

Review of *The Legacy of Zellig Harris: Language and information into the 21st century. Volume 1: Philosophy of science, syntax and semantics. Edited by Bruce Nevin*

Reviewed by John Goldsmith, The University of Chicago

1. Introduction

Zellig Harris (1909-1992) cast a long shadow across 20th century linguistics. In mid-century, he was a leading figure in American linguistics, serving as president of the Linguistic Society of America in 1955, just a year before Roman Jakobson. It is fair to say that during that decade – the years just before generative grammar came on the scene – Zellig Harris and Charles Hockett were the two leading figures in the development of American linguistic theory. Today, I daresay Harris is remembered by most linguists as the mentor and advisor to Noam Chomsky at the University of Pennsylvania – and the originator of transformational analysis.

But Harris was an extraordinarily deep thinker about linguistic theory, and he made important contributions to many fields of linguistics. Though I never met him myself, I have often felt that a great deal of my own work was exploration of territory where he had already been, and had left signs for later researchers, signs that noted where important problems were to be found and how they might best be treated. And three of my teachers worked closely with Harris (Lila Gleitman, Haj Ross, and Noam Chomsky), and been greatly influenced by him. Truth in advertising: Zellig Harris is a bit like a grandfather I never met.

Bruce Nevin, a student of Harris's, has now produced a two-volume tribute to Harris bringing together work by range of researchers in linguistics and the other disciplines where Harris did significant work. All of the areas of Harris's linguistic work are covered in these volumes, and a careful reading of them leaves this reader with the conclusion that there is no way to understand American linguistic theory through the second half of the 20th century without understanding Zellig Harris's thought. I will attempt to explain why this is so in this discussion of volume 1; volume 2 focuses on computational issues. As Nevin notes, what he produced is neither a festschrift nor a memorial volume, and as my purpose here is to better understand Harris's role in the development of 20th century linguistics and the relevance of his thought for linguists today, I have little to say about some of the contributions; contributors were apparently invited to this project, and asked to offer a paper that clearly represented "some relationship to Harris's work," but this relationship is not always evident.

Harris's work must be situated in terms of the conflict between two visions of linguistic science: the *mediationalist* view, which sees the goal of linguistic research as the discovery of the way in which natural languages link form and meaning, and the *distributionalist* view, which sees the goal as the fully explicit rendering of how the individual pieces of language (phoneme, syllable, morpheme, word, construction, etc.) connect to one another in the ways that define each individual language.¹ The

¹ These ideas are developed at length in Huck and Goldsmith 1996.

mediationalist view lurks behind most conceptions of language study, formal and non-, but it was Harris' view that each successive improvement in linguistic theory took us a step further *away* from the mediationalist view, much as advances in biology led scientists to understand that the study of living cells required no new forms of energy, structure or organization in addition to those which were required to understand non-living matter. Harris had no use for mediationalist conceptions of linguistics. For linguists in 2004, steeped as we are in an atmosphere of linguistic mediationalists, this makes Harris quite difficult to understand at first.

Harris's goal was to show that all that was worthwhile in linguistic analysis could best be understood in terms of distribution of components at different hierarchical levels, because he understood – or at least he believed – that there was no other basis on which to establish a coherent and general linguistic theory. His genius lay in the construction of a conception of how such a vision could be put into place concretely.

Harris' view, from his earliest work through his final statements in the early 1990s, was that linguistics' best foundational chances were to be found in establishing a science of *external linguistic facts* (such as corpora, though they would typically be augmented by other external facts, like speakers' judgments), rather than a science of internalized speaker knowledge. Today, in mainstream linguistics, that decision is generally described as an *empiricist* leaning on Harris's part, in contrast to an alternative *rationalist* leaning that lay at the heart of Chomsky's conception of generative grammar by the late 1950s. Harris's linguistics was a science of English, or of Hidatsa or of Hebrew, insofar as it was manifested in various actual linguistic creations in those languages, while Chomsky's linguistic was, or is, a science of the capacities of speakers who could utter or create sentences in those languages.

I believe that it is very important to understand the challenge that is implicit in developing these two perspectives. As Morris Halle has noted,² it was a growing consciousness of a commitment to a rationalist linguistics (in the sense just noted), and its difference from Harris's empiricist view, which led to the intellectual estrangement between Harris and Chomsky, in marked contrast to their close relationship up until around 1960.³ There are important subfields of linguistics that remain thoroughly committed to an empiricist view of linguistics: all corpora linguistics, to be sure, is based on such a view,⁴ and most contemporary computational linguistics is concerned with solving various problems that arise out of dealing with English (Hidatsa, Somali, etc.) in

² Personal communication, January 9, 2004.

³ There is very little in this volume that explicitly compares generative grammar and Harrisian linguistics. Nevin, in his introduction, expresses regret that Noam Chomsky declined to contribute to the volume (xxvi).

⁴ For a strong statement of this perspective, one may see Geoffrey Sampson 2003, Are we nearly there yet, Mum? <http://www.grsampson.net/Aawn.html>.

the real world, rather than with providing a model of how human minds or brains might deal with such problems.

I myself am of the opinion that an updated version of Harris's conception does indeed provide the working linguist with a solid theoretical foundation for linguistics, arguably a more solid foundation than that provided by the current rationalist view, and I will fill in a few of the details on that score.⁵ Harris's picture is one that provides linguistics, for better or worse, with a nearly radical insulation from interaction with psychology and biology, and such a move is one that will be (at least at first glance) anathema to many leading theoreticians in linguistics, ranging across a spectrum from, say Ronald Langacker and George Lakoff to Ray Jackendoff and, of course, Noam Chomsky – all of whom feel that the study of linguistic structures should meet up with the study of human cognitive structures before too long (and may already have done so).

2. Harris's big idea

Harris did not appear to make a great effort to make his conclusions easily accessible to the reader. And yet once his ideas are understood, it is hard to deny that his way of stating them is direct, elegant, and striking. Let us approach the central idea of all of Harris's work, as summarized by Harris himself in his introductory paper to this volume:

[T]he structure of language can be found only from the non-equiprobability of combination of parts. This means that the description of a language is the description of contributory departures from equiprobability, and the least statement of such contributions (constraints) that is adequate to describe the sentences and discourses of the language is the most revealing. (p. 9)

Picking this apart into pieces:

⁵ To offer a concrete formulation, consider Andrea Calabrese's formulation of *realism* in Calabrese 2004:

This realistic view of language is based on two indisputable facts: given an utterance, 1) there must be a long term representation of the elements intervening in it; 2) there is an articulatory representation of it before actual muscular implementation. Phonology investigates the system of knowledge that allows the concrete occurrence of the real time computational steps that convert the mnemonic representation of the utterance into the articulatory representation of it. This knowledge involves representations and computation that have a concrete spatio-temporal occurrence and allows the production of concrete articulatory events, and are part of the workings of an actual brain with all its limitations.

