heart of LP and today is the spine of minimalist syntax.

The existence of (English) affix classes was known since SPE (Chomsky & Halle 1968:84ff): affixes fall into two classes whose members share a number of morphological and phonological properties that are not observed for items of the other class. Class 1 items are (roughly) of Romance origin, while class 2 items belong to the Germanic stock of vocabulary. The former include *in-*, *-ity*, *-ic*, *-al* (adj.-forming), the latter *un-*, *-ness*, *-less*, *-hood*, *-ing*.

The most prominent phonological effect concerns stress placement, which is commonly admitted as the basic diagnostic for class membership. Class 1 affixes are reputed to be stress-shifting, while class 2 affixes are stress-neutral. Classical examples are words such as *párent*, *válid* and *átom*, which appear with right-shifted stress when occurring with a class 1 suffix (*parént-al*, *valíd-ity*, *atóm-ic*), but conserve their lexical pattern when followed by a class 2 suffix (*párent-hood*, *válid-ness*, *átom-ise*).

This much was known since SPE. The critical new generalization came from Dorothy Siegel's (1974) dissertation and is of morphological kind. Siegel observed that class 1 affixes always occur closer to the stem than class 2 affixes. That is, affixes of both classes can freely attach to stems that already contain an affix of the same class (class 1: *atom-ic*$_1$*-ity*$_1$, *univers-al*$_1$*-ity*$_1$, class 2: *atom-less*$_2$*-ness*$_2$, *beauty-ful*$_2$*-ness*$_2$, *guard-ed*$_2$*-ness*$_2$). In addition, class 2 affixes can hook onto a class 1 affix (*univers-al*$_1$*-ness*$_2$). However, sequences of class 2 - class 1 affixes do not occur (**atom-less*$_2$*-ity*$_1$, **piti-less*$_2$*-ity*$_1$, **guard-ed*$_2$*-ity*$_1$ etc.). This observation is known as the affix ordering generalisation.[17]

Interactionism affords to kill two birds with one stone: the phonological and the morphological effects of affix classes can be derived with the same mechanism. The idea is to intersperse word formation rules with phonological rules: first you apply phonology to a piece, then you concatenate an affix, then you do some more phonology on the new string created, then you concatenate another affix etc. This procedural scrambling of morphology

[17] Which has turned out to be wrong: Aronoff (1976) and Aronoff & Sridhar (1983,1987) have collected a substantial number of counter-examples such as *develop-ment*$_2$*-al*$_1$ or *organ-iz*$_2$*-at-ion*$_1$.

and phonology was proposed by Pesetsky (1979), Booij (1981), Williams (1981) and Kiparsky (1982a,b).

Interactionism accounts for affix ordering by assuming the existence of several procedurally defined levels (or strata) that the morpho-phonological derivation runs through without being able to loop back (a restriction known as level ordering): class 1 affixation is done at level 1, hence before class 2 affixes are concatenated at level 2. This guarantees that a stem may take on a (number of) class 1 affix(es) and then move on to level 2 where class 2 affixes join. The result are sequences of class 1 - class 2 affixes. The reverse order, however, cannot be generated because this would imply a return to level 1 once the derivation has added an affix at level 2.

On the phonological side, two-step affixation allows for a two-step interpretation, hence capturing the insight of cyclic derivation: starting with the stem, phonological interpretation works its way outwards through the affixes. For example, a root such as *párent* appears with regular penultimate stress when it occurs in isolation; adding the stress-shifting class 1 affix *-al* produces *parént-al*, while the stress-neutral class 2 item *-hood* yields *párent-hood*. Another way of looking at these facts is that both *párent* and *parént-al* bear transparent penultimate stress, while *párent-hood* illustrates an opaque non-penultimate pattern where stress behaves as if the suffix were not there. In other words, stress has been reassigned when *-al* was added (stress-shifting), but reassignment was blocked upon the merger of *-hood*. The task for the analyst is thus to organize underapplication of the stress rule, which must somehow be prevented from reapplying to strings that are headed by class 2 affixes.

Table (2) below shows the solution that is proposed by Lexical Phonology.

(2) *párent - parént-al* vs. *párent-hood* in Lexical Phonology

		parent	parént-al	párent-hood
lexicon		parent	parent	parent
level 1	concatenation	—	parent-al	—
	stress assignment	párent	parént-al	párent
level 2	concatenation	—	—	párent-hood
	rule application	—	—	—

The spine of Lexical Phonology is its stratal architecture: the lexicon contains underived roots, all class 1 affixes are concatenated at stratum 1 (level 1), while class 2 affixes join in at stratum 2 (level 2). After the concatenation is complete at each stratum, a stratum-specific phonology applies to the string as it stands. Rules are assigned to specific strata: in our example, the stress-assigning rule is a level 1 rule, which means that it is active at level 1, but absent from level 2. Another ground rule is that the derivation is strictly serial: given the order lexicon → level 1 → level 2, strings that are present at some level must run through all subsequent levels on their way to the surface. This means that they experience the computation at these levels.

Under (2), then, /parent/ in isolation receives stress at level 1, where stress assignment is active. This is also true for /parent-al/, since *-al* has been concatenated in time. Stress assignment to /parent-hood/, however, concerns only /parent/ since *-hood* has not yet joined in. After its concatenation at level 2, stress does not move since the stress rule is absent from this stratum. Note that this is critical: otherwise **parént-hood* would be produced. Underapplication of stress assignment at level 2 is thus achieved by the split of phonological computation into two morpheme-specific mini-grammars: one that assesses class 1 strings (where the stress rule is present), another that takes care of class 2 strings (where the stress rule is absent).

6.2. Interaction with syntax: post-lexical phonology

Affix classes and affix ordering have thus led to the interactionist stratal architecture described. The output of this process are words, and this is where the name of the theory comes from: Lexical Phonology recognizes an independent unit – the Lexicon – whose intertwined morphological and phonological computation produces words. Let us now look at how LP conceives of the relationship with syntax.

Beyond the Lexicon, words serve as the input to syntax. In the perspective of LP, syntax concatenates words, not morphemes: the internal structure of words is invisible to syntax. In SPE (Chomsky & Halle 1968:15), so-called bracket erasure eliminates internal morpho-syntactic divisions at the end of every cycle and thus affords a no look-back effect for transitions between all chunks, not just for transitions between morphemes and words. Based on this heritage, LP stresses the specific invisibility of word-internal structure for phonological processes that apply across word boundaries (e.g. Kaisse & Shaw 1985:4f, Mohanan 1986:23ff, Bermúdez-Otero forth:45ff). LP thus follows the traditional view whereby morphology and syntax are distinct computational systems with distinct grammars and possibly distinct vocabulary (Distributed Morphology has set out to show that the opposite is true). Only is the order reversed: on the traditional view, syntax is done first, and its output is the input to morphology.

Upon the availability of syntactic structure, i.e. after syntax has applied to the output of the Lexicon, the string is assessed by the postlexical block of (phonological and semantic) rules. Booij (1997:264, note 3) traces the distinction between lexical and postlexical phonology back to the Prague Linguistic Circle (Circle 1931) where the phonology of words (*phonologie du mot*) is opposed to the phonology of sentences (*phonologie de la phrase*). Praguian segregation was introduced into modern thinking by Rubach (1981).

LP thus distinguishes two kinds of phonological computational systems: postlexical phonology concerns processes that are either insensitive to any kind of extra-phonological information (morphological and syntactic alike), or conditioned by syntactic structure alone. In the latter case, they could not apply in the Lexicon because syntactic information had not yet been constructed. Conversely, rules that are sensitive to morphological information cannot apply postlexically because morphological structure is erased at the end of each stratum (by the aforementioned bracket erasure).

According to this system, lexical phonological processes (which are traditionally called internal sandhi) interpret morphological structure and allow for exceptions, i.e. may be opaque, while postlexical rules (which are known as external sandhi) interpret syntactic structure and are exceptionless. Today the term postlexical has acquired a lingua franca meaning that is quite independent of the particular theory in which it was born: phonologists talk about postlexical (or late, or surface-true) processes when they are exceptionless (such as the aspiration of English voiceless plosives).

The lexical-postlexical distinction is reminiscent of the structuralist opposition between the morphophonemic and the phonemic level, but not entirely congruent. Like postlexical phonology, computation that applies at the phonemic level is uninformed of any morphological information (or rather: ought to be, in practice this is circumvented, see section 3.3), but also of any syntactic information: the bottom-up discovery procedure (phones → phonemes → morpho-phonemes → clauses) disallows reference to higher level information because this information is not constructed yet. By contrast in LP, postlexical phonology is sensitive to syntactic structure. Related to this difference is a contrasting serial order of components: while the structuralist way is strictly bottom-up in the order mentioned, sentences are constructed before postlexical phonology applies in LP.

Another specific property of LP is that phonological computation in the Lexicon is cyclic (i.e. interspersed with morphological concatenation), while postlexical computation is non-cyclic: it applies only once syntactic action is completed and is thus not intertwined with syntactic concatenation. This contrast is contrary to SPE's conception of cyclic derivation which, recall, concerns morphemes as much as words ("the principle of the transformational cycle […] appl[ies] to all surface structure whether internal or external to the word" Chomsky & Halle 1968:27). Lexical Phonology imposes this major change of perspective without any discussion. Kiparsky (1982b:131f) simply decrees that the derivation of words is non-cyclic.

The question why sequences of morphemes should, but sequences of words should not be derived cyclically is an interesting and understudied issue which I call the word-spell-out-mystery (Scheer 2009, forth a). As a matter of fact, it seems that research on external sandhi has not produced any data or analysis that requires a cyclic treatment along the lines of affix classes and the like: while the cyclic spell-out of morphemes leaves ample phonological traces (this is the object of study of lexical phonology in LP), the supposedly cyclic spell-out of larger units – words – does not seem to provoke any phonological reaction. In other words, there is plenty of external sandhi, but there is no cyclicity-induced external sandhi.[18]

This massive generalization goes unmentioned in the literature (Kiparsky for example could have built on it in order to conclude that postlexical phonology is not cyclic), where syntactic influence on phonological processes is analyzed exclusively with representational tools (# in SPE, the Prosodic Hierarchy later on). Either the empirical generalization (absence of phonological traces of the cyclic spell-out of words) is true, in which case it needs to be made explicit and LP is right in making postlexical phonology non-cyclic; or it is wrong, and the absence of relevant data is due to the bias that was introduced when the non-cyclic character of postlexical phonology was set in stone by LP.

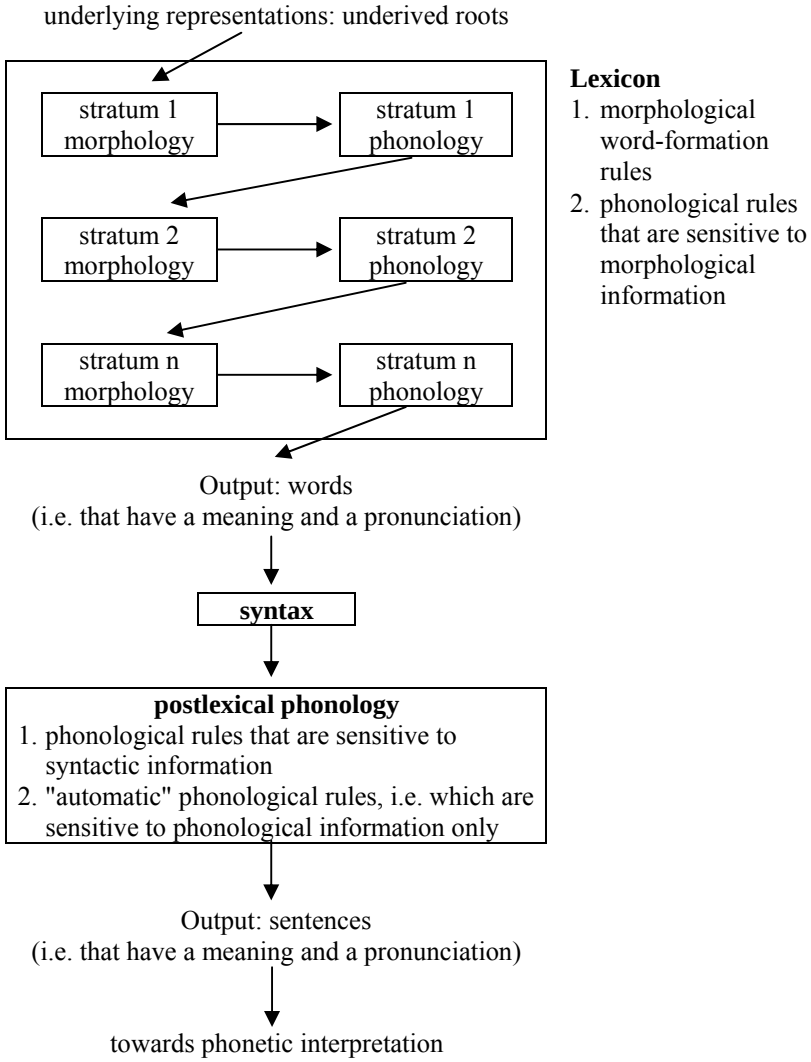
This issue is highly relevant for architectural properties of grammar, especially in current interface-focused minimalist syntax which is based on the piecemeal, i.e. cyclic spell-out of sentences that is called derivation by phase (Chomsky's 2000 original take is that CP and vP, maybe DP, are phase heads, i.e. provoke spell-out and hence interpretation). This is along the lines of SPE (where, recall, the derivation of all chunk sizes is cyclic) and hence in conflict with LP. If the SPE-minimalist option is on the right track, phonology is under piecemeal fire from syntax that hands down words and larger chunks, but refuses to react (while it does react on the piecemeal reception of morphemes). If this really turns out to be empirically true, the different treatment of morphemes and larger chunks begs the question and certainly puts the SPE-minimalist perspective in a delicate position. Therefore the existence of cyclicity-induced external sandhi is an open question that needs further inquiry.

In conclusion, the global architecture proposed by LP appears under (3) below.

---

[18] A special case needs to be made for intonation (phrasal stress): at least since Bresnan (1971), it is established that intonation directly depends on syntactic structure. The topic is covered by a rich syntactic literature, including Berman & Szamosi (1972), Cinque (1993), Kahnemuyipour (2004) and Adger (2006). The distribution of intonation is thus certainly a function of the syntactic structure of sentences, i.e. of external sandhi. It may be the case, though, that intonation is not a (primarily) phonological phenomenon: it typically does not interact with other phonological phenomena, it appears to be predictable solely from syntax and information structure, i.e. in absence of the actual lexical items that will be inserted in the terminal nodes (hence it cannot be phonological: phonology only works after lexical/vocabulary insertion), intonational structure is sometimes argued to be recursive (while recursion is known to be absent from phonology). Intonation is discussed at greater length in Scheer (2009, forth a).

(3)



6.3. Lexical Phonology hallmarks and further evolution

LP thus requires at least two distinct computational systems, lexical and postlexical, where "distinct" means that the set of rules (or the constraint ranking today) is not the same. In addition, those languages which have stratal phenomena (such as affix classes) also possess morpheme-specific mini-grammars (level 1, level 2, eventually more).[19]

It was mentioned that the critical innovation of LP is interactionism: the inverted T is left unquestioned, but the extra proviso according to which all concatenation must be completed before interpretation begins (see section 4.1) is abandoned. All to the good indeed, since this automatically makes brackets disappear from phonology. Recall that in SPE phonology received a string with labelled brackets, e.g. $[[[theatr]_N$ ic + al$]_A$ i + ty$]_N$, where brackets define cycles. Phonology was thus supposed to able to read/parse diacritic brackets. In an interactionist grammar, this unwarranted diacritic disappears because cyclic structure is

[19] The luxuriance of distinct computational systems that are around in the literature is not easy to disentangle (see Scheer forth a for further discussion): SPE had two kinds of rules, cyclic and word-level. The former applied to each cycle, while the latter were active only at one designated point in the derivation, i.e. the word level (in fact only stress rules were cyclic). This distinction does not correspond to anything that we have come across so far. It reincarnates in LP under the header of a specific proposal by Rubach & Booij (1984,1987), who distinguish a block of cyclic rules (SPE: cyclic) and a block of rules that also apply in the lexicon, but only after the last cyclic rule has applied (SPE: word-level rules).

expressed procedurally by the back-and-forth movement between morpho-syntax and phonology (see Scheer forth a).

The other fundamental innovation are morpheme-specific computational systems. SPE recognized chunk-specific mini-grammars, i.e. which are defined by the size of the items that are computed (cyclic vs. word-level rules, cf. note 19), and this line of thought is continued with the introduction of Praguian segregation (i.e. the distinction between morpheme-based and word-based phonology). Distinguishing between morpheme-specific computational systems, however, is a new quality.