Needless to say, perhaps, someone who did not agree with this position is not likely to want to call his position "non-realist"; but regardless of what names we apply to the positions, a person who (like Harris) felt that phonology need not view articulatory representations as the foundations of phonological theory might well reply that it may well be an "indisputable fact" that basic research in particle physics is carried out only with massive financial support from a central government, but that does tell us where the foundations of the scientific research lie.

1. Linguistic analysis consists of building a representation, built out of a finite number of formal objects.
2. The essence of any given language is the restrictions, or constraints, that it places on how the pieces may be put together – these may be phonemes, morphemes, constituents, what have you. If there were no structure, then pieces could be put together any which way; structure *means* – it is nothing more or less than – restrictions on how pieces can be put together.
3. These restrictions may be absolute (“no *pk* clusters are permitted in this language”) or – much more likely – they are statements of distribution, best expressed in the mathematics of probability. A crude reformulation of this would be in the language of markedness, which is arguably an informal way of talking about distributional frequencies. A better way is to use the mathematical vocabulary of distributions, which is to say, probability theory.
4. A formal system can be described formally in a multitude of ways. These are not equivalent: there is a priority among them based on their formal length. In general, one will be significantly shorter than the others, and knowing its length is important.

It is probably impossible to understand the intellectual pull of this research program if one does not appreciate the revolutionary character (and the perceived success) of the phoneme. If the phoneme today seems passé, the discarded error of an earlier generation, then today’s linguist should think of its descendant – for most of us, the idea of an underlying segment; I will return to this in a moment. By the mid 20th century, linguistics was arguably the most successful of the social sciences in dealing with the major challenge faced by anyone trying to bring the words “social” and “science” together: how to recognize and acknowledge the fact of human choice and action in the social world (we say what we want to, not what we are obliged to) and at the same time to apply a rigorous methodology that, at the end of the day, provides a revealing analysis. Phonemic analysis of natural language established once and for all that utterances in all the world’s languages were not just sequences of one damn sound after another,⁶ but were rather predictable realizations of a language’s choice of phonemes, a limited inventory chosen by a language with an eye to symmetry and perhaps even a concern for its history – for one thing that was certain was that the history of a language was not just a change of pronunciations, but a change in the organized structure of the sound system of the language, as manifested in its inventory of phonemes.

Today we are likely to think that classical phonemic analysis focused on finding a way to carve sound space into a language-particular set of boundaries; a vowel on one side of a particular boundary in English is a /e/, and a vowel on the other side is a /ɛ/. Learning the phonemes of a language (if we accept that view) amounts to learning a language-

⁶ Paraphrasing Elbert Hubbard and Arnold Toynbee.

particular sense of the distance between sounds (or phones): two phones not separated by any phoneme boundaries in a language are perceived by speakers of the language as identical, or nearly so, while sounds separated by a phonemic boundary are perceived as quite distant (hence, distinct). It is easy to think that this is the classical phonemic analysis, because this is what many phonemicists said, in so many words.

But in some respects, this view does not do justice to the notion of the phoneme, nor to phonology. Some sound-relationships are more complex than what we can get from establishing boundaries among the sounds of a language. Consider the flap of American English: it is systematically related to the released voiceless [t] and to the unreleased glottalized [tʔ], but the relationship is complex. Within words, the flap appears obligatorily when a stressed vowel precedes (with only the possibility of an /r/ interceding) and an unstressed vowel follows, and optionally when surrounded by unstressed vowels; when word-final, a flap appears optionally when the next word begins with a vowel, regardless of stress-level. Now the idea that *this* kind of depth and detail could govern the behavior of normal human beings in their everyday life was radical, and phonemics offered a method for discovering these relationships and for expressing the results in a simple format: there is a phoneme /t/, and its allophones are such-and-such, and each appears (obligatorily or optionally) in such-and-such contexts.

And if we no longer call this level of analysis *phonemics* – if we call it an analysis of the underlying inventory of phonological segments, or something else – we have not lost sight of the crucial insight. Zellig Harris had certainly not lost sight of the crucial insight; his genius (or one of them) was to understand what the methodology was that lay behind phonemic analysis, lay it bare, and extend it to other aspects of human language.

It was his view that the *important* relationship between sounds lay not in their phonetics, but in their *distribution*, along the lines that I sketched two paragraphs above for the flap. What tells us that the flap and the other *t*'s of English are realizations of a single phoneme /t/ is not the similarity of sound, but the complementarity and predictability of the distribution. How can we extend this style of analysis to morphology, to syntax, and beyond? *That* was Zellig Harris's program. This point is clearly noted by Ryckman ("Method and theory in Harris's Grammar of Information," 19-37) the central role played by the phoneme (p. 25), and its distributional definition, in the development of Harris's method and general understanding of linguistics; he also notes (p. 32) that in Harris's work "grammatical transformations are developed as a kind of 'extended morphophonemics', more powerful regularizing methods that enable even the derivation of tense and affixes."

Harris used the opportunity of a review of Trubetzkoy in 1941 to articulate his methodological disdain for phonetics in a well-known passage (Harris 1941, 346):

It is pointless to mix phonetic and distributional contrasts. If phonemes which are phonetically similar are also similar in their distribution, that is a result which must be independently proved. For the crux of the matter is that phonetic and distributional contrasts are methodologically different,

and that only distributional contrasts are relevant while phonetic contrasts are irrelevant.

This becomes clear as soon as we consider what is the scientific operation of working out the phonemic pattern. For phonemes are in the first instance determined on the basis of distribution. Two positional variants may be considered one phoneme if they are in complementary distribution; never otherwise. In identical environment (distribution) two sounds are assigned to two phonemes if their difference distinguishes one morpheme from another; in complementary distribution this test cannot be applied. ... [T]he distributional analysis is simply the unfolding of the criterion used for the original classification. If it yields a patterned arrangement of phonemes, that is an interesting result for linguistic structure.

From a broader perspective, there is an irony in choosing Trubetzkoy as the butt of this criticism: Trubetzkoy who, in his *Grundzüge* 1939, famously quoted Jakobson as saying that phonetics was to phonology as numismatics was to economics (in the context of a long discussion of the evolution of linguistic thought advancing towards a distinction between these sciences: an evolution which has continued *far* more in North America than it has in Europe, interestingly). But the review was an opportunity for Harris to express his own point of view.