In the mid-80s, LP was attacked by orthodox defenders of SPE on the grounds of interactionism: Halle & Vergnaud (1987) propose a non-interactionist version of LP, i.e. where all concatenation is done before all interpretation. This attempt at restoring SPE has rapidly gone down the drain: since the early 90s (Halle & Marantz 1993 et passim), Morris Halle is engaged in perfectly interactionist Distributed Morphology, and interactionism is the spine of minimalist syntax since Chomsky's (2000 et passim) derivation by phase.

This notwithstanding, Halle & Vergnaud (1987) have introduced an important new idea: interpretation-triggering affixes. Rather than being the property of a specific stratum, the fact of triggering interpretation is a lexical property of each affix that percolates into the morpho-syntactic tree and thereby creates projections that are interpretation-triggering (something that is called a phase head in current minimalist terminology). Also, interpretation-triggering is privative in Halle & Vergnaud's system: affixes are either interpretation-triggering or interpretation-neutral, which is quite different from the stratal architecture where all strata trigger the application of some mini-grammar.

In the 90s, this line of thought was further developed in Government Phonology (Kaye 1992, 1995).[20] The main innovations were of two kinds: on the one hand, interpretation-triggering affixes trigger the spell-out of their *sister* (*-hood*$_2$ triggers the interpretation of *parent* in [[parent] hood], rather than of their mother as in Halle & Vergnaud (1987) (where *-al*$_1$ triggered the interpretation of [parent-al], see $Scheer, 2008 #3284£ 2008 for discussion). On the other hand, Kaye (1992, 1995) introduces a no look-back device that inhibits the modification of strings that have already undergone computation on a previous cycle: on the outer domain of [[parent] hood], stress is not reassigned since *párent* has already received stress through phonological computation on the inner domain (while [parent-al] makes only one domain which receives regular penultimate stress). This restriction on computation is quite different from all versions of the Strict Cycle Condition that were entertained since Chomsky (1973), and namely from Kiparsky's (1982a,b). It is in fact what is known as the Phase Impenetrability Condition (PIC) in current minimalist syntax (see Scheer forth b).

During the same period of the 90s, LP was by and large absent from the discussion: the overview book edited by Hargus & Kaisse (eds.) (1993) closed the period of its dominance. An important reason for that was certainly the general hostility against anything that looked serial in general (on which more in section 8.1), and the advent of OT with its anti-derivational architecture in particular. Initiated by Rubach (1997) and Kiparsky (2000), OTed versions of LP are instrumental since the late 90s in challenging the anti-derivational mantra of OT, which has also produced anti-cyclic flowers (on which more below).

7. (Unintended) counterrevolution: autosegmental representations

7.1. A new player, ill-formedness, opens new horizons and buries the revolution

[20] On the basis of a collaboration between Jean-Roger Vergnaud and Jonathan Kaye ($Kaye, 1990 #1922£ 1990) which did not leave any co-authored traces in print (more detailed discussion appears in Scheer forth a).

In the second half of the 70s, the SPE-mainstream was drifting in unfriendly waters without real research programme and without any perspective of a significant evolution. At the same time, the revolution initiated by Natural (Generative) Phonology gained ground. Kiparsky's revisionist work offered some relief, but could not produce more than some kind of SPE-Light. It is then that two movements appeared out of the blue, i.e. without being prepared, solicited or envisioned, which acted as a fountain of youth that made generative phonology take a new start on the grounds of renewed premises. One was discussed in section 6: the initial spark of Lexical Phonology was Siegel's (1974) discovery of affix ordering, i.e. a phonology-external cause that allowed for the establishment of a new global architecture and the unification with Kiparsky's revisionist programme.

The other is to be discussed here: autosegmental representations (also called non-linear, multi-linear). These carry an entirely new dimension into phonological theory and certainly deserve to be called, also retrospectively, the most significant scientific advance that generative phonology has produced since its inception. A good indication for this status is the fact that autosegmental representations are the smallest common denominator of all phonological theories, present and past (including OT, at least in word, on which more below).

Autosegmental representations are groundbreaking because they introduce a new player that was absent in neogrammarian as much as in structuralist thinking: ill-formedness. It was mentioned in section 4.3 that SPE has structure (the material on which computation works, i.e. features and segments), but no representations: a representation is something that can be ill-formed for *grammar-internal* reasons. The potential of ill-formedness as an overgeneration-killer that restricts the expressive power of grammar was rapidly identified, and phonologists put much hope into this new perspective after the straining experience of the abstractness debate. Also, the overgeneration-killing virtue of autosegmental representations is *intrinsic* since their mere existence, associated with the statement of the (universal or language-specific) conditions of ill-formedness, dramatically cuts down possible results. Note that this does not mean that representations are a filter only on outputs: they restrict possible phonological forms wherever they occur: in the lexicon, in intermediate or in surface forms. A violation of well-formedness may either lead to the crash of the derivation, or cause phonology to act in order to repair the offending property (Goldsmith's 1976:27 initial tone-based conception only built on repair strategies, and this line of thought is also embodied in Paradis 1988 and subsequent work).

Representations are certainly more powerful an overgeneration-inhibitor than Kiparsky's revisionist programme, where a phonological computation could crash for computation-internal reasons. It now can also crash because it either works on or produces an ill-formed structure. This gives new flesh to Stephen Anderson's perspective on the equilibrium between structure and process: autosegmental representations take the generative cursor away from the computational extreme of SPE into some midfield position.

The notion of ill-formedness of course is not the only virtue of autosegmental representations (more on this below), and it goes without saying that they have not solved the overgeneration problem as such (for example, they had no effect on the "static" overgeneration inherent in feature matrices, see section 5.1.1). They have, however, brought home the promise of overgeneration-inhibition, at least partly (Lowenstamm 2006 provides a good overview of the credits).

The new autosegmental possibilities opened new horizons, and after an initial period of timid proliferation in the late 70s the entire field was turned upside down when everybody participated in the autosegmentalisation of all of areas of phonology. This emulation rapidly and completely eclipsed the revolutionary vigour of Natural Generative Phonology, which all of a sudden appeared toothless and simply petered out (see Laks 2006 on this decline). By

contrast, natural non-generative phonology continued its development outside of the generative paradigm.

7.2. Genesis and properties of autosegmental representations

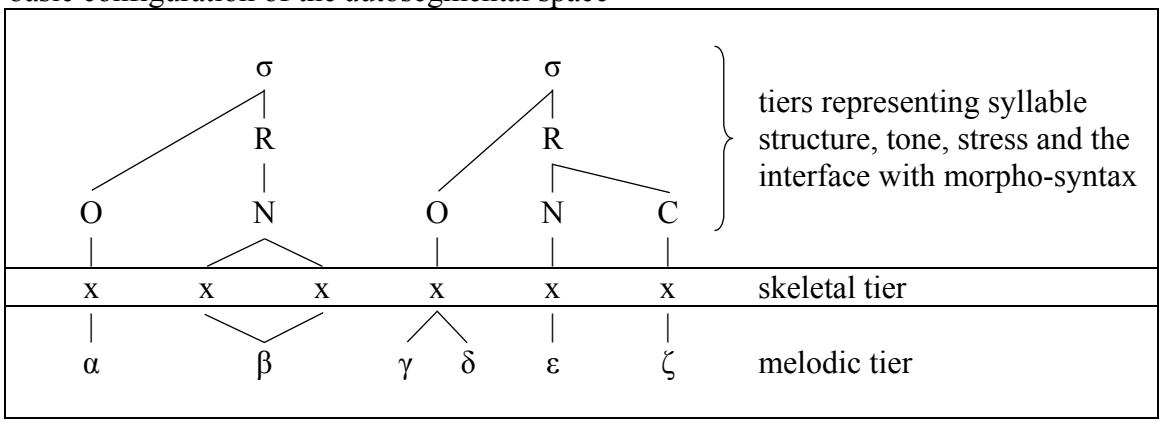
It is reasonable to daresay that Kiparsky's revisionist programme associated to Lexical Phonology would not have been able to capture the victory over the revolution alone. It is only autosegmental representations that could drive revolutionary concerns out of the field. Interestingly, this was completely unintended: autosegmental representations came into being as a problem-solving mechanism whose counterrevolutionary virtue and impact on overgeneration was only discovered as things unfolded.

Autosegmental representations have emerged in the analysis of a number of different empirical problems and quite independently: work on tone (Leben 1973, Goldsmith 1974, Williams 1976), syllable structure (Fudge 1969, Kahn 1976), stress and rhythm (Liberman 1975, Liberman & Prince's 1977), Semitic non-concatenative (templatic) morphology (McCarthy's 1979) and vowel harmony (Clements 1977) have played an important role. The emerging non-linear atmosphere was condensed into a general autosegmental theory by John Goldsmith's (1976a) Ph.D, which is usually credited as the initial spark of autosegmentalism (see Anderson 1985:347ff, Lowenstamm 2006 for a survey). This is correct in the sense that the decisive property of autosegmental representations beyond their graphic and non-linear aspect is introduced in Goldsmith's Ph.D: the idea that a structure may be ill-formed (Goldsmith 1976a,b defines well-formedness conditions).

In careful historiographic work, John Goldsmith himself as well as $Clements, 2000 #3490£ (2000) have also identified pre-generative inspiration for non-linear thinking. This is most obvious for the syllable, which is known at least since the neogrammarian school of the 19th century and widely acknowledged in structuralist work (e.g. Pike & Pike 1947, Hockett 1955:51ff; Pulgram 1970 offers an overview) as well as in Natural Generative Phonology (e.g. Hooper 1972). In addition to that, Goldsmith (1979) for example (see also Goldsmith 1976b, Goldsmith & Laks ms) mentions Harris' (1944) long components, Bloch (1948) and the prosodic analysis of the Firthian school of the 40s. While all of these precursors (from hindsight) heralded the non-linear idea by promoting this or that aspect of it, none of them held all pieces in hand, not to mention their constitution in terms of a formal system. This condensation into a frame that also provides a graphic identity to multilinear structure (which was absent in the structuralist and the NGP conception of the syllable, where only syllable boundaries "$" were inserted into the linear string) has only occurred in the second half of the 70s in generative phonology ($Halle, 1980 #923£ 1980 provide an early overview of the then recent history).

Table (4) below shows a typical autosegmental representation.

(4) basic configuration of the autosegmental space



The idea of autosegmental representations is that two phonological objects (features in SPE) may be related in such a way that no linear order can be determined. This is why SPE is called a linear model (all objects, i.e. features and feature matrices, obey an unambiguous precedence relation with all other objects), against autosegmental representations, which are called non-linear. Under (4) for example, there is no precedence relation between β and the two x-slots that it is attached to: β "belongs" to one as much as to the other. Phonological objects may thus occur on different levels, which are called autosegmental tiers (or lines): under (4) Greek letters belong to the melodic tier, x-slots to the skeletal tier, and higher items to the (or several) syllabic tier(s) (more on the specific identity of supra-skeletal tiers shortly). On every tier, items – which are called autosegments – obey a strict linear order. Non-linear effects are produced when items of different tiers are not associated one-to-one: association may be either one-to-many (e.g. β) or many-to-one (e.g. γ and δ).

This is the baseline that defines the minimal formal properties of autosegmental representations. Two more properties are agreed upon by all phonologists as far as I can see, although they do not follow from formal requirements. For one thing, autosegmental tiers are disposed in three-dimensional space, rather than only in two dimensions as under (4). That is, the melodic tier under (4) could well entertain associations with both the skeletal and the syllabic tier. This specific configuration does not occur in practice, though, because of the second broad agreement, that is the division of the autosegmental space into two basic areas, one accommodating melodic representations, the other constituent structure of various kinds: syllables, stress, interface with morpho-syntax (it is true that tones, which a priori are a piece of melody, commonly appear on the constituent side).

The red line that separates the melodic and the constituent area is the skeleton, which has a special status among autosegmental tiers. Skeletal slots are timing units that define the chronological progression of the overall structure: they are the instrument that defines the linear order of the non-linear structure. Items on other tiers may be related in various ways among themselves and with skeletal slots. When it comes to transform the structure into a linear phonetic signal, the skeleton is parsed slot by slot, and every autosegment of whatever tier that is attached to a given slot will be pronounced on this particular slot. Conversely, items that are not attached to any skeletal slot (such as so-called floating segments, on which more below) are not pronounced.

Indeed, an important ground rule that is shared by all versions of autosegmental representations is that a result on the phonetic side supposes three things: 1) a piece of melody, 2) a constituent that is itself integrated into the constituency above the skeleton and 3) an association line that relates both items. In absence of either, no phonetic trace is produced (e.g. $Itô, 1986 #918£ 1986). This is illustrated under (5) below, where the skeleton is used for the sake of exposition (in moraic theory, other constituents take over its role).

(5) autosegmental conditions on phonetic action

a. phonetic result: [α]	b. no phonetic result	c. no phonetic result	d. no phonetic result
x | α	α	x	x α

There are two autosegmental theories (and only two as far as I can see) that in one way or another do not fit the picture presented. Moraic theory (Hyman 1985, Hayes 1989) is based on the claim that the skeleton does not exist. Phonologically relevant units are based on weight (morae), rather than on chronological precedence. Moraic theory thus challenges standard syllable structure, which is x-slot-based. It assigns morae to segments that contribute syllabic weight, i.e. to vowels (and syllabic consonants) in all languages, parametrically to coda consonants in some, but never to onset consonants.

The other theory to be mentioned in this context is based on the so-called metrical grid (Prince 1983, Hayes 1984, Idsardi 1992, see Halle 1998 for an overview). Metrical grids are erected over segments and indicate their relative prosodic prominence, i.e. regarding stress and rhythm. Prominence is marked by piling up a number of diacritic *grid marks* (which are graphically represented as "*"). Metrical grids thus compete with regular autosegmental arborescence (feet) for the representation of stress and, depending on the particular implementation, also with other autosegmental structure (e.g. the prosodic hierarchy that is responsible for the interface with morpho-syntax in Selkirk 1984). Metrical grids clearly are a child of autosegmentalism and the prevailing non-linear atmosphere of the 80s; it is true, however, that they are not autosegmental themselves because they are in no way non-linear: every segment has its own pile of grid marks, which may shift from one segment to another through computation, but which do not interrupt the linear sequence of segments. That is, a given grid mark is never related to more than one segment, and a given segment never associates to grid marks that dominate neighbouring segments. I leave the autosegmental status of metrical grids an open question here.

Let us now consider ill-formedness, the inherent overgeneration-inhibiting virtue of autosegmental representations. A very early and intuitive condition on well-formedness that is in place until today is the ban on line-crossing (Goldsmith 1976a,b): two items of the same tier (not of different tiers, though, recall that the autosegmental space is three-dimensional) may not entertain associations with another tier that suppose the crossing of association lines. Under (4) for example, ε could not be associated to the x-slot of the coda, and ζ simultaneously to the x-slot of the nucleus. The autosegmental literature has produced a number of well-formedness conditions which have been debated over the years. Three prominent cases in point are the Obligatory Contour Principle (OCP), the Strict Layer Hypothesis (SLH) and the ban on two empty nuclei in a row. The OCP (Goldsmith 1980, McCarthy 1986, Yip 1988) prohibits two identical (or similar) autosegments in a row on the same tier. It has played the role of an output filter for derivations, and also of a trigger for various repair strategies.

The SLH concerns the constituency of the Prosodic Hierarchy (Selkirk 1981, Selkirk 1984:26f). It expresses the idea that a prosodic constituent of a given layer only dominates constituents of the immediately lower level, and is exhaustively contained in a constituent of the immediately higher level. Hence there can be no nested constituents, and no association line can bypass a layer.

Finally, a ground rule in Government Phonology (Kaye et al. 1990, Harris 1994) is that a structure which accommodates two empty nuclei in a row (in the same domain) is ill-formed.[21] Hence in French *il faut y revenir* "one needs come back to this" where the two e's of *revenir* are schwas, either can be left out (*r'venir* is ok as much as *rev'nir*), but not both (**r'v'nir* is agrammatical). On the analysis of Government Phonology, this is because the non-pronunciation of a vowel only means that the association between the relevant piece of melody and the nucleus is interrupted. The nucleus itself is still present and thus empty, and two empty nuclei in a row as in **r'v'nir* are ill-formed.

7.3. Autosegmentalisation of all areas of phonology, blooming theoretical diversity

The potential of autosegmental representations, regarding overgeneration but also a host of analytic advances, was well understood in the early 80s. Therefore early spartan representations were continuously enriched throughout the 80s, leading to rather complex structures. This evolution was parallel to the expansion of arboreal structure in (GB) syntax during the same period.

In contrast to the theoretical monoculture of the 90s and 00s (the large array of OT-internal variation not withstanding), a blooming variety of new theories emerged in the early and mid-80s on the autosegmental ticket. As far as I can see, this is the first time in the history of generative phonology that the mainstream divided into clearly distinguished, partly or completely competing theories that were defined by original assumptions and carried a cleaving name. The only clear precedent is Natural (Generative) Phonology – otherwise phonologists gathered rather around general tendencies or issues such as abstractness/concreteness.

A good witness of this extraordinary vitality are the first two volumes of the then newly established journal Phonology Yearbook (edited by Colin Ewen and John Anderson, today Phonology), which were published in 1984/85 and became the voice of the autosegmental project (the two volumes edited by Hulst & Smith (eds.) 1982 were also a focal point). The Phonology Yearbook 1/2 contains a number of contributions that give Lexical Phonology a more precise shape after Kiparsky's (1982a,b) founding papers (Paul Kiparsky, Jerzy Rubach & Geert Booij, Jerzy Rubach, K.P. Mohanan, Patricia Shaw, including also the first overview by Ellen Kaisse and Patricia Shaw), and they herald the new thematic focus that will dominate the 80s: the internal structure of segments.

The founding statement of the new autosegmental theories also appeared in the two volumes at hand. On the one hand, Feature Geometry will become the mainstream: Nick Clements 1985 proposes an autosegmental arrangement of a set of articulatorily defined binary features that is more or less the same as in SPE. Each feature resides on an autosegmental tier of its own, and features are bundled under labelled nodes that define natural classes of segments, an new and important notion (the standard geometrical system will be condensed in Elizabeth Sagey's 1986 Ph.D).

The alternative system of melodic representation rejects binary primes and instead favours so-called privative (or monovalent, holistic) melodic primes that 1) define items which are bigger than a single feature (e.g. the prime |i| represents the high front tongue body position) and 2) produce contrast by being either present or absent (rather than by being always present but having two distinct values). The privative idea was not new: it was first aired by Anderson & Jones (1974), but now received three distinct implementations in Dependency Phonology (Anderson et al. 1985, condensed in Anderson & Ewen 1987), Government Phonology (Kaye et al. 1985) and Particle Phonology (Schane 1984).

---

[21] This is only a rough and incomplete formulation of the phonological ECP, see Gussmann & Kaye (1993), Scheer (2004) for details.

It is to be noted that the privative option has rapidly gained ground: it entered Feature Geometry under the header of underspecification (Archageli 1988, Pulleyblank 1988). The idea is that plus-minus specifications of some (but not of all) features are absent from lexical representations and come into being through spreading from the context, or by feature-filling default rules that apply at the end of the derivation. Today the question is not so much whether there should be underspecification, but rather which features should be underspecified (see Hall 2007 for an overview).

This implementation of the privative idea is still quite different from the aforementioned theories that are direct heirs of Anderson & Jones (1974): here features are not just temporarily unspecified for a value, but missing altogether. Melodic primes are different in nature (bigger than a classical feature), and they are independently pronounceable, which is not the case of (underspecified) features. Finally, underspecified features often afford a three-way opposition (plus, minus, unspecified), while truly privative systems only oppose the presence to the absence of a prime.

Another important issue was the kind of relationship that melodic primes entertain: while everybody agreed that relations between linguistic objects, in phonology and elsewhere, are asymmetric and hierarchical, opinions diverged as to how this insight should exactly be implemented. The most obvious take is the one of Feature Geometry, which simply extends autosegmental trees to the infrasegmental area. The alternative is the heart of Dependency Phonology, i.e. a dependency relation: the idea is that one item is more important than the other(s) and contributes more to the result of the reunion (cf. also Liberman & Pince's 1977 labelled trees where sisters are either weak or strong). Government Phonology follows this line of thought (here the distinction is between heads and operators), while Particle Phonology expresses asymmetry by the number of copies of the same prime that contributes to the definition of a segment: [i] is the prime |i|, [e] identifies as |i+a|, [ɛ] as |i+a+a| and [æ] as |i+a+a+a| (the presence of several copies of the same prime does not have any impact on the result in other theories). Hulst & Ritter (1999, 2003) provide an informed overview of the various implementations of the dependency programme.

Melodic representation of course was not the only sub-field that was autosegmentalised: all areas of phonology were recast in autosegmental terms. Table (4) shows syllable structure above the skeleton, but this is only an arbitrarily chosen example. The representation of stress and the interface with morpho-syntax was also autosegmentalized: while the former is expressed in terms of feet since Libermann & Prince (1977) (with the aforementioned competing metrical grids), Prosodic Phonology has replaced linear carriers of morpho-syntactic information in phonology, i.e. SPE-type boundaries #, with a multi-layered arboreal structure, the Prosodic Hierarchy (Selkirk 1981, 1984, Nespor & Vogel 1986, see Scheer 2008, forth on this move). Prosodic Phonology also proposed a global unification of all types of constituent structure above the skeleton: the Prosodic Hierarchy encompasses (with some variation regarding e.g. morae and the clitic group) the following units of growing size in a single hierarchical structure (which is marshalled by the Strict Layer Hypothesis): mora < syllable (σ) < prosodic (phonological) word (ω, about the size of a word), prosodic (phonological) phrase (φ, about the size of an NP or a VP), intonational phrase (IP, an intonational unit), phonological utterance (U, about the size of a sentence).

Finally, it is to be noted that in parallel to the autosegmentalization of phonology, the new general research programme for generative linguistics that was defined by Chomsky's (1981) Pisa lectures and gave rise to the Government and Binding framework in syntax very rapidly seeped through into phonology. Grammar was thus divided into principles and parameters: the ban on line crossing for example was a (universal) principle, while there was a fair amount of parametric variation in the application of the OCP. The import of the syntactic model into phonology as a major change in perspective is especially (and explicitly)

recognized in Government Phonology ($Kaye, 1984 #1018£ 1984:123, Kaye et al. 1985:305, Lowenstamm & Kaye 1986:97f, 124ff), but also in Declarative Phonology (Scobbie 1991:7).

7.4. What about computation?

In this rush for representational expansion, a fair question to ask is what happened to computation. At first sight, the answer is: nothing at all. All through the 80s, computation was carried out exactly as before, only that the set of ordered rules now applied to representations, rather than to linear strings of segments and feature bundles. Computation was called to do labour that it did not do before, though, since the most widespread idea regarding representations above the skeleton was that they are absent from the lexicon and hence created by rule. This means that underlying representations were not really different from what they were in SPE (and form what they have returned to today in OT): a linear sequence of segments (except of course that the internal structure of segments was now autosegmental).

Syllable structure for example was erected by a so-called syllabification algorithm over unsyllabified lexical entries on the grounds of major category information provided by the segments. In further evolution, bits and pieces of constituent structure were stored in the lexicon as well. Rubach (1986) for example carries out the autosegmentalisation of vowel-zero alternations in Slavic languages (where alternating vowels are called yers), whose major advance was to distinguish alternating (Polish *pies - ps-a* "dog, NOMsg, GENsg") from stable (*bies - bies-a* "devil, NOMsg, GENsg") vowels that are phonetically and phonemically identical by having the latter, but not the former associated to the skeleton in the lexicon (the linear solution was to add extra items to the vowel inventory, ĭ and ĭ, that were then either transformed into [ε] or deleted, that is subject to an absolute neutralization). On this analysis, thus, alternating vowels are floating, while stable vowels are lexically associated to the skeleton. The very notion of *floating segment*, which is has rapidly gained ground, actually supposes the existence of some constituent structure in the lexicon: it only makes sense to talk about something that floats if there are other items that do not float, i.e. are associated.

Another typical case of floating segments that require the presence of constituent structure in the lexicon is French liaison, where word-final consonants are pronounced or not according to whether the following word begins with a consonant (unpronounced: *les [le] cafés* "the coffees") or a vowel (pronounced: *les [lez] enfants* "the children"). The standard autosegmental analysis of liaison is in terms of floating consonants (Encrevé 1983, 1988). Government Phonology (Kaye et al. 1990) has gone farthest in this direction by assuming fully syllabified lexical entries.

Finally, the nature of computation was significantly changed by the fact that, operating over representational material, all it could do was to link or to delink autosegmental items. Indeed, the typical computational operation in an autosegmental environment is to *spread* some item, i.e. to create an association line between it and some other unit. The palatalization of velars before front vowels for example is understood as the spreading of some palatal prime from the trigger into the target. This contrasts with pre-autosegmental rule application where the structural change operated the modification of a feature value. Therefore Harris (1994:111) for example holds that the "class of possible phonological processes is restricted to operations of delinking or spreading".

We will see in the next section that computation, which was by and large ignored in the autosegmental commotion, will come back like a boomerang in the 90s in order to set back the cursor on Anderson's structure vs. process-scale right to where it stood in the 60s (or even farther towards the computational extreme). Writing at the representational peak of the 80s or rather, on the way uphill, Stephen Anderson (1985) extrapolates that computation will take its

revenge: phonology stands at the dawn of a new computational, that is anti-representational round. Here is the last sentence of his book (p.350).

> "If current attention to the possibilities of novel sorts of representations leads to a climate in which the importance of explicit formulation of rule-governed regularities disappears from view, the depth of our knowledge of phonology will in all likelihood be poorer for it. We hope that this book has demonstrated that neither a theory of rules nor a theory of representations constitutes a theory of phonology by itself."

Little did he know how right he was, i.e. how far on the extreme computational end OT would take phonology a couple of years later.

8. Second revolution: anti-derivationalism

8.1. Anti-derivationalism without argument

The second revolution rages against derivationalism in general and ordered rules in particular. It is far more serious than the first revolution for two reasons: 1) it attacks one of the deepest layers of generative thinking and more generally of Cognitive Science (while the first revolution was only about a "technical" problem, overgeneration, and the distribution of labour between different modules, i.e. the question how much of what we see is dealt with in phonology); 2) it combines an internal concern, i.e. the growing discomfort with ordered rules, with an external solution, Parallel Distributed Processing (PDP, i.e. parallel computation), which is the spearhead of connectionism. Connectionism, however, is the neo-empiricist (neo-behaviourist) alternative that started to challenge the rationalist standard theory of Cognitive Science in the second half of the 80s (Rumelhart et al. 1986 et passim). Section 8.4 below discusses the connectionist import at greater length.

But let us start by looking at what anti-derivationalism actually is, and what its motivation was. Rooting in the properties of the universal Turing machine, derivationalism lies at the heart of the standard theory of Cognitive Science that emerged in the 50s, and whose application to linguistics produced generative grammar. It is the idea that computation in the mind involves a set of instructions that act on the input in such a way that this input experiences step-by-step modifications that occur in a chronological and logical order where the output of step n-1 is the input to step n (see section 2). Serialism (perhaps more appropriate a word than derivationalism), then, boils down to the existence of a set of extrinsically ordered instructions that produce a chronological and logical order of events (and hence of the action of computation).

In generative grammar, serialism incarnates as extrinsically ordered rules in phonology, and in early syntax as extrinsically ordered transformations. These were abandoned in the early 80s when GB replaced them by so-called move α, i.e. a system where movement (computation) is free in itself, but marshalled by constraints on representations (e.g. Newmeyer 1986:163ff). Move α represents an important turn in syntactic theory away from restrictions on computation itself (Chomsky's 1973 original Strict Cycle Condition, extrinsically ordered rules) in favour of a central role of well-formedness constraints on representations such as barriers, the ECP, case checking and so forth. The autosegmental evolution in phonology that was described in section 7.2 follows the same track: representations are marshalled by well-formedness conditions such as the OCP. On the first page of their manuscript, Prince & Smolensky (1993) explicitly draw on the evolution in syntax and declare that their new theory extrapolates the timid phonological precedent into a formal system: "[o]ur goal is to develop and explore a theory of the way that representational well-formedness determines the assignment of grammatical structure. […] The basic idea we will explore is that Universal Grammar consists largely of a set of constraints on representational well-formedness" Prince & Smolensky 1993:1f).

There is thus a red line running from the emergence of autosegmental representations over well-formedness conditions to constraint-based computation.[22] In practice, Prosodic Morphology (McCarthy & Prince 1986), which fully explores the autosegmental tool, was instrumental as a precursor of constraint-based computation. It developed at least two central devices of OT: correspondence theory and alignment of prosodic and morphological constituents (the foreword to the 2001 edition of the manuscript, McCarthy & Prince 2001, explains this evolution in greater detail). We will see below that for reasons to be determined, the implementation of the constraint-based programme in OT has initiated the dissolution of representations as an autonomous actor in grammar.

For the time being, it is enough to take stock of the fact that while generative syntax has abandoned serialism since 1981, the representational blossoming of the early 80s has left serialism entirely untouched in phonology.[23] In the second half of the 80s, though, a diffuse and inscrutable discomfort with ordered rules arose, which quickly turned into a vigorous and lasting antipathy. I have read through the literature of that period in search for indications why serialism is supposed to be wrong, or why it aroused scepticism – without success. I have also asked phonologists who have lived through that period: all confirm that there was indeed a deeply rooted antipathy against ordered rules, and that this feeling was shared by about everybody across theories, but that it somehow remained below the surface. The broad reference to the evolution in syntax set aside, nobody could name, and I was unable to hunt down sources in print that would have explained why serialism is wrong, and why phonology needs to engage into a major cultural break.

I happen to be aware of the reason why Government Phonology participated in the anti-serialist movement: because Jonathan Kaye considered that extrinsic rule ordering was empirically vacuous. According to Kaye, examples where serial ordering of instructions is alleged to be critical are either based on non-existent data, involve misanalysis or concern processes whose properties (no plausible relationship between trigger and effect, exceptions, appeal to morphological information) disqualify them as instances of phonological computation. An example for non-existent data is Martin Joos' famous dialect B of Canadian English for which there is no evidence ($Kaye, 2008 #3638£ 1990, 2008), but which was used by Bromberger & Halle (1989) as the litmus test for rule ordering (see note 24). Examples for processes that are not phonological in nature are trisyllabic shortening (laxening) or other traces of the great vowel shift and velar softening (see the aforementioned electri[k] - electri[s]-ity). The trouble is that apart from Kaye (1990) which is only a short notice about the non-existence of dialect B, there is no trace of this programme in print (Lowenstamm & Kaye 1986:97 mentions that the model of the authors is referred to as the "no-rule approach", but does not say why).

In contrast to the non-overt sources of anti-serialism, the origin of parallel computation is evident: connectionism. Since Rumelhart et al. (1986), the central argument in favour of parallel computation was clearly made and pedagogically repeated (e.g. Rumelhart 1989:134ff): the implementation of serialism in a neuron-based environment appears to be unrealistic given the computational complexity that would be required and the time that it

---

[22] Constraints as such, however, are probably as old as linguistic analysis: so-called morpheme structure constraints defined what a possible morpheme is in SPE, and structuralist as much as neogrammarian thinking appealed to restrictions, requirements and prohibitions.

[23] There was an attempt to do away with extrinsically ordered rules in the 70s around the work of Andreas Koutsoudas ($Koutsoudas, 1974 #3562£ et al. 1974, $Koutsoudas, 1976 #616£ 1976), who claimed that its effects can be derived (and predicted) from other factors. In other words, there is no empirical basis for extrinsic rule ordering, which is misanalysis at best. The evidence presented was not judged convincing, though, and the proposal did not have any posterity. In any event, it was entirely unrelated to the representationally dominated environment of the 80s in general and the antipathy against derivationalism in particular.

would take to carry out all intermediate steps one by one. We know that neurons are not serially ordered in the brain, but rather multiply interconnected. Hence in order to get to grips with a realistic implementation in the brain, several things must be done simultaneously, just like many neurons fire at the same time and thus transmit information simultaneously. This is why we need "brain-style computation" (Rumelhart 1989).

We are thus in presence of an argument that calls for the modification of grammar, which is a model of competence, under the pressure of performance, i.e. neural implementation. It is interesting to observe that Prince & Smolensky (1993:215f) categorically reject this kind of argument ("[i]t is not incumbent upon grammar to compute"), which they consider a category mistake. The rejection of performance-based reasoning is their general-purpose shield against the numerous objections that have been raised against the astronomic inflation of computational complexity in OT (see section 8.3 below).

It is thus idle to speculate where the general antipathy against serialism came from, how it spread in absence of written record, and how an entire field could throw over board a fundamental piece of its identity without any discussion of the reasons. In his article *In defense of serialism*, Clements (2000:193f) makes the obvious point that the rise of connectionism on the Cognitive Science scene has played an important role in the development of anti-derivationalism in phonology. It remains to be seen, though, whether phonologists were closely enough following the development in Cognitive Science, which was only unfolding when anti-serialism was already widespread among phonologists.