3. Discovery procedures

All linguists are aware today that a major difference between Chomsky's generative grammar and Harris's earlier distribution-based grammar lay in the role they assigned to what Chomsky (but not Harris) called *discovery procedures*, a term coined by Chomsky to contrast with *evaluation procedures*, a notion central to generative grammar (up until Chomsky 1981, when it was replaced by principles and parameters). Understanding this difference is both difficult and important.

The distinction transparently alludes to an issue in the philosophy of science that was raging in the 1950s, one that had its roots in the 1930s, and which has not yet died down. In the philosophy of science, it referred to the distinction between the *context of discovery* and the *context of justification*, terms coined by Hans Reichenbach 1938. The context of discovery of a scientific theory concerns history and personal psychology; the context of justification concerns logic and the relationship between a scientific statement and observation statements about events in space and time. A legitimate positivist should not confuse the two, and ultimately the scientist-positivist cares only about the context of justification, not that of discovery.

What is the relationship of a linguist's work to his data? Are the data merely entries in his diary that provide some understanding of how he got to his eventual insight, or are they the stuff on which the theory stands or falls? This is the question an exegesis of Zellig Harris's work must answer. The rationalist today understands the linguist's data as being entries in the linguist's diary, of interest to historians of linguistics; the real subject of the

work is the theory and the models. The empiricist sees the data which is collected by the larger community of linguistic researchers as an integral part of the work of the scientific community; no data, no science. It's not that it's simply *unlikely* the scientist will stumble upon the right theory without looking carefully at the data; there is no *right* theory to speak of except insofar as theory is united with data.

If we accept that perspective, then we can better understand what Harris's procedures were for and were about. They were proposed as methods which were systematic, which could be defined so as to be used in identical ways by different analysts, and which would provide insights into a wide range of human languages in ways that other methods would not. Developing such a method was understood as the primary goal of the scientific methodologist.⁷

In "On discovery procedures" (69-86), Francis Y. Lin focuses on the status of discovery procedures in Harris's work. He offers the following as a typical example:

[W]e take a form A in an environment C—D and then substitute another form B in the place of A. If, after such substitution, we still have an expression which occurs in the language concerned, i.e., if not only CAD but also CBD occurs, we say that A and B are members of the same substitution-class, or that both A and B fill the position C—D, or the like (Harris 1946:102, cited by Lin 72).

Lin argues that the entire panoply of discovery procedures, such as those Zellig Harris presented, should be understood as a scientific theory of how language is learned by the child. Lin himself feels that we are astride a dilemma: the discovery procedures can be conceived of either as a set of procedures that a linguist (either in theory or in practice) is to follow, or else as a set of procedures that the child-language-learner is to follow. But Harris, says Lin, could not have meant the procedures to apply to the linguist: the linguist knows the phonemes, the morphemes, the transformations of the language before even starting the procedures, after all (80), and "no serious scientists would ever think of such an idea" as that of constructing an algorithmic procedure to analyze a set of data which the scientist him/herself is capable of analyzing by hand and for which no known algorithm that can replace the scientist exists at present; in this, it appears he is accepting, and following, Chomsky's suggestion (81), which Lin calls the 'oddity' argument: that 'there are few areas of science in which one would seriously consider the possibility of developing a general, practical, mechanical method for choosing among several theories, each compatible with available data.'

I'm afraid Lin's interpretation is unconvincing. There is no place in Harris which Lin can cite that supports this view, for starters. More to the point, this interpretation is the

⁷ T. A. Ryckman notes in his contribution (I, 22-37) notes: "Once in conversation, when I had referred to him as a 'linguist', Harris demurred, disclaiming any title as linguist, and said he thought of himself as a 'methodologist'." (22).

Chomskian interpretation: it is Chomsky who will, by the mid 1960s, be pushing linguistic theory as a theory of language acquisition by human children. Harris had every opportunity to make explicit that this had been his intention too – but it had not been, as I read Harris (which is to say, conservatively). Just as importantly, Harris, I think, would never have agreed to let linguistics become just a branch of psychology, which is to say, a cognitive science.⁸

Now this, I am sure, will strike many a reader as odd beyond measure. There is a belief widely held among linguists along the following lines: to be a science, linguistics must be *about something*, something in the real world; the only thing in the world it can be about is human brains; therefore, linguistics is a science that deals with biological (or psychological) facts about the brain or it is no science at all. Harris, I believe, would have rejected this logic, probably on the grounds that the scientific status of linguistics derives from its adherence to scientific *method*, not to whether there was a physical *thing*, one that we could point to, that was what linguistics was *about*. It was the development of a *method* that was Zellig Harris's lifework. Harris's framework of procedures, those called discovery procedures by Chomsky and others, was the heart of the method he was developing.

4. The autonomy of linguistics

This view of Harris was one facet of a radically autonomous conception of linguistics. If Harris was anti-cognitivist (less anachronistically, anti-psychologist) and anti-logicist (for he rejected all explanation of language on semantic grounds) and anti-phoneticist (as noted above), it was – at least at one level – because he was bound and determined to see linguistics become and remain an autonomous science, one whose conclusions were not dependent upon methods and conclusions of any other discipline, whether that discipline be psychology, logic, acoustics or biology. It seems to me – and I cannot underscore this enough – that Harris took a clear and defensible position on this, one of the most important foundational questions in our field, and we have been rather lax in coming to grips with these questions, by and large. The challenges come up time and again: in the late 1960s, psychologists took seriously generative grammar's claim to be a psychological model, and then they were disappointed, and perhaps even shocked, to find that linguists were not interested in changing their minds about linguistic analysis when psychological evidence could be brought to bear on selecting grammatical models. Linguists continued to act as if their linguistic methods were independent of psychologists' criticisms, just as Harris would tell them to.