In any event, the fact is that the defenders of serialism – of which Morris Halle is the most prominent figure whose position has not varied since the 50s – reacted on the anti-derivational atmosphere by exposing arguments in favour of ordered rules. The article by Sylvain Bromberger and Morris Halle published in 1989 (Bromberger & Halle 1989) discusses the question whether the abandon of ordered instructions by GB syntax and its replacement by the Principle and Parameters approach should lead phonology to follow the same track. The authors reject this perspective because, as they try to show,[24] the subject matter of phonology and syntax is intrinsically different. Bromberger & Halle thus defend serialism against an extra-phonological trend, and also mention the fact that extrinsically ordered rules have come under fire in phonology. Significantly, though, they can come up with only two anti-derivational references: Lowenstamm & Kaye (1986) and a 1987 LSA presentation by B. Majdi & David Michaels. This is good indication for the fact the anti-derivational atmosphere was by and large absent from print: Bromberger & Halle were fighting against an invisible enemy.

[24] Bromberger & Halle argue along the lines of the strongest case, which they claim is the comparison of the well-attested phenomenon of Canadian Raising in Canadian English (dialect A) with another Canadian dialect (dialect B) whereby the only difference is the order of application of two rules. Unfortunately, the only source for dialect B is a three-page article by Joos (1942), whose informants were among his pupils in an Ontario public school. Despite extensive study of Canadian varieties by Canadian dialectologists such as Chambers (1973), though, no trace of dialect B could be found some thirty years later. Despite this demonstration by Kaye (1990, 2008), the Canadian raising case continues to spook though the literature as alleged support for rule ordering.

8.2. Consequences of the latent anti-derivationalism: new theories in the early 90s

The latent anti-derivationalism of the late 80s was the driving force of the events in the early 90s that was now made explicit, but there was still no discussion of comparative merits with serialism, or of the reasons why serialism is a bad thing to have. Rather, parallel computation is typically introduced as an alternative to ordered rules, but leaves it at this juxtaposition (e.g. Prince & Smolensky 1993).

zzzThree theories that have emerged in the early 90s (or late 80s) are based on the anti-derivational mantra: Optimality Theory (Prince & Smolensky 1993), Declarative Phonology (Scobbie 1991, Scobbie et al. 1996) and Government Phonology (Kaye et al. 1990, which was prepared by work quoted earlier since the mid-80s). The computation in all three cases is based on constraints, which however do not have the same status: while they are ranked and violable in OT, they are absolute (i.e. non-violable) in Declarative Phonology. Computation in Government Phonology was not explicitly regulated for some time (the only indication that could be found in print was Kaye's 1992:141 statement according to which processes "apply whenever the conditions that trigger them are satisfied"), but its constraint-based character is obvious since the mid-90s (Licensing Constraints, Charette & Göksel 1994, 1996, Kaye 2001).

Constraints in Government Phonology thus apply whenever a form may be modified by them, but with no extrinsic ranking or ordering, and without being able to be violated: constraints are (simultaneously and) iteratively applied to the string that is submitted to interpretation, and computation ends when no further modification can be made (this is an obvious parallel with Harmonic Serialism, on which more below). To use serial vocabulary, this system is thus able to handle a feeding relationship (the input for the application of a constraint is created by the modification of the string by another constraint), but no other type of rule interaction (bleeding, counter-feeding, counter-bleeding). A difference must therefore be made between serial computation (GP computation is serial in the sense that constraints may apply to the same string several times) and serialism (there is no extrinsic or logical ordering of instructions, i.e. classical extrinsic rule ordering). Also, there is no ranking or prominence relationship among constraints: all instructions are equally important. I am not aware of discussion in the GP literature of what happens when different constraints conflict, or whether such a situation is provided for at all.

Declarative Phonology is directly affiliated with HPSG on the syntactic side and therefore does not stand on generative grounds: under the pressure of overgeneration, HPSG has taken the radical step to eliminate computation altogether. The result are so-called monostratal representations, which are fully informed of morpho-syntactic, semantic and phonological information that is available at any point in the derivation (talking about a derivation is actually improper because monostratalism rejects the existence of distinct underlying and surface representations). The major issue that HPSG has with the generative approach is thus about modularity: there are simply no modules in the HPSG landscape, which is a fully scrambled everything-is-one environment.

Regarding OT, it is to be noted that the juxtaposition of serial and parallel computation was really a matter of indecision in the early days: Prince & Smolensky (1993) consider so-called Harmonic Serialism all through their manuscript: Harmonic Serialism works like regular OT, only that the candidate set produced by GEN is much more local, and the winner of the strictly parallel evaluation procedure is fed back into GEN. This procedure is repeated until no harmonic improvement can be achieved anymore. Hence Prince & Smolensky (1993:6) write that "[d]efinitive adjudication between parallel and serial conceptions, not to mention hybrids of various kinds, is a challenge of considerable subtlety, as indeed the debate

over the necessity of serial Move-α illustrates plentifully […], and the matter can be sensibly addressed only after much well-founded analytical work and theoretical exploration".

A more direct application of connectionism to phonology, John Goldsmith's Harmonic Phonology (Goldsmith 1992, 1993, Larson 1992), also follows this track: the successive application of rules progressively increases harmony, and the well-formedness of representations is measured in gradient, rather than in categorical terms. Finally, Harmonic Grammar (Legendre et al. 1990, Smolensky & Legendre 2006) is also a direct application of connectionism which is most closely related to the interests of Cognitive Science: it is based on so-called weighted constraints, which like Goldsmith's gradual well-formedness are a direct transcription of the central connectionist notion of connection weight (and the activation level of neurons which defines their output). In this perspective, the relationship between constraints is one of lesser or greater prominence, rather than of strict dominance: less important constraints can league together and outrank a more important constraint on account of their cumulated weight.

8.3. How computation works: rules vs. constraints, and eventual hybrid solutions

Serialism, which stood unchallenged and without alternative since the 50s, has thus first fallen into disgrace for reasons that are not clear and have not been made explicit, and was then replaced by parallel computation when this option was available as an import from connectionism (more on this import in the following section).

Parallel computation has an important corollary: it can only be done on the basis of constraints. We have seen in section 8.1 that autosegmental representations and well-formedness conditions have initiated an independent bias in favour of constraints. Evicting rules as the basic carrier of computational instructions in favour of constraints was thus the result of the conjoint action of parallel computation (which in turn was the instrument of anti-derivationalism) and the increasing importance of well-formedness conditions.

Rules and constraints have different properties:[25] rules are made of a structural change (the part on the lefthand side of the slash in A → B / C__D) and a structural description (the part on the righthand side). The latter defines a string, CAD, that is in need of modification. Constraints do the same thing (if in different vocabulary, see below): they issue general requirements or prohibitions, in our case for example *CAD. However, they do not specify what should be done in order to satisfy the requirement or prohibition at hand. This is why constraint-based computation is said to be output-oriented: they specify how things should or must not look like, but do not give any indication how the desired or prohibited state of affaires should be achieved. In sum, thus, constraints divorce the structural description and the structural change of rewrite rules. Goldsmith & Laks (ms) point out that this split was already suggested by Sommerstein (1974).

Hale & Reiss (2008:195ff) discuss the formal difference between rules and constraints at length. They describe rules as a function that maps an input representation to an output representation (i.e. something is modified), while a constraint maps an input representation to the binary set "violation" or "no violation", i.e. without modifying anything. Eventual modifications of the input are operated elsewhere. Hale & Reiss (1998:196) build on this difference in order to make a point against constraints from the logical and cognitive point of view: a grammar ought not to contain explicit statements against monsters (they use the NOBANANA example in order to show that there is no point in explicitly excluding real bananas from UG by an explicit statement therein). Constraints, however, only inform some

[25] $Mohanan, 2000 #3642£ (2000:146) argues that they are not because one can always be translated into the other. The point, however, is not whether the "correct result" can be achieved, but whether *the way* in which this result is achieved is different. See the informative discussion in Hale & Reiss (2008:195ff) on this issue.

other part of the grammar that a given representation is ill-formed, and there is an infinite number of ways in which a representation can be ill-formed. Hence explicit statements (other constraints of the NOBANANA kind) are needed in order to rule out the monsters.

Another important difference between constraints and rules is the vocabulary in which they are stated: wile rules can only refer to the specific vocabulary items that phonology is made of (features or other items of autosegmental representations), constraints are made of prose statements and can express anything that prose can express (including very broad instructions such as "be lazy!", which is the formulation of the constraint LAZY that Kirchner 1998 believes is the motor for lenition; *STRUCTURE is another case in point). This loss of reference to a specific phonological vocabulary is meaningful in terms of Cognitive Science: we will see below that while so-called domain-specificity is a defining property of cognitive modules (which thus operate over a specific and proprietary vocabulary), it is denied by connectionism, where computation is content-free.

The fact that the formulation of constraints is not constrained in any way is a well-known and oft-mentioned property of this carrier of computational instruction, which to date however has failed to produce any effect: as far as I can see, no attempt is made in order to define what an (im-)possible formulation of constraints looks like.[26]

Obviously, the freedom to express anything that prose can express dramatically increases overgeneration, which takes us back to the post-SPE debate. OT has often been charged with computational irresponsibility: in principle GEN produces an infinite set of candidates that cannot even be stored, let alone computed, and the number of grammars that a set of, say, two- or three hundred universal constraints (which is an extremely conservative estimate) produces given free ranking is astronomical (even when logically impossible rankings and those that produce identical patterns are counted out). Factorial typology is commonly advertised as a trump for modelling dialectal variation, but the hundreds, thousands or more systems that are produced without empirical echo are usually not mentioned. Also, the concern for limiting the number of constraints that was present in early OT has more or less disappeared from the agenda: new constraints are proposed every day, and hardly anybody today is able to establish a comprehensive inventory.

From the beginning, the answer of OT to criticisms regarding computational complexity (and overgeneration) was to call on the competence-performance distinction (but recall from section 8.1 that it is precisely performance-based arguments that are used by connectionism in order to promote parallel, i.e. brain-style computation):

> "It is not incumbent upon a grammar to compute, as Chomsky has emphasized repeatedly over the years. A grammar is a function that assigns structural descriptions to sentences; what matters formally is that the function is well-defined. […] Grammatical theorists are free to contemplate any kind of formal device in pursuit of these goals; indeed, they *must* allow themselves to range freely if there is to be any hope of discovering decent theories. Concomitantly, one is not free to impose arbitrary additional metaconstraints (*e.g.* computational plausibility) which could conflict with the well-defined basic goals of the enterprise." Prince & Smolensky (1993:215f, emphasis in original)

Since its inception, generative grammar indeed followed this line of thinking: Chomsky has always argued that competence is not about implementation, and that implementational arguments have no bearing on the properties of the model of competence that linguists build. The minimalist programme (Chomsky 1995, 2000 et passim), however, clearly suspends with this kind of reasoning: grammar must respond to implementational requirements. The whole point of the minimalist approach is to make grammar evolve in response to extra-grammatical factors. Phase Theory for example cuts the computation of a full sentence into independent

[26] The only exception that I am aware of is Oostendorp & Weijer's (2005) attempt to define what they call a universe of discourse for the expression of instructions in OT.

pieces for reasons of computational economy regarding the limited availability of active memory, a costly cognitive resource (e.g. Chomsky 2000:101, 2001:15).

On the other hand, the greater generality of constraints with respect to rules and their dissociation from a specifically phonological vocabulary has crucially contributed to the success of OT and parallel computation. de Lacy (2007b:14ff) for example recapitulates the advantages of parallel computation over serialism: ordering paradoxes, global conditions, conspiracy (the latter notion is known at least since Kisseberth 1970). The question, then, is whether the generalizing remedy is not worse than the original disease.

A worrisome property of the OT literature and of the self-understanding of OT is the constant attempt to assess a tabula-rasa interpretation of the year 1993: in his introductory chapter to the Handbook of Phonology that he has edited, de Lacy (2007b) for example is much concerned with showing in which way OT is different from anything that phonology has produced before 1993 in general, and with making OT antithetical with respect to SPE in particular (de Lacy 2007b:13 opposes OT to "the dominant theories before OT - SPE and its successors"). The tabula-rasa claim may help to assess a new theory, to raise funds and to gain academic positions, as it did in the 60s when generative phonology rolled over the structuralist establishment – but it is as wrong now as it was then. Even a cursory look through Anderson's structure-process prism is enough to be convinced: OT takes phonology way back to the 60s when representations did not exist and computation was king (more on this in section 8.5). Hulst & Ritter (2000) describe the SPE-heritage of OT in greater detail. Also, it was shown in sections 8.1 and 8.2 that OT roots in autosegmentalism-promoted well-formedness conditions. Prince & Smolensky's (1993) original manuscript does not really license the picture that present-day OT-based story telling tries to establish.

This is not to say, of course, that there is nothing new in OT: our conception of computation has profoundly changed since OT and other theories have introduced constraints. The merit of OT is the application of parallel computation to phonology, as well as the promotion of the specific view whereby constraints are ranked and violable.

In this context, it is interesting to compare the evolution of computation in syntax and phonology: as Hale & Reiss (2008:202ff) point out, the Pisa turn syntax and the movement against extrinsically ordered rules in phonology were not only abolishing the same kind of serialism in both areas, and were not only doing away with rules. They also replaced rule-based serialism by free generation-cum-filters. In GB syntax, morphemes could "freely" be concatenated, and Move-α could "freely" apply. Ill-formed results were then filtered out by global constraints (on locality, case etc.). The minimalist version of this conception of computation focuses on the interfaces, PF and LF, which impose conditions that make the derivation either crash or converge. Constraint-based computation in phonology follows the same track: in OT GEN does the free generation, and constraints filter out the optimal candidate. Unlike in syntax where filters are equal-righted and inviolable, however, constraints are ranked and violable in OT.

Given this overall picture, an interesting question is certainly whether the correct solution for computation is only binary, i.e. either rule-only or constraints-only. Calabrese (2005) for example holds the principled position that a sound theory of phonology must have both serially ordered rules and constraints: while the former are instructions to create a given configuration, the latter specify which configurations must be avoided (constraints apply only to Kisseberth's 1970 conspiracies). This is indeed the naturally-grown state of the art of the 80s (with enlarged competences for well-formedness constraints, which may also concern purely melodic configurations) where well-formedness constraints (i.e. output filters), a direct consequence of (autosegmental) representations, cohabitated with traditional rule-based computation (see sections 7.2 and 8.1). Hale & Reiss (2008:209ff) offer instructive discussion

of the rules-cum-constraints option (concluding that both hybrid and constraint-only systems are inaccurate: computation must be purely rule-based in their opinion).

8.4. OT and its connectionist endowment

It was mentioned in section 2 that generative grammar is the linguistic outgrowth of the cognitive revolution of the 50s, which was anti-behaviourist, anti-empiricist and dualistic in nature. Connectionism is the modern version of the exact reverse intellectual position. Prince & Smolensky (1993:217ff) provide a good discussion of the relationship of OT with connectionism. They argue for cherry-picking: while parallel computation is taken over, other major connectionist tenets are rejected. That is, Prince & Smolensky do not accept the connectionist anti-symbolic stance according to which the only relevant level where decisions are made is the neuronal level: neurons only react on activation levels, hence cannot parse, or distinguish between, symbolic objects (e.g. Fodor & Pylyshyn 1988, Dinsmore 1992).

Just like the standard theory of Cognitive Science, OT recognizes a symbolic level of representation. The place for connectionist non-symbolic computation, then, is an intermediate level between the symbolic level and the physiologically neural functioning of the brain. This conciliatory position that rejects reductionism (the denial of the mind as an independent level of analysis) where the connectionist level mediates between the mind and the brain is defended by Paul Smolensky since his earliest work (Smolensky 1986, 1987, 1988, 1991) and up to the present day (Smolensky & Legendre 2006).

Also, Prince & Smolensky reject the neo-behaviourist take of connectionism regarding acquisition according to which "knowledge of language can be empirically acquired through statistical induction from training data" (Prince & Smolensky 1993:217).

Although Prince & Smolensky (1993) do not mention this issue, OT obviously also rejects the connectionist claim that there is no difference between computation and storage. In the connectionist perspective, the "experience" of a neural network – the equivalent notion of memory – is acquired when the patterns of connectivity change: neurons may develop new connections (synapses), may lose old connections, or modify the strength (weight) of existing connections (the two former are often viewed as a special case of the latter). The computational units themselves have no variable behaviour that contributes to the properties of the whole, which is exclusively determined by the connective network (see e.g. Stillings et al. 1995:114ff on connectionist models of memory).

All linguistic theories since Antiquity of course rely on the assumption that there is a lexicon that exists independently of grammatical activity which transforms lexically stored objects into actual speech. The linguistic mirror of the connectionist non-separation of storage and computation is so-called "Cognitive" Grammar,[27] which was founded by Langacker (1987) und is overtly empiricist, (neo-)behaviourist and anti-generative (see Taylor 2002): Langacker (1987 Vol.1:42) talks about the "rule/list fallacy". The phonological offspring of this line of thought is represented by exemplar- and usage-based approaches in general, and by Joan Bybee in particular, who writes that "[l]inguistic regularities are not expressed as cognitive entities or operations that are independent of the forms to which they apply, but rather as schemas or organisational patterns that emerge from the way that forms are associated with one another in a vast complex network of phonological, semantic, and sequential relations." Bybee (2001:20f).