⁸ Space does not permit a discussion of the history of psychologism, and its rejection, by logicians and philosophers since the late 19th century. Psychologism is the drive to reduce various fields of study (e.g., mathematics, linguistics, logic) to the study of how the human mind deals with the objects of those fields (numbers, sentences, deductions); it has waxed and waned several times in the last 150 years. A very brief remark along these lines by S. Shaumyan appeared on LinguistList in 1996: <http://www.linguistlist.org/issues/7/7-1478.html>

Harris adopted what seems to me to be the most honest and perhaps the most legitimate perspective on the relationship of linguistics to its neighboring fields: there is an interface between linguistics proper and such neighboring fields as phonetics, semantics, and psychology, and where there is disagreement, the differences have to be hashed out as best they can without a prior guarantee that both sides will come to an amicable agreement. When (distributional) linguistics tells us something about the sound structure of a language, such as that *angma* does not appear after silence in English (for this example, see Harris 1951 p.21), we may observe that “English speakers will in general find difficulty in pronouncing them.” But “all such predictions are outside the techniques and scope of descriptive linguistics. Linguistics offers no way of quantifying them. Nevertheless, taking the linguistic representation as a clear and systemic model of selected features of speech, we may find that this model correlates with other observations about the people who do the speaking.” Harris then directs the reader to Sapir’s classic observations about speaker intuitions in his “*La réalité psychologique des phonèmes.*” It is no accident, I would suggest, that this is precisely the same perspective that Chomsky took on the relationship of semantics to syntax during the 1960s, a subject discussed in detail in Huck and Goldsmith 1996: meaning is not part of the business of syntax or grammar, but a well carried out program of grammatical analysis is likely to have much to offer to neighboring disciplines that care about meaning and logical structure. Robert Longacre, in his contribution (“Some implications of Zellig Harris’s discourse analysis” 117-135) remarks on Harris’s unwillingness to see distribution-based analysis of discourse as yielding “information as to its meaning,” and one gets the impression that Longacre finds him being either overly modest or perverse (or perhaps ornery). Longacre rebuffs him in a thoroughly friendly way, saying, “Surely [if we follow Harris’s procedure] we may conclude that the result of [Discourse Analysis] is insight into the *meaning* of the articles so analyzed.” I am sure that if Harris were to agree to that statement, it would be with the proviso that insights are outside of the scientific content of the theory proper. Longacre then provides a useful comparison of Harris’s analysis of a modern fable by James Thurber and his own analysis

T.A. Ryckman, in his contribution to this volume cited above, discusses the philosophical context in which Harris embraces grammar and rejects logic: as Ryckman puts it, the “entry point...is the methodological postulate that there is no standpoint outside the data of language from which to advance theoretical inquiry.” (21) – and this statement is paraphrased as the title of Maurice Gross’s brief contribution (57-67) “Consequences of the metalanguage being included in the language.”

Ryckman offers a rather long description of a parallel between the challenge of finding a foundation for logic, on the one hand, and finding a foundation for grammar, on the other. In the case of logic, once understood as the analysis of sound inference in thought, all reference to psychological foundations has been abandoned since the discoveries of the early 20th century, in favor of a purely formal account of how logic operates. How does this evolution bear on linguistics?⁹ Ryckman offers some possibilities: a behaviorist

⁹ In the paper by Bar-Hillel I cite below, Bar-Hillel suggests that Carnap and Bloomfield were doing much the same thing, in trying to do their job without fuzzy semantics: “It is an interesting fact that, deserving the

Bloomfield is anti-mentalist, hence uninterested in an explanation founded on logic; Boas and Sapir were suspicious of categories historically derived from Indo-European languages. And Harris? What about Harris? Ryckman suggests that Harris cared more about patterns in language than in translation of language into another language, a logical language, and this is certainly true. But let's not forget that all academic conversation is dialog between opposing points of view. I think Ryckman's point can be strengthened by taking note of the context in which Harris was working in the 1940s and 1950s.

Readers of this journal fifty years ago were offered an explanation of this context by Yehoshua Bar-Hillel (1954), who began by pointing out to linguists that Rudolf Carnap (who Bar-Hillel regarded as "one of the greatest philosophers of all time" (1964, 4)) had thirty years earlier, in *The Logical Syntax of Language*, argued for a coming together of formal syntax and formal logic: by *formal*, he meant analysis ignoring meaning and considering only categories and combinations of symbols; by *syntax*, the rules by which items are combined to form expressions (sentences); and by *logic*, the rules by which valid inferences from one sentence to another can be made. The contrast between syntax and logic was dubbed by Carnap (in English) as the difference between *formation* rules and *transformation* rules – an interesting terminological suggestion, and one that may have later influenced Harris. Bar-Hillel's paper starts – quite literally from the very first sentence – as an attack on Zellig Harris's claims about the sufficiency of distributional methods for discovering the rules of both formation rules and transformation rules – and this from someone who had once been quite sympathetic to Harris's project. Regardless of whether we are convinced or not by Bar-Hillel's arguments today, *Harris* clearly was not. Bar-Hillel clearly thought that the innovations in logic brought about in the early 1930s by the Warsaw-Lwów school (Kotarbinski, Tarski, Adjukiewicz) were sufficient to dispel Bloomfieldian doubts about the dangers inherent in basing grammatical analysis on formal logic; Harris did *not* agree. Paul Mattick, in his contribution to this volume ("Some implications of Zellig Harris's work for the philosophy of science", 39-55), notes that the studies that had given rise to Bar-Hillel's optimism were quickly subject to withering criticism, and that Harris did in later years allude to his reasons for not taking the Carnapian approach seriously: "Not a few of the difficulties in the philosophy of language and in neighboring areas of philosophy arise from starting with the equipment which had been developed for truth systems and using it to analyze the information system that language represents." (Harris 1976, cited on p. 43). Mattick, in another context, summarizes one of the conclusions of Harris's later work on scientific language, and notes that "although logical structures are to be identified in natural language

attention of sociologists of science, that at approximately the same time, but in complete independence of each other, Bloomfield and Carnap were fighting the psychologism that dominated their respective fields, linguistics and logic. They both deplored the mentalistic mud into which the study of meaning had fallen, and tried to reconstruct their fields on a purely formal-structural basis. I think it is correct to say that the difference between the structural linguist and the formal logician is one of stress and degree rather than of kind. (p. 43 in Bar-Hillel 1964)." Complete independence seems to this writer unlikely: Carnap and Bloomfield were, after all, colleagues in the Humanities Division at the University of Chicago from 1936 (when Carnap arrived) to 1940 (when Bloomfield left for Yale); Bar-Hillel later visited Carnap at Chicago when Bloomfield was no longer there.

discourse, as in all cognitive activity, natural language is a much richer system of representation than logical calculi.” (p. 51).