Finally, another important connectionist headline is the aforementioned Parallel Distributed Processing (PDP), which contrasts with the regular assumption that computation

[27] I use quotation marks in order to refer to this framework because its name fallaciously suggests that it has a copyright on cognitive aspects of grammar, and that anything which is non-Langackerian must be non-cognitive.

is serial. While on the count of standard Cognitive Science the output of one computation is the input to another, several computations take place in parallel in the connectionist perspective. Also, the units that carry out computation – neurons, or clusters thereof – are not specialised for any particular computational task, or for a particular input material (computation is non-symbolic). Rather, neurons are all-purpose computational units that are able to perform any computation on the grounds of any type of information submitted. This is why connectionist computation is called *distributed*.

A corollary of distributed computation is the claim that there are no stable computational units made of clusters of neurons that can be delineated in the brain: computation is opportunistic and does not need any specialisation of its support units, the neurons. The modular theory of the mind (Fodor 1983), which incarnates as the inverted T model in the generative architecture of grammar, works with the exact reverse assumption: there are stable, genetically endowed, content-sensitive (i.e. symbolic) computational units that are devised to a very narrow and specific function, which can only work with a particular type of input vocabulary, and can do nothing else than what they have been designed for.

In sum, the connectionist perspective may be characterized by the fact that computation is content-free: all other properties follow from this assumption. That is, the mind does not know what it is doing when computation takes place: computation is only general-purpose, that is non-specialised for any task or function (i.e. the reverse of the modular perspective); it works without reference to any symbolic code, which would make the operations specific to a particular domain or content since symbols are symbols of something, and may be opposed to symbols of a different kind.

The question is whether the cherry-picking of items in this densely interrelated network of assumptions that Prince & Smolensky (1993) propose is viable: the genetic code of OT rejects basically all tenets of connectionism save one, parallel computation. Parallel computation is represented by the two Ps in PDP (Parallel Distributed Processing), but Prince & Smolensky do not address the question of the D, which is anti-modular. We will see in section 8.6 below that the D appears to be a direct consequence of parallel computation: it is constantly working on OT practice (if without explicit discussion) and has induced what I call the scrambling tropism, i.e. the creeping dissolution of modular contours. The same holds true for content-free computation, which has led to make representations irrelevant and interchangeable, and finally to dissolve them in computation (representations are "emergent", rather than given). This computation-tropism is discussed in the following section.

The conclusion, then, is that parallel computation has probably entered the generative paradigm with some more empiricist luggage, and the question for further development is whether a theory can be designed that holds up the rationalist and anti-empiricist core of generative grammar while implementing constraint-based and parallel computation.

8.5. Grammar reduces to computation: representations are demoted to decoration

Knowing about the connectionist roots of OT helps to understand the extreme computational orientation that phonology has taken under its lead since 1993. It is sometimes rightly recalled that OT is a theory of constraint interaction, not of constraints. This means that OT does not supply any substance itself: there are genuine vocabulary items in structuralism (phonemes), SPE (segments) and autosegmental theory (autosegmental structure), but there are no OT-specific representational items. OT uses whatever representational material comes the way, and may well produce the same result with entirely different (and incompatible) vocabulary.[28]

[28] For example, Lombardi (2001:3) writes with respect to melodic representation: "the tenets of OT, regarding constraint violability and ranking, make no particular claims about phonological representations. We could, for example, do OT with any kind of feature theory: SPE feature bundles or feature geometric

It is difficult not to establish a direct relationship between the fact that OT is a purely computational theory where representations make no sovereign contribution to the definition of grammaticality (which is decided by constraint interaction alone) and its content-free connectionist prototype.

The Handbook of Phonology that Paul de Lacy (2007a) has edited in general, and de Lacy's (2007b) introduction to the volume in particular, document the global trend from representation to computation in much detail. It is hard to find a thematic chapter of the book that does not insist on this evolution. A good example is John Harris's contribution on representations. The author, who is not exactly known for anti-representational positions, defends the categorical, i.e. non-gradient character of representations. In the end, however, the reader learns why it is that representations can be non-gradient at all: because they do not exist. The world, including grammar, is definitely gradient and computational: categorical objects can exist at best as a product of gradient computation (structure is emergent, rather than given: "categorical behaviour is emergent rather than an inherent property of these descriptors" Harris' conclusion p.137). Rather than about representations, Harris thus ends up talking about "descriptors" since this is the only role that they can play in a world where all decisions are made by constraint interaction (and by nothing else): representations "still have an important heuristic value as descriptors to be used in the building and experimental testing of models of phonological grammar" (p.137). In other words, representations are decoration: they may help the linguist to picture the result of constraint interaction – but they have no impact on grammar at all.

This is indeed the conclusion that follows from the OT-tenet that the *only* means to determine grammaticality is constraint interaction. Hence whatever items of the representational furniture of the 80s are used, they are mere decoration that do not contribute any sovereign arbitral award, and do not have anything to say regarding grammaticality (a structure with line-crossing for example may be the optimal candidate if all other candidates violate higher-ranked constraints).

Given Stephen Anderson's prism and the constant see-saw movement observed, a fair question to be asked in the face of the new OT-loop back into the 60s is *why* theories go down the representational or the computational road. When autosegmental representations were developed, the motivation was clear: gain of insight (tone spreading, the possibility to characterize the coda disjunction __{#,C} as a single phonological object etc.) and the promise of an efficient instrument against the plague of overgeneration. De Lacy (2007b) examines the question why OT has progressively replaced representations by computation. The answer is more or less that representations are eliminated because their function *can* be taken over by computation: "OT has allowed the burden of explanation to move from being almost exclusively representation-based to being substantially constraint-based" (p.24). The question why there *should* be such a movement (which, recall, is not what Prince & Smolensky 1993 have envisioned at all) is left unanswered: we do X not so much because we want to do it and have good reason to, but simply because we *can* do it. In other words, OT is a theory of computation and promotes what it is competent for without looking left or right. The progressive elimination of representations is thus but a side-effect of this computational tropism, which roots in the decision that the *only* thing that determines grammaticality is constraint interaction.

This proviso, however, does not follow from OT: OT is a theory of parallel computation that uses ranked and violable constraints. This does not lay any claim on how much of the explanative pie is computational: the view that 100% is has been installed without discussion or comment and today is part and parcel of OT. Yet it is just one possible attitude. Another

---

representations, privative or binary features, and so on."

view is expressed in a small but growing body of literature to which Marc van Oostendorp (Oostendorp 2002, 2003, 2005, 2006) has contributed a good deal, and which is condensed in Blaho et al. (eds.) (2007); Optimal Domains Theory is another case in point ($Cole, 1994 #3558£ 1994, $Cassimjee, 1998 #3557£ 1998). Blaho et al. (eds.) challenges Freedom of Analysis: you ought not to be free to do what you want with representations. In terms of the classical OT grammar, this means that there are restrictions on GEN, which produces only a subset of logically possible candidates.

The idea that OT is a complete theory of grammar has been tacitly entertained since its inception. The bare existence of versions of OT that place restrictions on GEN show that this view is overstated: OT is not a theory of grammar; it is a theory of a piece of grammar, computation. Anderson (1985) says that it takes more than just computation to make a grammar. The smallest common denominator of OT, then, is parallel computation that uses ranked and violable constraints. All the rest is free and a matter of choice of the analyst, who may or may not be a generativist, may or may not be a functionalist, may or may not assume a modular architecture, may or may not be representationally oriented, may or may not believe in the virtue of serial ordering of phonological (and/or grammatical) events (see below), may use this or that representational system, and so forth.

8.6. The scrambling tropism

Closely related to the tropism for computation is the pervasive reflex of OT to make distinct things indistinct; that is, to put them in the same constraint hierarchy, to intersperse them and to assess them in one go. Of course there are many different degrees of scrambling, but the tendency is clearly towards a single grammatical space where anti-derivationalism is enforced globally. The most visible result of this tropism is anti-cyclicity, i.e. the rejection of inside-out interpretation because of its serial character (e.g. Kager 1999:277). OT has produced a whole anti-cyclicity literature (which has an important intersection with the anti-opacity literature) that proposes alternative, strictly parallel ways of communicating with morpho-syntax: co-phonologies, indexed constraints, OO faithfulness and so-called interface constraints.

Other diagnostics for OT's misty (and largely unreflected) relationship with modularity are the following: 1) customary and uncontradicted violations of Indirect Reference (i.e. the prohibition to make reference to untranslated morpho-syntactic categories that was established by Prosodic Phonology). That is, ALIGN and WRAP constraints make constant reference to morpho-syntactic structure and labels; interface constraints such as FAITH-root and FAITH-affix make reference to designated morpho-syntactic categories (even reference to individual morphemes is not a problem, Anttila 2002 provides an overview) and thereby reincarnate the SPE-practice of supplementing rules with morphological diacritics. Also, 2) mapping (of morpho-syntactic into phonological prosodic categories) is done *in* the phonological constraint hierarchy (rather than outside of the phonology as was the case in Prosodic Phonology): ALIGN and WRAP are interspersed with purely phonological constraints. Finally, 3) constraints whose formulation combines phonological and morphological instructions are commonplace (see Yip 1998 on this issue).

Kingston (2007) discusses the scrambling tropism on the example of the abandon of the distinction between phonetic and phonological constraints. He also points out the causal relationship between the move from serial to parallel computation on the one hand and the everything-is-the-same programme on the other: "[r]eplacing serial derivation by parallel evaluation removes the barrier to phonetic constraints being interspersed among and interacting with phonological constraints. […] Future research will determine whether phonological and phonetic constraint evaluation are a single, integrated process, as advocated by Steriade and Flemming or instead sequential, as advocated by Zsiga" (Kingston 2007:432).

There is thus reason to believe that the commitment to non-serialism is the driving force behind OT's scrambling tropism. Despite this in-built tendency, however, adhering to modular-destructive indistinction in grammar is a personal choice of the analyst, not a fatality. Those who argue with anti-derivationalism in order to challenge cyclic derivation so to set up a single constraint hierarchy where phonetic, phonological, morphological and even syntactic constraints (sic, Russell 1999) are interleaved make a category mistake: derivation and computation is not the same thing (see Scheer forth a). OT is committed to parallel *computation*, and in generative grammar the unit where computation takes place is the module. Grammar is made of several modules, each with a distinct computation that works on distinct (i.e. domain specific) vocabulary. Hence nothing withstands a perspective where all linguistic computation is perfectly parallel, but distributed over distinct and serially ordered computational systems (modules). It is only when the non-derivational claim is laid on the entire grammar that the scrambling tropism appears.

That one can resist the scrambling tropism is shown by the fact that there are derivational versions of OT that allow for serial communication among modules while observing strictly parallel modular-internal computation: DOT (Rubach 1997 et passim) and Stratal OT (Kiparsky 2000, Bermúdez-Otero forth) are reincarnations of Lexical phonology in the new constraint-based environment.

Beyond the dilution of modular contours in one single constraint chamber, another target of the scrambling tropism is the lexicon. Richness of the Base prohibits the introduction of any distinction in the lexicon that goes beyond melodic contrast (this is one of a number of readings, the literature offers various interpretations) because GEN must not be marshalled in any way when creating candidate-variation. Hence nothing can be hard-wired in the lexicon, and somehow somebody has decided that melodic contrast (but nothing else) is an exception to this rule: this is why input forms in OT typically look like underlying forms in SPE: they are made of a linear chain of segments. In practice, this means that the progressive integration of constituent structure into the lexicon that was undertaken in 80s (see section 7.4) is ruled out, and the corresponding labour is transferred to the constraint chamber. But even more generally there is a trend in OT to introduce more and more unpredictable information by constraint. This movement is documented for example by de Lacy (2007b:19f); it contributes to the computationalisation of grammar, and is also a step in the direction of the connectionist indistinction between computation and storage (whose linguistic expression is Langacker's 1987 Vol.1:42 and Bybee's 2001:20f "rule/list fallacy").

Finally, two other types of static information are also expressed by constraints, rather than by some hard-wired (lexical) recording: parameters and inventories. The transformation of the lexical recording of language-specific parameters into a variable constraint ranking (children do not set parameters on-off, but rank relevant constraints) is one of the most basic ambitions of OT, and also one that is generally held to be an area where OT is successful: cross-linguistic variation is described in terms of different rankings of the same set of constraints (factorial typology).

Regarding inventories (see also section 3.5), the idea is to kill two birds with one stone: both inventory properties and some of the language-specific processing are derived from the same source, a set of constraints. Steriade (2007) provides a well-informed overview of this project, which is also developed in Government Phonology since the mid-90s (where the instrument are so-called Licensing Constraints, see Charette & Göksel 1994, 1996, Kaye 2001). In my opinion, the computationalisation of parameters and of inventory information has a different quality when compared to other aspects of the trend towards scrambling: it does not destroy any distinction that is central to generative thinking (as is the case with modular contours), nor does it undo analytic advances (as is the case with the prohibition of constituent structure in the lexicon); rather, it gives a different expression of something that

must be stated somehow anyway (parameters), or provides a new analytic perspective for the (belated) re-integration of the heart of structuralist thinking (inventories, i.e. the systemic pressure that is also at the origin of phonological processes, see section 3.5).

In conclusion, it appears that all letters in PDP (Parallel Distributed Processing) make sense and are interrelated. This casts doubt on the viability of cherry-picking of connectionist tenets: recall that Prince & Smolensky (1993) import parallel computation into OT, but do not mention its distributed aspect. It looks like the distributed character of computation, i.e. the connectionist everything-is-the-same programme, is a direct (ineluctable?) consequence of parallel computation.

9. Counterrevolution #2: neo-serialism and representations

9.1. Serialism *within* phonological computation: OT-CC

The latest evolution within OT has initiated the second counterrevolution: John McCarthy (2007), who was at the forefront of the anti-derivational movement in the 90s and 00s, has made a radical about-turn and now promotes serialism. Indeed, OT-CC has an entirely different quality than OTed versions of Lexical Phonology (DOT, Stratal OT) where the relationship between modules is serial, but phonological computation itself strictly parallel: OT-CC holds that phonological computation itself is done on the basis of serially ordered instructions. This does not mean, however, that parallel computation is abandoned, or that McCarthy goes back to ordered rules. Rather, OT-CC is a version of Harmonic Serialism (see section 8.2, and actually is more and more referred to under this name): the overall computation of an input is cut into a series of different evaluations where the output of EVAL is fed back into GEN, which adds harmonically improved candidates on each round. Candidate chains thus reflect the chronology of computational events (like intermediate forms in a serial SPE-derivation). EVAL is strictly parallel (and invariable) itself; it evaluates candidate chains, rather than single candidates, and so-called PREC constraints, a new constraint family which is actual the locus of ordered instructions, assign violation marks on the grounds of the comparison of different members of the candidate chain (PREC constraints react on the order of faithfulness violations).

There is an obvious parallel to computation as conceived of in Government Phonology (see section 8.2) where input strings are also modified by a fixed set of instructions (which however are unordered and unranked) in iterative fashion: the output of computation also loops back in order to become the new input, and the derivation also ends when the string cannot be further modified ("improved").

OT-CC is the capitulation of anti-serialism in the face of opacity: John McCarthy has tried hard to find a solution for opacity within parallelism (Sympathy Theory, Comparative Markedness), but was not any more successful than the half dozen of other opacity-killers that can be found in the literature (e.g. targeted constraints, enriched inputs, F&F conjunction, M&F conjunction, turbidity). McCarthy has studied one particular case of opacity (Bedouin Arabic) in depth to make sure that there is no lexical, morphological or other parasitic conditioning, and also that the processes at hand are really synchronically active and productive.

In this context, we may anticipate on the inventory of non-OT theories that have continued to exist, or have emerged, since 1993: one of them continues the SPE-tradition of extrinsically ordered rules (e.g. Vaux 2003, Calabrese 2005, Halle & Matushansky 2006, Raimy & Cairns (ed.), Hale & Reiss 2008) and thus shares with OT-CC the original tenet of generative grammar that (phonological) computation is serial and requires an extrinsically ordered set of instructions.

9.2. Representations: you cannot compute nothing

The other strand of the counterrevolution are representations that are not just mere decoration. Meaningful representations contribute a sovereign and unoutrankable arbitral award. That is, their verdict is absolute and not subject to any further evaluation: an ill-formed representation is ill formed no matter what and independently of constraint interaction. An ill-formed structure where, say, association lines cross is out; it cannot turn out to be the optimal candidate because all other candidates violate some higher ranked constraint. An ill-formed representation is either repaired, or the derivation crashes. In other words, representations that deserve their name break with the OT-baseline according to which the only way to determine (relative) grammaticality is constraint interaction.

Another property of meaningful representations is that they are not necessarily the result of computation: structure may be given, and neither its genesis nor its arbitral award has to be "emergent". Representations of this kind are also meaningful because they are not interchangeable. It was mentioned in section 8.5 that OT is a theory of computation and therefore has no opinion on representations: it has not developed any and can (actually does) work with any furniture of the 80s, including competing and incompatible structure (see note 28). Constraint interaction will always squeeze out the right result. The effect is that representations first become decoration, and then completely disappear.[29] Since representations are irrelevant, they have stopped being developed: there is no discussion regarding the comparative merits of competing representational theories ("I do this analysis using morae, but if you prefer x-slots that will do as well, you will just have to change the constraints a little").

Theories that develop meaningful representations include Government Phonology (Kaye et al. 1990, Charette 1991, Harris 1994, 1997, Kaye 2005, Pöchtrager 2006), Dependency Phonology (Hulst & Ritter 1999a, Botma 2004, Hulst 2005), Substance-Free Phonology (Hale & Reiss 2008, Blaho 2008) and the aforementioned rule-based approaches.

Especially Government Phonology is often regarded as a representation-oriented theory. It has continuously worked on the development of representations since the mid-80s, namely in the area of syllable structure. The genuine idea defended is that constituent structure is lateral, rather than arboreal: in Standard Government Phonology (see the references in the preceding paragraph), whether or not a consonant can be a coda depends on the availability of licensing from a following onset. That is, r in Vr.tV can be a coda, but word-final consonants cannot be codas since there is no following onset: therefore r in Vr# is an onset.

So-called CVCV (or strict CV, Lowenstamm 1996, Scheer 2004, Szigetvári 2002, Szigetvári & Scheer 2005, Cyran 2003), a development of Standard Government Phonology, takes the lateral idea to its logical end: instead of tolerating the cohabitation of arboreal and lateral structure, constituent structure boils down to a strict sequence of non-branching onsets and non-branching nuclei (no branching constituents, no rhymes, no codas). In this environment, the coda behaviour of a consonant is a consequence of the fact that it is followed by an empty nucleus (while consonants that show onset behaviour are followed by filled nuclei), which is unable to license its onset.

The flat, i.e. non-arboreal constituent structure that is the result of strict CV offers an in-built explanation why there is no recursion in phonology: the absence of arboreal structure witnesses the fact that phonology has no access to Merge (or some equivalent tree-building

[29] An eloquent example is the "representation" of melody that is currently practised in OT today: binary features are about as sophisticated as they were in SPE (i.e. an amorphous, unarticulated set), and constraints that manipulate melody boil down to "star [feature]" (e.g. *dorsal etc.). Clements (2001) and Hall (2001) bemoan this evolution.

device), which however is supposed by recursion (Scheer 2004:xlix ff, Neeleman & van de Koot 2006).

Finally, it was already mentioned that there is also a representation-rehabilitating strand within OT that was condensed in Blaho *et al.* (2007); Optimal Domains Theory is also a case in point (Cassimjee & Kisseberth 1998). Oostendorp & Weijer (2005) for example make the argument based on vocabulary: OT needs a universe of discourse, you cannot compute nothing (or interchangeable phantoms). Regarding melodic representation, Hall (2001) and Clements (2001) may certainly be counted as expressions of this line of thought as well.

10. Conclusion

The chapter is already much too long despite the fact that many noteworthy aspects of more and less recent generative development in phonology (such as for example the work of William Labov on social variation and experimental phonology, Moreton 2008) could not even be mentioned. This is probably the price to pay for an overview that spans some 40 or 50 years. Below I try to condense what I take to be the important issues of the field, as well as its true advances that probably still play a role in 50 years. Table (6) first looks at the former category.

(6) important issues
- a. the system
  systemic pressure on phonological events and their implementation into a formal system.
- b. type of relatedness of alternating forms
  1. computational: by a phonological process (common underlying form) or by a non-phonological process (allomorphy)
  2. non-computational: two separate lexical entries
- c. melodic primes
  binary, privative or a blend thereof?
- d. phonological computation
  1. type of computational instruction
     rules (specifying the modification of the input and its triggering conditions) or constraints (general requirement or prohibition with no modificational instruction)?
  2. serial, parallel or a blend thereof? If serial, what kind of serialism: extrinsically ordered rules, harmonic serialism? If parallel, what kind of parallelism: hard or violable constraints, if violable, dominance expressed by weight or by rank?

None of these issues is settled, and I doubt that (6)b will ever be. (6)a is not about whether or not phonological processes are under the spell of systemic pressure: hardly anybody doubts that they are; the question is just how this could be expressed in a formal system.

The list under (7) below is what I submit may be booked as true achievements and advances of generative phonology.

(7) achievements and advances
   a. cyclic derivation
   b. modular architecture of grammar
   c. interactionism
   d. autosegmental representations
      i.e. which can be ill-formed, issue a sovereign arbitral award and are not necessarily the result of computation
   e. privative melodic primes
   f. constraints and parallel computation

This list may be divided into two groups. On the one hand cyclic derivation and the modular architecture of grammar are the generative baseline: they were present upon inception of the generative enterprise, and they concern architectural properties of the grammar that lie beyond narrowly phonological concerns (but recall that while modularity is shared with structuralism, cyclic derivation is a genuinely generative discovery). It is difficult to see how anything could be called generative that does not endorse them. It was shown that OT in its present state freely violates modularity, and that this appears to be the application of the D (distributed: the scrambling tropism) of PDP, which itself is a direct consequence of the P (parallel computation). The future will have to show whether Prince & Smolensky's (1993) cherry-picking (we are true generativists and hence reject the empiricist freight of connectionism, except for parallel computation) can be brought home.

The other group of the list under (7) encompasses items that have appeared in the course of the development of generative theory. Two of them, (7)e,f, also appear in the list of issues, but these do not concern their existence, which is not really doubted. Rather, the question is how much of melodic representation is privative (maybe all), and to which extent computation is parallel and based on constraints. While constraints are used by all theories in one form of another (recall that well-formedness conditions of the 80s such as the OCP are a form of constraints), it is true that strictly rule-based approaches may reject the idea of parallel computation altogether. In any event, the parallel idea is an entirely new perspective on computation that has injected fresh air into the settled waters of the serial model of Standard Cognitive Science in general, and of its linguistic outgrowth in particular. Note that unlike modularity, serialism is not an essential of the rational conception of the mind: it is simply the default of the 50s that was introduced by the Turing/von Neumann machine. All other things being equal (and this is the question raised by cherry-picking), computation could be parallel as well. Generative syntax has shown that non-serialism is possible within the generative realm (note, however, that the constraints on syntactic well-formedness are hard, rather than ranked and violable).

Finally, the two other items of the second group, (7)c,d, are the major scientific advances of generative phonology in my opinion. Again, one concerns an architectural property: interactionism was invented by phonologists but has made a brilliant career in syntax: it is the spine of current minimalist thinking. The other item is the face of modern phonology: autosegmental representations are certainly the smallest common denominator of all currently practiced theories, even if they are demoted to pure decoration in (some versions of) OT.

This record appears to be quite positive and encouraging. It needs to be tuned down, though, when the fact is considered that generative phonology seems to go in circles with respect to some fundamental issues, and spends quite some time reinventing the wheel. The systemic question was thrown over board in the 60s without argument and was reintegrated into the research agenda only very recently. But worse is the situation regarding Anderson's (1985) concern for a balance between representation and computation. Generative phonology

has made a complete loop (computational SPE → super-representational 80s → super-computational OT) and is maybe engaging into another round (cf. the still timid return of representations within and outside of OT). The parallel with syntax in this area is quite striking: like phonology, syntax was anti-lexicalist and very computationally oriented in the 60s, explored the lexicalist and representational track in the 70s and 80s, but has gone back to extreme anti-lexicalism and proceduralisation since the turn of the minimalist programme.

It would be nice if phonology (and syntax for that matter), like adult science, could be said to follow a linear trajectory from less to more knowledge, in a cumulative movement that builds on and learns from the experience and errors of the past. This has certainly not been the case thus far (regular assertions of the contrary from OT quarters notwithstanding), but maybe the see-saw movement can be broken this time: we do not know where exactly the red line runs, but there is good reason to believe that both representations and computation are needed in order to make a grammar.

References

Anderson, John & Colin Ewen 1987. Principles of Dependency Phonology. Cambridge: Cambridge University Press.

Anderson, John, Colin Ewen & Jørgen Staun 1985. Phonological structure: segmental, suprasegmental and extrasegmental. Phonology Yearbook 2: 203-224.

Anderson, John & Charles Jones 1974. Three theses concerning phonological representations. Journal of Linguistics 10: 1-26.

Anderson, Stephen 1981. Why phonology Isn't "Natural". Linguistic Inquiry 12: 493-539.

Anderson, Stephen 1985. Phonology in the Twentieth Century. Chicago: University of Chicago Press.

Anderson, Stephen 2000. Reflections on `On the Phonetic Rules of Russian'. Folia Linguistica 34: 11-28.

Anttila, Arto 2002. Morphologically conditioned phonological alternations. Natural Language and Linguistic Theory 20: 1-42.

Archangeli, Diana 1988. Aspects of underspecification theory. Phonology 5: 183-208.

Aronoff, Mark 1976. Word Formation in Generative Grammar. Cambridge, Mass.: MIT Press.

Aronoff, Mark 1980. The Treatment of Juncture in American Linguistics. Juncture, edited by Mark Aronoff & Mary-Louise Kean, 29-36. Saratoga: Anma Libri.

Aronoff, Mark & S. Sridhar 1983. Morphological levels in English and Kannada, or Atarizing Reagan. Chicago Linguistics Society 19, Papers from the Parasession on the interplay of Phonology, Morphology and Syntax, edited by J. Richardson, M. Marks & A. Chukerman, 3-16. Chicago: Chicago Linguistics Society.

Aronoff, Mark & S. Sridhar 1987. Morphological levels in English and Kannada. Rules and the Lexicon, edited by Edmund Gussmann, 9-22. Lublin: Katolicki Universytet Lubelski.

Auroux, Sylvain, E.F.K. Koerner, Hans-Josef Niederehe & Kees Versteegh (eds.) 2006. History of the Language Sciences, vol.3 (HSK 18.3). Berlin: de Gruyter.

Bermúdez-Otero, Ricardo forth. Stratal Optimality Theory. Oxford: Oxford University Press.

Blaho, Sylvia 2008. The syntax of phonology. A radically substance-free approach. Ph.D dissertation, University of Tromsø.

Blaho, Sylvia, Patrick Bye & Martin Krämer (eds.) 2007. Freedom of analysis? Berlin: Mouton de Gruyter.

Bloch, Bernard 1948. A set of postulates for phonemic analysis. Language 24: 3-46.

Bloomfield, Leonard 1936. Language or ideas? Language 12: 89-95.

Bloomfield, Leonard 1939. Menomini morphophonemics. Travaux du Cercle Linguistique de Prague 8: 105-115.

Booij, Geert 1981. Rule ordering, rule application and the organization of grammars. Phonologica 1980, edited by Wolfgang Dressler, Oskar Pfeiffer & John Rennison, 45-56. Innsbruck: Institut für Sprachwissenschaft.

Booij, Geert 1997. Non-derivational phonology meets Lexical Phonology. Derivations and Constraints in Phonology, edited by Iggy Roca, 261-288. Oxford: Oxford University Press.

Botma, Bert 2004. Phonological Aspects of Nasality. An Element-Based Dependency Approach. Ph.D dissertation, University of Amsterdam.

Bromberger, Sylvain & Morris Halle 1989. Why Phonology Is Different. Linguistic Inquiry 20: 51-70.

Bybee, Joan 2001. Phonology and Language Use. Cambridge: Cambridge University Press.

Calabrese, Andrea 2005. Markedness and Economy in a Derivational Model of Phonology. Berlin: Mouton de Gruyter.

Campbell, Lyle 1981. Generative phonology vs. Finnish phonology: retrospect and prospect. Phonology in the 80's, edited by Didier Goyvaerts, 147-182. Ghent: Story-Scientia.

Carr, Philip 2000. Scientific Realism, Sociophonetic Variation, and Innate Endowments in Phonology. Phonological Knowledge. Conceptual and Empirical Issues, edited by Noel Burton-Roberts, Philip Carr & Gerard Docherty, 67-104. Oxford: Oxford University Press.

Carvalho, Joaquim Brandão de 2002. De la syllabation en termes de contours CV. Habilitation Thesis, Ecole des Hautes Etudes en Sciences Sociales, Paris.

Carvalho, Joaquim Brandão de 2004. L'analogie est-elle un fait fonctionnel ou grammatical ? Le cas de l'apophonie portugaise. Corpus 3: 101-123.

Chambers, J. K. 1973. Canadian Raising. Canadian Journal of Linguistics 18: 113-135.

Charette, Monik & Asli Göksel 1994. Vowel Harmony and Switching in Turkic languages. SOAS Working Papers in Linguistics and Phonetics. 4: 31-52. Also in Kardela, Henryk, Bogdan Szymanek (eds.), A Festschrift for Edmund Gussmann, 29-56. Lublin 1996: University Press of the Catholic University of Lublin.

Charette, Monik & Asli Göksel 1996. Licensing constraints and vowel harmony in Turkic languages. SOAS Working Papers in Linguistics and Phonetics 6: 1-25. Also in Cyran, Eugeniusz (ed), Structure and Interpretation. Studies in Phonology, 65-88. Lublin 1998: Folium.

Chomsky, Noam 1951. The Morphophonemics of Modern Hebrew. Masters thesis, Université de Pennsylvania.

Chomsky, Noam 1955-56. The Logical Structure of Linguistic Theory. New York: Published 1975, Plenum.

Chomsky, Noam 1959. Review of Skinner's Verbal Behavior. Language 35: 26-58.

Chomsky, Noam 1964. Current Issues in Linguistic Theory. The Hague: Mouton.

Chomsky, Noam 1965. Aspects of the Theory of Syntax. Cambridge, Mass.: MIT Press.

Chomsky, Noam 1966. Cartesian Linguistics. New York & London: Harper & Row.

Chomsky, Noam 1970. Remarks on nominalisation. Readings in English Transformational Grammar, edited by R.Jacobs & P.Rosenbaum, 184-221. Walthamm, Mass.: Ginn.

Chomsky, Noam 1973. Conditions on Transformations. A Festschrift for Morris Halle, edited by Stephen Anderson & Paul Kiparsky, 232-286. New York: Holt, Rinehart & Winston.

Chomsky, Noam 1981. Lectures on Government and Binding. 7th edition 1993, Dordrecht: Foris.

Chomsky, Noam 1986. Knowledge of language: its nature, origin, and use. New York:

Praeger.
Chomsky, Noam 1995. The Minimalist Program. Cambridge, MA: MIT Press.
Chomsky, Noam 2000. Minimalist inquiries: the framework. Step by Step. Essays on Minimalist Syntax in Honor of Howard Lasnik, edited by Roger Martin, David Michaels & Juan Uriagereka, 89-155. Cambridge, Mass.: MIT Press.
Chomsky, Noam 2001. Derivation by Phase. Ken Hale: A Life in Language, edited by Michael Kenstowicz, 1-52. Cambridge, Mass.: MIT Press.
Chomsky, Noam & Morris Halle 1968. The Sound Pattern of English. Cambridge, Mass.: MIT Press.
Chomsky, Noam, Morris Halle & Fred Lukoff 1956. On Accent and Juncture in English. For Roman Jakobson. Essays on the occasion of his sixtieth birthday, edited by Morris Halle, Horace Lunt, Hugh McLean & Cornelis van Schooneveld, 65-80. The Hague: Mouton.
Circle, Prague Linguistic 1931. Projet de terminologie phonologique standardisée. Travaux du Cercle Linguistique de Prague 4: 309-323.
Clements, George 1977. The autosegmental treatment of vowel harmony. Phonologica 1976, edited by Wolfgang Ulrich Dressler & Oskar Pfeiffer, 111-119. Innsbruck: Universität Innsbruck.
Clements, George 1985. The geometry of phonological features. Phonology 2: 225-252.
Clements, George 2000. In defense of serialism. The Linguistic Review 17: 181-197.
Clements, George 2001. Representational economy in constraint-based phonology. Distinctive Feature Theory, edited by Tracy Hall, 71-146. Berlin, New York: Mouton de Gruyter.
Clements, George 2002. Economie de traits et représentations lexicale. Paper presented at the 4è Journées Internationales du GDR Phonologies, Grenoble 6-8 June 2002.
Cole, Jennifer 1995. The Cylcle in Phonology. The Handbook of Phonological Theory, edited by John Goldsmith, 70-113. Oxford: Blackwell.
Cyran, Eugeniusz 2003. Complexity Scales and Licensing Strength in Phonology. Lublin: KUL.
de Lacy, Paul (ed.) 2007a. The Cambridge Handbook of Phonology. Cambridge: CUP.
de Lacy, Paul 2007b. Themes in Phonology. The Cambridge Handbook of Phonology, edited by Paul de Lacy, 5-30. Cambridge: CUP.
Dinnsen, Daniel 1980. Phonological rules and phonetic explanation. Journal of Linguistics 16: 171-191.
Dinsmore, John (ed.) 1992. The symbolic and connectionist paradigms: closing the gap. Hillsdale: Erlbaum.
Donegan, Patricia & David Stampe 1978. The Syllable in Phonological and Prosodic Structure. Syllables and Segments, edited by Alan Bell & Joan Bybee Hooper, 25-43. Amsterdam: North Holland.
Donegan, Patricia & David Stampe 1979. The study of natural phonology. Current approaches to phonological theory, edited by Daniel Dinnsen, 126-173. Bloomington: Indiana University Press.
Dresher, Elan 1981. Abstractness and Explanation in Phonology. Explanation in Linguistics. The logical problem of language acquisition, edited by Norbert Hornstein & David Lightfoot, 76-115. London, New York: Longman.
Dresher, Elan 2003. Contrast and asymmetries in inventories. Asymmetry in Grammar, Volume 2: Morphology, Phonology, Acquisition, edited by Anna-Maria di Sciullo, 239-257. Amsterdam: Benjamins.
Dresher, Elan 2003. The Contrastive Hierarchy in Phonology. Toronto Working Papers in Linguistics 20: 47-62.