Returning to Ryckman’s development of Harris’s overall program, he notes that there is a natural connection between the emphasis in the late 1940s and 1950s on information theory, on the one hand, and Harris’s central view of language as a system of many levels, in which items at each level are combined according to their local principles of combination, and in which relatively simple cross-level principles could be established. What Ryckman does not say – but what is very clear from the quotation from Harris that I cited near the beginning of this review – is that Harris’s method naturally leads one to a probabilistic grammar: *the description of a language is the description of contributory departures from equiprobability*, says Harris; hence, with a bit of mathematical work, to be sure, any utterance can be assigned a probability, and it can thus be assigned an information content (through the basic definition that information content is the base 2 logarithm of the reciprocal of the probability of the message; this is a well-known expression in information theory). Ryckman then goes on to discuss Harris’s idea of a “least grammar”, and he makes the connection to the ideas of computational complexity developed by Kolmogorov, Solomonoff, and others. Here I think Ryckman is right in broad outline but off the mark in the details.¹⁰ He suggests that “since by definition, language structure is a structure of the restrictions on combinations of elements recognized by the speakers of the language in question, it is imperative that the characterization of this structure (in a ‘grammar’) not contribute to the redundancies of combination, the bearers of information in the language, to be described. This is not just the general methodological virtue of economy of means. If language structure *is* informational structure, the requirement of a ‘least grammar’ is not a nicety, it is a necessity.” (31), and here Ryckman cites Harris himself: “the grammatical description [must be kept] as unredundant as possible so that the essential redundancy of language, as an information –bearing system...not be masked by further redundancy in the description itself.” (Harris 1982, 10-11) Sympathetic though I am to the enterprise, this seems to me to be incorrect. One can have two grammars, each complete and each assigning exactly the same probabilities to exactly the same set of sentences, and at the same time, one grammar can be more redundant than the other, however one chooses to define redundancy (“wordier”, longer in its formulations): for example, one of the two may distinguish between a generalization holding in main clauses and the same generalization holding in embedded clauses, while the other states the generalization once for all clauses. From an information-theoretic point of view, both grammars perform precisely the same function as long as they assign the same probabilities, and neither Ryckman nor Harris explain why the shorter, less redundant version is superior to the longer. Additional considerations need to be brought to bear here, it seems to me, of the sort offered by Minimum Description Length analysis (Rissanen 1989) and others; I have tried to address this question in a Bayesian fashion in Goldsmith 2004, where I suggest

¹⁰ I have tried to deal with this subject in Goldsmith 2004, and also in a particular context, that of morphology, in Goldsmith 2001.

that the motivation for formal accounts of empirical data is based implicitly on a prior distribution over formal accounts.¹¹

5. Learning

Lila Gleitman (“Verbs of a feather flock together II”, 209-229) explores an exegesis of Harrisian doctrine from a related perspective, but rather than telling us what Harris actually meant, she offers a personal overview of what we have learned in the last several decades about child language-learning, and she highlights the many aspects of that learning which have turned out to be *distribution-learning* in important ways. There are several ways that this could be understood: one is that this exegesis is what Harris really meant, and he was right when he meant it this way, and if he didn’t tell the *whole* story, at least he got some significant part of it right (this interpretation would be a continuation of Lin’s paper, and it’s not what Gleitman intends); a second interpretation would be that Harris’s ideas were important in a closely related field that he didn’t happen to be all that interested in (which is closer to what Gleitman means, I think); and a third interpretation is that Harris’s ideas, grounded as they were in a deep intuition about the nature of language, gave a picture of any *possible* theory of language analysis, so it’s not surprising that *one* class of language-learners, children, follow many of his proposals. I would opt for the third interpretation of why Harris’s ideas are as relevant as they are to the psycholinguistics of language acquisition: Harris was interested in determining what procedures *in principle* could lead to a deep understanding of a natural language system, so it shouldn’t be surprising that the one existing system that actually acquires a natural language should display a set of behaviors that resemble in interesting ways a Harrisian system.

We can go a step further and inquire (so to speak) of Gleitman what aspects of contemporary linguistics are of the most value in understanding child language acquisition. Some mainstream theorists have opted in strong and even strident terms for the position that language acquisition is not learning at all, but is rather akin to the maturation of a biological organ like a liver or an eye. It seems to me that at a point like this, there must be an amicable parting of the ways between such theorizing and the study of child language acquisition: such linguistic theory has nothing to say of interest to psycholinguists. Gleitman goes on to give an overview of some of the most interesting results in language acquisition results of the last decade, and not surprisingly, they involve, well, learning: real learning, from evidence, data in the language at hand.

She points to work done during the 1990s by Kuhl, Jusczyk and others demonstrating the early acquisition by children of knowledge of the distribution of sounds in acoustic space. This work suggests that in some fashion children recognize that the linguistics sounds

¹¹ In my opinion, there is another important idea here that neither Ryckman nor Harris have exploited, but which is the heart of Minimum Description Length analysis, and that is that both information content and grammar conciseness are notions that can be quantified in the same unit, the (information-theoretic) bit. One can thus coherently ask of an analysis that it *minimize* the sum of the grammar length (in bits) and the adequacy of the data-modeling (also measured in bits).

they hear are not uniformly distributed over acoustic space, but rather are distributed in a more finely articulated way – in a distribution that is the sum of several gaussians (several normal distributions), perhaps, where each gaussian represents the sound realization of a phoneme, or rather something at a slightly finer level of structure; armed with such a parametric model of sound distribution, children deform the metric of acoustic space in such a way that two sounds close to the center of the same gaussian are perceived as being extremely close to one another. If this view is correct, the learning that is involved is based on distribution, but not the distribution of a structuralist linguist such as Harris, but rather of the statistician who enforces a parametric distribution on the data.

Gleitman points as well to the influential and widely cited Saffran, Aslin, and Newport 1996, which provide evidence that children learn something like words even without knowing that they are symbolic, that is, that they are associated in any sense with meaning. Experiments have been constructed in which within chunks of word-size length, the mutual information between successive segments is high, while across these chunks the mutual information is zero (“mutual information” here measures the degree to which there is a statistical connection between the values taken on by successive phonemes, crudely put). Mutual information can be understood as a way of modeling what linguists would call phonotactics, and the laboratory evidence suggests that learners are capable of using this information in identifying language chunks.

How do children learn about categories of words, such as *noun* and *verb*? Gleitman notes,

The most familiar modern statement of a distributional learning algorithm for grammatical categories is from Maratsos & Chalkley (1980) whose schema directly follows linguists including Bloomfield (1933) and Harris (1951). They proposed that children could sort words into grammatical categories by noting their co-occurrences with other morphemes and their privileges of occurrences in sentences. thus *-ed* is (probabilistically speaking) a verb-follower and *the* is a noun preceder. That these analyses are carried out by children even where the supportive semantic evidence is absent ...makes a strong case for the continuing power of distributional analysis at higher linguistic levels....Just as Harris proposed for linguistics-internal descriptive purposes, there is by now abundant evidence that probabilistic distributional analyses operate again and again to build up phonetic, syllabic, word, and sentence-like units and – by virtue of the same procedures – to grasp the distributional properties within each level.