Dressler, Wolfgang 1974. Diachronic Puzzles for Natural Phonology. Papers from the parasession on Natural Phonology, edited by A.Bruck, R.Fox & M.La Galy, 95-102. Chicago: Chicago Linguistic Society.

Dressler, Wolfgang 1984. Explaining Natural Phonology. Phonology Yearbook 1: 29-51.

Durand, Jacques 2006. La phonologie générative jusqu'en 1975. History of the Language Sciences, vol.3 (HSK 18.3), edited by Sylvain Auroux, E.F.K. Koerner, Hans-Josef Niederehe & Kees Versteegh, 2265-2270. Berlin: de Gruyter.

Dziubalska-Kołaczyk, Katarzyna 2002. Beats-and-binding phonology. Frankfurt am Main: Lang.

Encrevé, Pierre 1983. La liaison sans enchaînement. Actes de la Recherche en Sciences Sociales 46: 39-66.

Encrevé, Pierre 1988. La liaison avec et sans enchaînement: phonologie tridimensionnelle et usages du français. Paris: Seuil.

Encrevé, Pierre 1997. L'ancien et le nouveau. Quelques remarques sur la phonologie et son histoire. Langages 125: 100-123. An English translation has appeared as Encrevé, Pierre 2000. The Old and the New: Some Remarks on Phonology and its History. Folia Linguistica 34: 57-85.

Ewen, Colin & Harry van der Hulst 2001. The Phonological Structure of Words. Cambridge: Cambridge University Press.

Fodor, Jerry 1983. The modularity of the mind. Cambridge, Mass.: MIT-Bradford.

Fodor, Jerry & Zenon Pylyshyn 1988. Connectionism and Cognitive Architecture: A Critical Analysis. Cognition 28: 3-71.

Foley, James 1977. Foundations of theoretical phonology. Cambridge: CUP.

Fudge, Erik 1969. Syllables. Journal of Linguistics 5: 253-286.

Gardner, Howard 1985. The Mind's New Science: A History of the Cognitive Revolution. New York: Basic Books.

Giegerich, Heinz 1999. Lexical Strata in English. Cambridge: Cambridge University Press.

Goldsmith, John 1974. An autosegmental typology of tone: and how Japanese fits in. Proceedings of NELS 5, edited by Ellen Kaisse & J. Hankamer, 172-182. Cambridge, MA: Harvard University Linguistics Department.

Goldsmith, John 1976a. Autosegmental Phonology. Ph.D dissertation, MIT.

Goldsmith, John 1976b. An overview of autosegmental phonology. Linguistic Analysis 2: 23-68.

Goldsmith, John 1979. The aims of autosegmental phonology. Current Approaches to Phonological Theory, edited by Daniel Dinnsen, 202-222. Bloomington: Indiana University Press.

Goldsmith, John 1980. English as a tone language. Phonology in the 1980's, edited by Didier Goyvaerts, 287-308. Gent: Story-Scientia.

Goldsmith, John 1992. Local modeling in phonology. Connectionism: Theory and practice, edited by Steven Davis, 229-246. Oxford: OUP.

Goldsmith, John 1993. Harmonic Phonology. The last phonological rule, edited by John Goldsmith, 21-60. Chicago: University of Chicago Press.

Goldsmith, John 2005. Review of The Legacy of Zellig Harris: Language and Information into the 21st Century. Language 81: 719-736.

Goldsmith, John 2008. Generative Phonology in the late 40s. Phonology 25: XXX.

Goldsmith, John & Bernard Laks (eds.) 2000. The History of Phonology in the Twentieth Century. Folia Linguistica 34.1/2.

Goldsmith, John & Bernard Laks forth. Generative phonology: its origins, its principles, and its successors. The Cambridge History of Linguistics, edited by Linda Waugh & John Joseph. Cambridge: CUP.

Goyvaerts, Didier 1981. Introduction. Phonology in the 1980's, edited by Didier Goyvaerts, 1-26. Ghent: Story-Scientia.

Green, Anthony 2007. Phonology Limited. Ms., University of Surrey Roehampton. [ROA#605].

Gussmann, Edmund 1980. Studies in Abstract Phonology. Cambridge Mass.: MIT Press.

Gussmann, Edmund 2007. The Phonology of Polish. Oxford: Oxford University Press.

Gussmann, Edmund & Jonathan Kaye 1993. Polish notes from a Dubrovnik Café: I. The yers. SOAS Working Papers in Linguistics and Phonetics 3: 427-462.

Hale, Mark & Charles Reiss 2008. The Phonological Enterprise. Oxford: OUP.

Hall, Daniel 2007. The Role and Representation of Contrast in Phonological Theory. Ph.D, University of Toronto.

Hall, Tracy 2001. Introduction: Phonological representations and phonetic implementation of distinctive features. Distinctive Feature Theory, edited by Tracy Hall, 1-40. Berlin, New York: Mouton de Gruyter.

Hall, Tracy 2007. Segmental Features. The Cambridge Handbook of Phonology, edited by Paul de Lacy, 311-334. Cambridge: CUP.

Halle, Morris 1959. The Sound Pattern of Russian. The Hague: Mouton.

Halle, Morris 1978. Formal vs. functional considerations in phonology. Studies in the Linguistic Sciences 8: 123-134. Reprinted in Brogyányi, B. (ed.), Studies in diachronic, synchronic any typological linguistics: Festschrift for O. Szemerényi, 325-341. Amsterdam: Benjamis.

Halle, Morris 1998. The Stress of English Words 1968-1998. Linguistic Inquiry 29: 539-568.

Halle, Morris 2005. Palatalization/velar softening: what it is and what it tells us about the nature of language. Linguistic Inquiry 36: 23-41.

Halle, Morris & Alec Marantz 1993. Distributed Morphology and the Pieces of Inflection. The view from Building 20. Essays in Linguistics in Honor of Sylvain Bromberger, edited by Kenneth Hale & Samuel Keyser, 111-176. Cambridge, Mass: MIT Press.

Halle, Morris & Ora Matushansky 2006. The Morphophonology of Russian Adjectival Inflection. Linguistic Inquiry 37: 351-404.

Halle, Morris & Jean-Roger Vergnaud 1987. An Essay on Stress. Cambridge, Mass.: MIT Press.

Hargus, Sharon & Ellen Kaisse (eds.) 1993. Studies in Lexical Phonology. New York: Academic Press.

Harris, John 1994. English sound structure. Oxford: Blackwell.

Harris, John 2007. Representation. The Cambridge Handbook of Phonology, edited by Paul de Lacy, 119-138. Cambridge: CUP.

Harris, Zellig 1944. Simultaneous components in phonology. Language 20: 181-205.

Harris, Zellig 1951. Methods in Structural Linguistics. Edition 1960 entitled Structural Linguistics. Chicago & London: University of Chicago Press.

Hayes, Bruce 1984. The phonology of rhythm in English. Linguistic Inquiry 15: 33-74.

Hayes, Bruce 1989. Compensatory Lengthening in Moraic Phonology. Linguistic Inquiry 20: 253-306.

Hayes, Bruce 1995. On what to teach undergraduates: some changing orthodoxies in phonological theory. Linguistics in the Morning Calm 3, edited by Ik-Hwan Lee, 59-77. Seoul: Hanshin.

Hellberg, Staffan 1978. Unnatural phonology. Journal of Linguistics 14: 157-177.

Hill, Archibald 1954. Juncture and syllable division in Latin. Language 30: 439-447.

Hirschfeld, Lawrence & Susan Gelman 1994. Toward a topography of mind: an introduction to domain specificity. Mapping the mind: Domain specificity in cognition and culture, edited by Lawrence Hirschfeld & Susan Gelman, 3-35. Cambridge: CUP.

Hockett, Charles 1948a. A note on "structure". International Journal of American Linguistics 14: 269-271.
Hockett, Charles 1948b. Biophysics, linguistics, and the unity of science. American Scientist 36: 558-572.
Hockett, Charles 1955. A Manual of Phonology. Baltimore: Waverly Press.
Hooper, Joan 1972. The Syllable in Linguistic Theory. Language 48: 525-540.
Hooper, Joan 1975. The archi-segment in Natural Generative Phonology. Language 51: 536-560.
Hooper, Joan 1976. An Introduction to Natural Generative Phonology. New York: Academic Press.
Huddleston, Rodney 1972. The development of a non-process model in American structural linguistics. Lingua 30: 333-384.
Hulst, Harry van der 2005. Why phonology is the same. Organizing Grammar. Studies in Honor of Henk van Riemsdijk, edited by Hans Broekhuis, Norbert Corver, Riny Huybregts, Ursula Kleinhenz & Jan Koster, 252-262. Berlin: Mouton de Gruyter.
Hulst, Harry van der & Nancy Ritter 1999. Head-Driven Phonology. The syllable, Views and Facts, edited by Harry van der Hulst & Nancy Ritter, 113-167. Berlin, New York: de Gruyter.
Hulst, Harry van der & Nancy Ritter 2000. The SPE-heritage of Optimality Theory. The Linguistic Review 17: 259-289.
Hulst, Harry van der & Nancy Ritter 2003. Levels, constraints and heads. Asymmetry in Grammar, edited by Anna Maria Di Sciullo, 147-188. Amsterdam: Benjamins.
Hulst, Harry van der & Norval Smith (eds.) 1982. The structure of phonological representations. 2 Vols. Dordrecht: Foris.
Hurch, Bernhard & Richard Rhodes (eds.) 1996. Natural Phonology: The State of the Art. Berlin: Mouton de Gruyter.
Hyman, Larry 1985. A Theory of Phonological Weight. Dordrecht: Foris.
Idsardi, William 1992. The computation of prosody. Ph.D dissertation, MIT.
Jakobson, Roman 1939. Observations sur le classement phonologique des consonnes. Proceedings of the 3rd International Congress of Phonetic Sciences, Ghent, edited by Edgard Blancquaert & Willem Pée, 273-279. Ghent: Laboratory of Phonetics of the University of Ghent.
Jakobson, Roman, Gunnar Fant & Morris Halle 1952. Preliminaries to speech analysis. Cambridge, Mass.: MIT Press.
Jones, Daniel 1967. The phoneme: its nature and use. Cambridge: Heffer.
Joos, Martin 1942. A phonological dilemma in Canadian English. Language 18: 141-144.
Kager, René 1999. Optimality Theory. Cambridge: Cambridge University Press.
Kahn, Daniel 1976. Syllable-based generalizations in English phonology. Ph.D dissertation, MIT. Published by Garland Press, New York 1980.
Kaisse, Ellen & Patricia Shaw 1985. On the theory of Lexical Phonology. Phonology Yearbook 2: 1-30.
Katz, Jerrold 1964. Mentalism in linguistics. Language 40: 124-137.
Kaye, Jonathan 1990. What ever happened to dialect B ? Grammar in Progress: GLOW Essays for Henk van Riemsdijk, edited by Joan Mascaró & Marina Nespor, 259-263. Dordrecht: Foris.
Kaye, Jonathan 1992. On the interaction of theories of Lexical Phonology and theories of phonological phenomena. Phonologica 1988, edited by Uli Dressler, Hans Luschützky, Oskar Pfeiffer & John Rennison, 141-155. Cambridge: Cambridge University Press.
Kaye, Jonathan 1995. Derivations and Interfaces. Frontiers of Phonology, edited by Jacques

Durand & Francis Katamba, 289-332. London & New York: Longman. Also in SOAS Working Papers in Linguistics and Phonetics **3**, 1993, 90-126.

Kaye, Jonathan 2001. Working with licensing constraints. Constraints and Preferences, edited by Katarzyna Dziubalska-Kołaczyk, 251-268. Berlin & New York: Mouton de Gruyter.

Kaye, Jonathan 2005. "GP, I'll have to put your flat feet on the ground". Organizing Grammar. Studies in Honor of Henk van Riemsdijk, edited by Hans Broekhuis, Norbert Corver, Riny Huybregts, Ursula Kleinhenz & Jan Koster, 283-288. Berlin: Mouton de Gruyter.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud 1985. The internal structure of phonological representations: a theory of Charm and Government. Phonology Yearbook 2: 305-328.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud 1990. Constituent structure and government in phonology. Phonology 7: 193-231.

Kean, Mary-Louise 1974. The Strict Cycle in Phonology. Linguistic Inquiry 5: 179-203.

Kenstowicz, Michael & Charles Kisseberth 1977. Topics in Phonological Theory. New York: Academic Press.

Kenstowicz, Michael & Charles Kisseberth 1979. Generative Phonology. Description and Theory. San Diego: Academic Press.

Kingston, John 2007. he phonetics-phonology interface. The Cambridge Handbook of Phonology, edited by Paul De Lacy, 435-456. Cambridge: CUP.

Kiparsky, Paul 1968-1973. How abstract is phonology? Manuscript circulated since 1968 and published 1973 in: Three Dimensions of Linguistic Theory, edited by Osamu Fujimura, 5-56. Tokyo: TEC.

Kiparsky, Paul 1973. Abstractness, opacity and global rules. Three Dimensions in Phonological Theory, edited by Osamu Fujimura, 57-86. Tokyo: TEC.

Kiparsky, Paul 1974. On the Evaluation Measure. Papers from the parasession on Natural Phonology, edited by A.Bruck, R.Fox & M.La Galy, 328-337. Chicago: Chicago Linguistic Society.

Kiparsky, Paul 1982a. Lexical morphology and phonology. Linguistics in the morning calm, edited by In-Seok Yang, 3-91. Seoul: Hanshin.

Kiparsky, Paul 1982b. From Cyclic Phonology to Lexical Phonology. The structure of phonological representations I, edited by Harry van der Hulst & Norval Smith, 131-175. Dordrecht: Foris.

Kiparsky, Paul 1985. Some Consequences of Lexical Phonology. Phonology Yearbook 2: 85-138.

Kiparsky, Paul 1993. Blocking in nonderived environments. Studies in Lexical Phonology, edited by Sharon Hargus & Ellen Kaisse, 277-313. San Diego: Academic Press.

Kiparsky, Paul 2000. Opacity and cyclicity. The Linguistic Review 17: 351-365.

Kirchner, Robert 1998. An effort-based approach to consonant lenition. Ph.D dissertation, University of California at Los Angeles.

Kisseberth, Charles 1970. On the functional unity of phonological rules. Linguistic Inquiry 1: 291-306.

Koerner, E.F.K. 2002a. The Chomskyan 'Revolution' and its historiography. Toward a History of American Linguistics, edited by E.F.K. Koerner, 151-209. London: Routledge.

Koerner, E.F.K. 2002b. American structuralist linguistics and the "problem of meaning". Toward a History of American Linguistics, edited by E.F.K. Koerner, 75-104. London: Routledge.

Koerner, E.F.K. 2002c. Toward a History of American Linguistics. London: Routledge.