Here I have some trouble with the direction of the discussion. As far as I can see, psycholinguists do not have the tools to develop what is or isn't a good “noun preceder,” and if Bloomfield or Harris thought that *the* is a good noun preceder, it may mean that they didn't look at what words actually follow *the* in English. Most of the highest frequency words following both *the* and *a* are actually not nouns, but adjectives and adverbs. In the Brown Corpus, for example, the five most popular words after *the* are *first*, *same*, *most*, *other*, and *new*; after *a*, the top ten are: *few*, *little*, *new*, *man*, *good*, *small*, *great*, *very*, *long*, and *number* (all in these orders). Computational linguists are working these days on developing

perspectives that will categorize words into a pattern that resembles traditional linguistic labeling; the results are either promising or disappointing, depending on your expectations (I lean towards the former).

But the heart of Gleitman's argument in this paper deals with a point that speaks to one of the central points in Harris's program. It involves a skill that English speakers have that Gleitman has been interested in since she was a student of Harris's (and I remember her presenting it to my introductory linguistics class in 1969, when she called it *The Great Verb Game*). Tell a group of linguistically sophisticated people that you are thinking of a verb, and present them with one syntactic subcategorization frame (NP __ NP *from* NP, e.g.). Rarely will that be enough information to guess the verb. Give them *another* frame in which the same verb appears, and then another, and generally someone will guess the right verb after three or four such clues. Furthermore, the set of verbs that people guess become more and more semantically limited and homogeneous as they are given more syntactic frames. Why is this? What does this tell us about the importance of distributional characteristics of words in language? And can it be linked to other things that we know about the acquisition of language, such as the robust and highly significant finding that in early language acquisition, nouns dominate the child vocabulary, a result "robust across individuals and across languages" (216).

Gleitman's Harrisian idea (which she calls *syntactic bootstrapping*) is that knowledge of *verbs* is based largely on the syntactic environments in which they appear: such knowledge is a prime component of the relevant input to the learning procedure needed to learn verbs, in *any* language. Knowledge of nouns, by way of contrast, is much less dependent on syntactic context, and knowledge of the environment of language use can play a much more robust role in learning the meaning and function of nouns – at least, of concrete nouns. And, in her words, "the required 'sophisticated' linguistic representations [necessary to learn verbs] must themselves be constructed by using lower-level representations as the scaffolding."

Gleitman and her colleagues presented video snippets of mother-child interactions without sound to college-age students, and told them that the mother uttered a common noun or verb (she told them which) at the moment that the subject hears a beep along with the video. Thus the subject has most of the pragmatic context that the child did that she might use to infer what the *meaning* was that the mother intended, at that moment; the subject, furthermore, already *knows* the words and merely needs to pick out the one she thinks the mother must have meant in that context.

The results from the experiment are very clear: the task is *much* easier when the subject is asked to guess the noun the mother used, compared to when the task is to guess the verb. At the same time, further evidence strongly suggests that the most significant factor in account for this difference is not the grammatical category difference, but the difference in *imageability*, a gradient property on which verbs as a whole uniformly score worse than nouns as a whole (though

there is additional refinement to the scale that goes well beyond grammatical category). Gleitman concludes that a person can learn a lot about noun-word meanings, given real world interaction but no knowledge of linguistic context – but a person cannot learn much about verb meanings without the linguistic context. Once a person knows enough about the language to be able to identify nouns in the speech stream, it becomes possible to learn the crucial properties of verbs, especially those that help identify many of the subtle semantic characteristics of a verb’s meaning.

This is an subtle and extremely appealing vision of some of the critical aspects of basic language-learning, and bring to the picture a pair of highly correlated distinctions (imageability on the one hand, and noun/verb on the other) whose relevance to acquisition was not foreseen by linguistic theory (Ryckman presents Harris’s imagined scenario for the way children project from reality to words (32ff), and there is little hint that nouns and verbs function according to different principles).¹² I can only conclude that this is an important consequence of Harrisian sensitivity to the importance of distributional analysis, one that Harris himself might never have countenanced. Gleitman ends her paper with a thought about what Harris’s reaction to her paper might have been. “Obviously,” he would have responded. “I told you so.”

6. Other matters: syntactic structure, sublanguages

The late Maurice Gross (“Consequences of the metalanguage being included in the language”, 57-67) suggested that “Harris demonstrated exceptional intellectual courage in abandoning the notion [of *subject*] and adopting for the description of sentences the general schema $N_0 V W$.” He goes on to explore a few of the non-sentential structures found in English (*Good night, Merry Christmas, The hell with that!*), and the difficulty of establishing generalizations without significant exceptions.

On quite another hand, Pieter A. M. Seuren discusses the difference between arguments and adjuncts, exploring a number of related questions, such as differences between pairs of sentences given in (2):

¹² Needless to say, there are enormous additional steps that need to be taken, including one place where I think the present paper understates the level of the challenge for language learning; Gleitman notes that “a correlated language-internal cue to subjecthood is that different categories of nouns probabilistically perform different thematic roles....Specifically, animate nouns are vastly more likely than inanimates to be the causal agents in events. Once the position of the sentence subject is derived by matching up the observed agent with its known noun label, the young learner has a pretty good handle on the clause-level phrase structure of the exposure language.” (223) Actually, I don’t think this could be called a good handle on the data needed to infer phrase structure. To make this point in detail would be long; but imagine how difficult the task would be of writing an automatic algorithm that would take in an unlabeled and unanalyzed text in an unknown language and infer the phrase structure rules of the language. Hard way past what we know how to do at present. Now imagine the same text, but the phrases corresponding to the most animate referents in the discourse have been marked as such (by some oracle, so to speak). Does that make the task any easier? Not in any way we know today.

- (2)
- a. She gave her sister a hug.
 - b. *She gave a hug to her sister.
 - c. She wrote a letter to the Pope.
 - d. She wrote the Pope a letter.
 - e. John loves Africa.
 - f. Africa is loved by John.

But it is not hard to find many sentences of the sort exemplified by 2b on the internet, such as *The other day, I gave a hug to a vendor named Jerome in an awkward moment; Howie's mother gave a hug to the girl walking out of the kitchen; or Even more strangely I noticed that Sharon gave a hug to the head of the Psychology department, who would eventually be my professor that second semester;* and so on, to the point where one wonders why someone might have starred (2b). Seuren suggests that (2d), unlike (2c), evokes a relation of familiarity between the Pope and the letter writer, yet we find many examples of the following sort on the internet, presupposing no such familiarity: *An American delegation arranged to go to see the Pope. I couldn't go, but organizers said that if I wrote the Pope a letter, they would deliver it personally.* Or again: *With a quick Google search we found the Pope's address in Vatican City. We wrote the Pope a letter and we are now anxiously awaiting his reply.*¹³ Seuren offers an analysis which provides for a deep syntactic distinction (with important semantic consequences) between the location of NPs complement to prepositions and those that are syntactic arguments of the verb; this reader, at least, would have been more convinced by an analysis which predicted more robust grammatical differences. Indeed, without wanting to get deeply into a question of the appropriate origins of data for grammatical analysis, I have already indicated that one major difference between Zellig Harris's conception of linguistic analysis and that of generative grammar's conception is that Harris took the task to be analysis of what Chomsky has called E-language, the totality of intersubjective linguistic events (which includes, to be sure, much of what the internet has to offer), while generative grammar takes the task to be the description of I-language, the internalized knowledge of grammar possessed by humans. The Harrisian E-view has been criticized for being, in some critical cases, too sparse: where, oh where, are the examples of parasitic gaps on the internet? – but by the same token, we are becoming increasingly sensitive to the fallibility of data on which the Chomskian I-view is based; for a recent discussion, see Bresnan 2004.