Koerner, E.F.K. 2003. Remarks on the origins of morphophonemics in American structuralist

linguistics. Language and Communication 23: 1-43.
Koerner, E.F.K. 2004a. Essays in the History of Linguistics. London: Benjamins.
Koerner, E.F.K. 2004b. On "influence" in linguistic historiography: morphohonemics in American structuralism. Essays in the History of Linguistics, edited by E.F.K. Koerner, 65-101. London: Benjamins.
Koutsoudas, Andreas 1980. The question of rule ordering: some common fallacies. Journal of Linguistics 16: 19-35.
Kuhn, Thomas 1962. The structure of scientific revolutions. Chicago: University of Chicago Press.
Kuryłowicz, Jerzy 1948. Le sens des mutations consonantiques. Lingua 1: 77-85.
Lacan, Jacques 1966. Ecrits. Paris: Seuil.
Laks, Bernard 2001. Un siècle de phonétique et de phonologie : quelques questions permanentes. Modèles Linguistiques 22: 75-103.
Laks, Bernard 2003. Phonologie Cognitive. Handout, Ecole thématique Phonétique et Phonologie, Porquerolles avril 2003.
Laks, Bernard 2005. Approches de la phonologie cognitive. Nouvelles approches en phonétique et en phonologie, edited by V. Rey & Noël Nguyen, 291-319. Paris: Hermès.
Laks, Bernard 2006. La phonologie générative naturelle et la phonologie naturelle. History of the Language Sciences, vol.3 (HSK 18.3), edited by Sylvain Auroux, E.F.K. Koerner, Hans-Josef Niederehe & Kees Versteegh, 2271-2280. Berlin: de Gruyter.
Langacker, Ronald 1987. Foundations of Cognitive Grammar, 2 Vols. Stanford: Stanford University Press.
Larson, G. 1992. Dynamic computational networks and the representation of phonological information. Ph.D dissertation, University of Chicago.
Leben, William 1973. Suprasegmental Phonology. Ph.D dissertation, MIT.
Leben, William & Orrin Robinson 1977. Upside-down Phonology. Language 53: 1-20.
Legendre, Geraldine, Yoshiro Miyata & Paul Smolensky 1990. Harmonic Grammar - a formal multi-level connectionist theory of linguistic well-formedness: theoretical foundations. Proceedings of the twelfth annual conference of the Cognitive Science Society, edited by the Cognitive Science Society, 388-395. Cambridge, MA.: Erlbaum.
Lévi-Strauss, Claude 1958. Anthropologie structurale. Paris: Plon.
Liberman, Mark 1975. The Intonational System of English. Ph.D dissertation, MIT.
Liberman, Mark & Alan Prince 1977. On Stress and Linguistic Rhythm. Linguistic Inquiry 8: 249-336.
Lightner, Theodore 1978. Generative Phonology. A Survey of Linguistic Science, edited by William Orr Dingwall, 1-32. 2nd edition Stamford, Connecticut: Greylock.
Lightner, Theodore 1981. New explorations into Derivational Morphology. Phonology in the 80's, edited by Didier Goyvaerts, 93-99. Ghent: Story-Scientia.
Lightner, Theodore 1985. On universal rules. Lingvisticae Investigationes 9: 419-420.
Lombardi, Linda 2001. Introduction. Segmental phonology in Optimality Theory. Constraints and Representations, edited by Linda Lombardi, 1-9. Cambridge: Cambridge University Press.
Lowenstamm, Jean 2006. Développement de la phonologie métrique. History of the Language Sciences, vol.3 (HSK 18.3), edited by Sylvain Auroux, E.F.K. Koerner, Hans-Josef Niederehe & Kees Versteegh, 2303-2312. Berlin: de Gruyter.
Lowenstamm, Jean & Jonathan Kaye 1986. Compensatory Lengthening in Tiberian Hebrew. Studies in Compensatory Lengthening, edited by Leo Wetzels & Engin Sezer, 97-132. Dordrecht: Foris.

Martinet, André 1955. Economie des changements phonétiques. Berne: Francke.
Martinet, André 1960. Eléments de linguistique générale. Nouvelle édition remaniée et mise à jour 1980. Paris: Armand Colin.
Martinet, André 1968. La linguistique synchronique. Paris: PUF.
Mascaró, Joan 1976. Catalan Phonology and the Phonological Cycle. Ph.D. dissertation. MIT.
McCarthy, John 1979. Formal problems in Semitic phonology and morphology. Ph.D dissertation, MIT.
McCarthy, John 1986. OCP Effects: Gemination and Antigemination. Linguistic Inquiry 17: 207-263.
McCarthy, John 2007. Hidden Generalizations: Phonological Opacity in Optimality Theory. London: Equinox.
McCarthy, John & Alan Prince 1986. Prosodic Morphology. Ms. University of Massachusetts et Brandeis University. Downloadable at ROA (#482).
McCarthy, John & Alan Prince 2001. Prosodic Morphology. Constraint Interaction and Satisfaction. Ms, ROA #482.
McMahon, April 2000. Lexical Phonology and the history of English. Cambridge: Cambridge University Press.
McMahon, April 2007. Who's afraid of the vowel shift rule? Language Sciences 29: 341-359.
Mohanan, Karuvannur 1986. The Theory of Lexical Phonology. Dordrecht: Reidel.
Moreton, Elliott 2008. Analytic bias and phonological typology. Phonology 25: 83-127.
Moulton, William 1947. Juncture in Modern Standard German. Language 23: 212-226.
Neeleman, Ad & Hans van de Koot 2006. On syntactic and phonological representations. Lingua 116: 1524-1552.
Nespor, Marina & Irene Vogel 1986. Prosodic Phonology. Dordrecht: Foris.
Newmeyer, Frederick 1986. Linguistic theory in America. 2nd edition New York: Academic Press.
Newmeyer, Frederick 1996. Generative linguistics: a historical perspective. London: Routledge.
Newmeyer, Frederick 1998. Language Form and Language Function. Cambridge, Mass.: MIT Press.
Newmeyer, Frederick 2005. Some Remarks on Roeper's Remarks on Chomsky's "Remarks". SKASE Journal of Theoretical Linguistics (www.skase.sk) 2: 26-39.
Oostendorp, Marc van 2002. The phonological and morphological status of the Prosodic Word Adjunct. Linguistische Berichte, Sonderheft 11: 209-235.
Oostendorp, Marc van 2003. Ambisyllabicity and Fricative Voicing in West-Germanic Dialects. The Syllable in Optimality Theory, edited by Caroline Féry & Ruben van de Vijver, 304-337. Cambridge: Cambridge University Press.
Oostendorp, Marc van 2005. The first person singular in Dutch dialects. Proceedings of the Thirty-Fifth Annual Meeting of the North East Linguistic Society, edited by Leah Bateman & Cherlon Ussery, 1-12. Amherst, MA: GLSA.
Oostendorp, Marc van 2006. A theory of morphosyntactic colours. Ms., Meertens Instituut.
Oostendorp, Marc van & Jeroen van de Weijer 2005. Phonological alphabets and the structure of the segment. The internal organisation of phonological segments, edited by Marc van Oostendorp & Jeroen van de Weijer, 1-23. Berlin: Mouton de Gruyter.
Paradis, Carole 1988. On constraints and repair strategies. The Linguistic Review 6: 71-97.
Pesetsky, David 1979. Russian Morphology and Lexical Theory. Ms, MIT. Available at http://web.mit.edu/linguistics/www/pesetsky/russmorph.pdf.
Pike, Kenneth 1947. Grammatical prerequisites to phonemic analysis. Word 3: 155-172.
Pike, Kenneth & E. Pike 1947. Immediate constituents of Mazatec syllables. International Journal of American Linguistics 13: 78-91.

Pöchtrager, Markus 2006. The structure of length. Ph.D dissertation, University of Vienna.
Prince, Alan 1983. Relating to the Grid. Linguistic Inquiry 14: 19-100.
Prince, Alan & Paul Smolensky 1993. Optimality Theory. Constraint Interaction in Generative Grammar. Ms, Rutgers University, University of Colorado (ROA version August 2002).
Pulgram, Ernst 1970. Syllable, Word, Nexus, Cursus. The Hague: Mouton.
Pulleyblank, Douglas 1988. Underspecification, the feature hierarchy and Tiv vowels. Phonology 5: 299-326.
Raimy, Eric & Charles Cairns (ed.) forth. Architecture and Representations in Phonological Theory. Cambridge, Mass.: MIT Press.
Rubach, Jerzy 1981. Cyclic Phonology and Palatalization in Polish and English. Warsaw: Wydawnictwa Universytetu Warszawskiego.
Rubach, Jerzy 1986. Abstract vowels in three dimensional phonology: the yers. The Linguistic Review 5: 247-280.
Rubach, Jerzy 1997. Extrasyllabic consonants in Polish: Derivational Optimality Theory. Derivations and Constraints in Phonology, edited by Iggy Roca, 551-581. Oxford: Clarendon.
Rubach, Jerzy & Geert Booij 1984. Morphological and prosodic domains in Lexical Phonology. Phonology Yearbook 1: 1-27.
Rubach, Jerzy & Geert Booij 1987. Postcyclic versus postlexical rules in Lexical Phonology. Linguistic Inquiry 18: 1-44.
Rubach, Jerzy & Geert Booij 2003. Lexical Phonology - Overview. International Encyclopedia of Linguistics, edited by William Frawley, 443-447. Second edition Oxford: Oxford University Press.
Rumelhart, D., J. McClelland & the PDP Research Group (eds.) 1986. Parallel Distributed Processing: Exploration in the Micro-Structure of Cognition. 2 vols. Cambridge, Mass.: MIT Press.
Rumelhart, David 1989. The Architecture of Mind: A Connectionist Approach. Foundations of Cognitive Science, edited by Michael Posner, 134-159. Cambridge, Mass.: MIT Press.
Russell, Kevin 1999. MOT: sketch of an OT approach to morphology. Ms., University of Manitoba.
Sagey, Elizabeth 1986. The representation of features and relations in nonlinear phonology. Ph.D dissertation. MIT.
Sapir, Edward 1933. La réalité psychologique du phonème. Journal de psychologie normale et pathologique 30: 247-265.
Schane, Sanford 1984. The fundamentals of particle phonology. Phonology Yearbook 1: 129-155.
Scheer, Tobias 2004. A Lateral Theory of Phonology. Vol.1: What is CVCV, and why should it be? Berlin: Mouton de Gruyter.
Scheer, Tobias 2008. Why the Prosodic Hierarchy is a diacritic and why the Interface must be Direct. Sounds of Silence: Empty Elements in Syntax and Phonology, edited by Jutta Hartmann, Veronika Hegedüs & Henk van Riemsdijk, 145-192. Amsterdam: Elsevier.
Scheer, Tobias 2009. Intermodular Argumentation and the Word-Spell-Out-Mystery. Explorations of Phase Theory: Interpretation at the Interfaces, edited by Kleanthes Grohmann, 23-65. Berlin: Mouton de Gruyter.
Scheer, Tobias forth a. How morpho-syntax talks to phonology. A survey of extra-phonological information in phonology since Trubetzkoy's Grenzsignale. Berlin: Mouton de Gruyter.
Scheer, Tobias forth b. Intermodular argumentation: morpheme-specific phonologies are out

of business in a phase-based architecture. The Sound Pattern of Syntax, edited by Nomi Shir & Lisa Rochman. Oxford: OUP.
Scobbie, James 1991. Towards Declarative Phonology. Edinburgh Working Papers in Cognitive Science 7: 1-26.
Scobbie, James, John Coleman & Steven Bird 1996. Key aspects of declarative phonology. Current Trends in Phonology: Models and Methods. Vol.2, edited by Jacques Durand & Bernard Laks, 685-709. Salford, Manchester: ESRI.
Selkirk, Elisabeth 1981. On the nature of phonological representation. The cognitive representation of speech, edited by J. Anderson, J. Laver & T. Meyers, 379-388. Amsterdam: North Holland.
Selkirk, Elisabeth 1981 [1978]. On prosodic structure and its relation to syntactic structure. Nordic Prosody II, edited by Thorstein Fretheim, 111-140. Trondheim: TAPIR.
Selkirk, Elisabeth 1984. Phonology and Syntax: The Relation between Sound and Structure. Cambridge, Mass.: MIT Press.
Siegel, Dorothy 1974. Topics in English Morphology. Ph.D. dissertation, MIT.
Skousen, Royal 1981. Analogical sources of abstractness. Phonology in the 80's, edited by Didier Goyvaerts, 55-92. Ghent: Story-Scientia.
Smolensky, Paul 1986. Information processing in dynamical systems: Foundations of Harmony Theory. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations, edited by D. E. Rumelhart, J. L. McClelland & the PDP Research Group, 194-281. Cambridge, MA: MIT Press, Bradford Books.
Smolensky, Paul 1987. Connectionist AI, symbolic AI, and the brain. Artificial Intelligence Review 1: 95-109.
Smolensky, Paul 1988. On the proper treatment of connectionism. Brain and Behavioural Sciences 11: 1-74.
Smolensky, Paul 1991. Connectionism, constituency and the language of thought. Meaning in Mind: Fodor and his critics, edited by B. Loewer & G. Rey, 201-227. Oxford: Blackwell.
Smolensky, Paul & Geraldine Legendre 2006. The Harmonic Mind. From Neural Computation to Optimality-Theoretic Grammar, 2 Vols. Cambrdige, Mass.: MIT Press.
Sommerstein, Alan 1974. On phonetically motivated rules. Journal of Linguistics 10: 71-94.
Stampe, David 1972. How I Spent my Summer Vacation. Ph.D dissertation, University of Chicago.
Steriade, Donca 1987. Locality Conditions and Feature Geometry. Proceedings of NELS 17, edited by Joyce McDonough & Bernadette Plunkett, 595-617. Amherst: GLSA.
Steriade, Donca 2007. Contrast. The Cambridge Handbook of Phonology, edited by Paul De Lacy, 139-157. Cambridge: CUP.
Stillings, Neil, Steven Weisler, Christopher Chase, Mark Feinstein, Jay Garfield & Edwina Rissland 1995. Cognitive Science. An Introduction. Cambrdige, Mass.: MIT Press.
Szigetvári, Péter 2001. Dismantling syllable structure. Acta Linguistica Hungarica 48: 155-181.
Szigetvári, Péter & Tobias Scheer 2005. Unified representations for the syllable and stress. Phonology 22: 37-75.
Taylor, John 2002. Cognitive Grammar. Oxford: Oxford University Press.
Tranel, Bernard 1981. Concreteness in Generative Phonology. Evidence from French. Berkeley: University of California Press.
Trubetzkoy, Nikolai Sergeyevich 1939. Grundzüge der Phonologie. 6th edition 1977, Göttingen: Vandenhoeck & Ruprecht.
Vaux, Bert 2003. Syllabification in Armenian, Universal Grammar, and the Lexicon.

Linguistic Inquiry 34: 91-125.
Vennemann, Theo 1971. Natural Generative Phonology. Paper presented at the Annual Meeting of the Linguistic Society of America, St. Louis.
Vennemann, Theo 1972a. Sound change and markedness theory: On the history of the German consonant system. Linguistic change and generative theory. Essays from the UCLA Conference on historical linguistics in the perspective of transformational theory (1969), edited by R.P. Stockwell & R.K.S. Macaulay, 230-274. Bloomington: Indiana Univ. Press.
Vennemann, Theo 1972b. On the theory of syllabic phonology. Linguistische Berichte 18: 1-18.
Vennemann, Theo 1974a. Phonological concreteness in natural generative grammar. Toward tomorrow's linguistics, edited by Roger Shuy & Charles Bailey, 202-219. Washington, DC: Georgetown University Press.
Vennemann, Theo 1974b. Words and syllables in natural generative grammar. Papers from the parasession on Natural Phonology, edited by A.Bruck, R.Fox & M.La Galy, 346-374. Chicago: Chicago Linguistic Society.
Vennemann, Theo 1976 [1971]. Vowel alternations in English, German and Gothic: remarks on realism in phonology. Linguistic and Literary Studies in Honor of Archibald A. Hill, vol. I: General and Theoretical Linguistics, edited by Muhammad Ali Jazayery, Edgar C. Polomé & Werner Winter, 337-359. Lisse: de Ridder.
Wells, Rulon S. 1949. Automatic alternations. Language 25: 99-116.
Williams, Edwin 1981. On the notions "Lexically Related" and "Head of a Word". Linguistic Inquiry 12: 245-274.
Williams, Edwin 2007. Dumping Lexicalism. The Oxford Handbook of Linguistic Interfaces, edited by Gillian Ramchand & Charles Reiss, 353-381. Oxforf: OUP.
Yip, Moira 1988. The OCP and Phonological Rules: A Loss of Identity. Linguistic Inquiry 19: 65-100.
Yip, Moira 1998. Identity Avoidance in Phonology and Morphology. Morphology and its Relation to Phonology and Syntax, edited by Steven Lapointe, Diane Brentari & Patrick Farrell, 216-246. Stanford: CSLI Publications.
Zwicky, Arnold 1974. Homing in: on arguing for remote representations. Journal of Linguistics 10: 55-69.