Morris Salkoff (“Some new results on Transfer Grammar,” 167-178) presents several fascinating cases that arise from a systematic study of the patterns of translation between French and English, work done in the context of a machine translation project. Salkoff extends some of Harris's suggestions on the general theme with striking illustrations.

¹³ Another, longer example: *Two monks, both heavy smokers, were unable to stop smoking during prayers. To solve this dilemma they each wrote the Pope a letter to get his opinion on the subject. Three weeks later they got together again and found to their surprise that the Pope had given only one of them permission to smoke. On comparing their letters, they found the explanation. The first monk had asked, "May I smoke during prayers?" and the answer was, "No." The second monk, however, had framed his question differently: "May I pray while I am smoking?"*

Particularly interesting are the cases where two languages (here, French and English) have a semantic contrast that is in rough correspondence, such as the French *être en train de* and the English progressive *be+ing*. It will be true for most English verbs that they can appear in the English progressive form if and only if their best French translation can appear after *être en train de*. And yet there are exceptions at the edges: *Max est en train de comprendre le problème* is fine (he is beginning to understand the problem), we do not find ??*Max is understanding the problem*.¹⁴ And, we might add, there are mismatches the other way: if McDonald's in English can declare, *I'm lovin' it*, McDo (in France) cannot translate this as ??*Je suis en train de l'aimer*.¹⁵

There are two contributions to this volume, "Grammatical specification of scientific sublanguages," by Michael Gottfried and "Classifiers and reference," by James Munz, which deal with Harris's treatment of sublanguages, especially as detailed in Harris et al 1989. This reviewer found the papers hard going. If I understand the challenge here, it is this: how can linguistic analysis shed light on a set of documents which are all about a similar theme and part of a common style of discourse? This is not a question linguists generally ask, but whether it is a reasonable question to ask depends mainly on whether insightful or useful answers follow, and it is not easy for this reader to say whether the results are revealing in the case of scientific sublanguages. In the same section, Carlota Smith ("Accounting for subjectivity (point of view)" 137-163) offers a number of insightful comments on indicators of perspective and point of view, based in part on a number of examples from fiction. Daythal L. Kendall ("Operator grammar and the poetic form of Takelma texts", p. 261-278) gives a fascinating analysis of a Takelma text "Coyote's Rock Grandson," with a range of structures binding the sections of the story. The late Fred Lukoff ("A practical application of string analysis", 279-304) discusses the usefulness of grammatical simplifications, identified using string analysis, for learners of Korean.

7. Phonology

Frank Harary and Stephen Helmreich, in their paper "On the bipartite distribution of phonemes," (241-258) return to the subject of a paper by Harary and Paper 1957 on the distributional properties of vowels and consonants. They explore a radically distributionalist view of what it means to distinguish between vowels and consonants, adumbrated in Harris 1951. This view says that the proper division of segments in a language into vowels(V) and consonants (C) is that division for which the largest number of transitions from one segment to the next (in the utterances of the language) are transitions from vowel to consonant, or consonant to vowel; alternatively put, it is that division for which the smallest number of transitions are C to C or V to V. Harary and

¹⁴ Actually, we do find it. There are hundreds of examples (though not with the name *Max*) on the internet, most of them quite natural, like this one: *Throughout, Rebecca has been nodding or saying OK, letting the mediator know she is understanding.*

¹⁵ In fact, McDo does not translate it, for most of their publicity in France, leaving it in English; but when they do translate it, the sentence becomes *C'est tout ce que j'aime*.

Helmreich suggesting associating with any such division a number that they call the *bipartness ratio*: this is the ratio of the transitions in a corpus that are from V to C, or C to V, divided by the total number of transitions in the corpus; this ratio is thus bounded from above by 1.0.

Digressing slightly from the paper, from a mathematical point of view, this is naturally expressed in the following terms: the vowel/consonant distinction can be thought of as a *characteristic function*, a function whose domain is the phonemes, and whose range is the set $\{0,1\}$; phonemes mapping to 1 are vowels, and those mapping to 0 are consonants. An analysis that looks for the function that maximizes a mapping from functions to real numbers is referred to as a solution of a *variational* problem.

Harary and Helmreich's procedure is brute force, in the sense that they consider each of the 2^{k-1} ways of dividing k phonemes into 2 groups. For a language with as many as 73 phonemes, as is the case with Achumawi, one of the languages they explore, it is not practical to search all of them. I will return to an alternative strategy below.

The authors explore Hawaiian, Esperanto, Spanish, and Achumawi, and obtain reasonable results in each case, but the results do depend rather heavily on whether the digraph that links adjacent words is counted or not,¹⁶ and on whether the relative frequencies of transitions are counted (as I have suggested thus far they are) or not: in the latter case, a phoneme-to-phoneme transition is counted as occurring once whether it has occurred once or 10,000 times. The best analysis is one in which word-word transitions are ignored, and the frequencies of transitions are taken into consideration – not surprisingly, in my view.

What is the point of this? The authors suggest some practical applications, and speculate on how these procedures could be extended to further phonological questions. But the point of this work within a book such as this one is that it serves to justify Harris's position of the radical autonomy of linguistics even vis-à-vis phonetics: the vowel/consonant distribution can be justified as the optimal solution of a problem conceived in purely linguistic terms, or rather, in terms that can be reduced to representations that consist of sequential and discrete sequences of symbols chosen from a phonetic alphabet, that is, phonetic transcriptions.

Harary and Helmreich employ the terminology of graph theory in their exposition, as the term *bipartness* (above) suggests: the nodes of the graph in question are the segments of the language, the edges are the phonologically permitted transitions, and the bipartness ratio is a quantitative measure of the degree to which the graph satisfies the conditions for being a bipartite graph (one in which the nodes can be divided into two proper subsets in

¹⁶ That is, since there is in most languages no local phonological principle that governs the transition from the last phoneme of a word to the first phoneme of the next, that transition is governed by word transition probabilities and the distribution of vowels vs. consonants in word-initial and word-final positions. Some languages, such as Spanish or Swahili, have strong biases towards vowels in word-final positions, while others, such as English, do not.

such a way that no edge has both vertices in the same subset). There are developments in graph theory in the past two decades that provide what I think are rather superior techniques and results to those described in this paper, however. Spectral graph theory employs techniques of linear algebra, applied to matrices derived from graphs, and one of the most important basic results in this area is that the eigenvector corresponding to the second smallest eigenvalue of the laplacian of a graph satisfies the variational problem that Harary and Helmreich seek to solve (see Chung 1997, or Godsil and Royle 2001 for a gentler introduction to spectral graph theory).¹⁷ Two things follow from this last observation. First, brute force total search methods are not needed; a very simple and quick eigenvector decomposition does the trick, from a computational point of view. Second, the results that are obtained distributes the segments over a continuum, with roughly half assigned negative numbers and roughly half assigned positive numbers.¹⁸ If we sort the phonemes of a language by their associated coordinate in this eigenvector, we get a graded statement of the cline from the most vowel-like to the most consonant-like phonemes of the language, a neat trick: in effect, a purely distributionalist account of the sonority hierarchy.

Leigh Lisker (“The voiceless unaspirated stops of English,” 233-240) presents experimental evidence regarding the phonetic realization of /b/ after /s/ (“this bout”), of /p/ after /s/ (“the spout”) if this is indeed a /p/, and a word-final /p/ before vowel (“[dro]p out”). He presented the phonetic material that immediately follows the labial closure in each case, and asked the subjects to determine which of three natural source sentences (“Did he win this bout?”, etc.) each sample came from. The result was that speakers could not distinguish the three, suggesting a single phonetic notation is appropriate for these three cases.

8. Finally: Why is Harris’s work so unknown?

Why is Harris’s work after *Structural Linguistics* (1951) so little known – especially his work in the years after 1960? Surely it is not because there was little of it, nor because it was of no inherent interest. It did not help that Harris did not attend linguistics conferences – but then again, neither has Chomsky. Perhaps it would have helped if Harris had come to LSA meetings. Chomsky’s views have been consistently presented at linguistics conferences by his students, but Harris’s views, by and large, have not been. Perhaps it would have helped if Harris’s students had defended his ideas at LSA

¹⁷ I explore such an approach in the context of syntactic distribution in Belkin and Goldsmith 2002, and in the context of phonology in a paper in progress.

¹⁸ The result depends on being able to perform the eigenvector decomposition on a symmetric matrix, which requires some non-obvious manipulations. With appropriate manipulations we can restrict our attention to undirected graphs, hence symmetric matrices, hence everywhere real eigenvalues, and a situation in which all eigenvectors are orthogonal. The explicit coordinates of the eigenvector associated with the lowest eigenvalue takes on values that reflect the frequencies of the phonemes (that is, of the nodes), and these are all positive; since all other eigenvectors must be orthogonal to this first eigenvector, we know that the coordinates of the other eigenvectors distribute themselves with roughly half their values greater than zero, and roughly half less than zero, as noted in the text.

meetings. Even so, even if these things had occurred, it would have been necessary to go to Harris's writings, and they are not easy to read, nor are they captivating. Perhaps what they lack most of all is what Thomas Kuhn (1962) originally referred to as *paradigms* (before the term took on additional freight): good examples of analyses that are striking and that make the reader want to copy them, in a creative sort of way. It is not that Harris was incapable of providing such paradigms; his idea in "From phoneme to morpheme" (1955) about detecting morpheme boundaries by counting possible phoneme successors had just that degree of insight, cleverness, and elegance, and there are linguists who continue to find research inspiration in that suggestion.

So one possible answer lies in the failure of the Harrisian school to produce a cadre of dedicated young scholars willing to proselytize; another in the failure to develop a product that could be mass-marketed. A third surely lies in the poor fit between Harris's style of linguistics and the superfield of cognitive science that held sway during the last quarter of the 20th century.

Zellig Harris's work offered more opportunities to the style of work that was done not by linguists over the last three decades, but by computational linguistics over the last fifteen years. But that is a story for another time, and the focus of volume 2 of Nevin's presentation of Harris's legacy.

References

- Bresnan, Joan. 2004. CLS paper.
- Bar-Hillel, Yehoshua. 1954. Logical syntax and semantics. *Language* 30:230-237.
- Bar-Hillel, Yehoshua. 1964. *Language and Information: Selected Essays on their Theory and Application*. Reading MA: Addison-Wesley.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chung, Fan. 1997. *Spectral Graph Theory*. American Mathematical Society.
- Godsil, Chris and Gordon Royle. 2001. *Algebraic Graph Theory*. Berlin: Springer Verlag.
- Goldsmith, John. 2001. Unsupervised Learning of the Morphology of a Natural Language. *Computational Linguistics* 27:2 pp. 153-19
- Goldsmith, John. 2004. From algorithm to generative grammar and back again. *Papers from the Fortieth Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistics Society.
- Harary, Frank and Herbert H. Paper. 1957. Toward a general calculus of phonemic distribution. *Language* 33: 143-169.
- Harris, Zellig S. 1941. Review of Trubetzkoy's *Grundzüge der Phonologie*. *Language* 17.4:345-9.
- Harris, Zellig S. 1946. From morpheme to utterance. *Language* 22(3): 161-183.
- Harris, Zellig S. 1951. *Methods in Structural Linguistics*. Chicago: University of Chicago Press.
- Harris, Zellig S. 1955. From phoneme to morpheme. *Language* 31.2: 190-222.
- Harris, Zellig S. 1976. A theory of language structure. *American Philosophical Quarterly* 13:237-255.
- Harris, Zellig S. 1982. *A Grammar of English on Mathematical Principles*.
- Harris, Zellig, M. Gottfried, Thomas Ryckman, Paul Mattick, Jr., Anne Daladier, T.N. Harris and S. Harris. 1989. *The Form of Information in Science: Analysis of an immunology sublanguage*. Dordrecht: Kluwer.
- Huck, Geoffrey and John Goldsmith. 1996. *Ideology and Linguistic Theory*. London: Routledge.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.

Reichenbach, H., 1938, *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge*, The University of Chicago Press, Chicago, Illinois, U.S.A.

Rissanen, Jorma. 1989. *Stochastic Complexity in Statistical Inquiry*. Singapore: World Scientific.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month old infants. *Science*, 274, 1926-1928.

Trubetzkoy, 1939. *Grundzüge der Phonologie*. Travaux du Cercle Linguistique de Prague 7.