
This book was developed from a lecture given at the Courant Institute of Mathematical Sciences, and is addressed primarily to mathematicians. It might thus escape the notice of a large number of linguists. This would be a pity, for there is a great deal of material in it that will interest the linguist. Indeed, the book stands out as a significant and unusual one in the literature of structural linguistics, for it is about the properties that languages are found to possess rather than about the methods linguists employ or the concepts and attitudes they adopt; as Humberstone (1971) has remarked, structural linguists in general have not written at book length about what they claim is their subject matter, but have written rather textbooks introducing the student to the profession of linguist and counseling him in objective observation and rigorous analysis. "Strange that, after all this, there should be nothing left to say", Humberstone comments (115).

In this book Zellig Harris, author of what is surely the most important book on structuralist method (1951) has a lot more to say, and it would be a mistake to think that because Harris claims from page 1 onward to be developing "mathematical linguistics" (and has, furthermore, published in a series devoted exclusively to mathematics) his work is in any way marginal to linguistics as a whole. As Martin Joos notes about his own paper (in Joos, 1957: 356), there was "exactly nothing new" in 1950 about describing linguistics as "a discontinuous or discrete mathematics" (349). Harris's book may be seen as a preliminary presentation of the "theory of natural language-design" that Joos was looking forward to.

Harris's aim is "to show how one can arrive at an abstract system which characterizes precisely natural language" (v). He stresses (for the benefit of the mathematician reader) that he is interested "not in investigating a mathematically definable system which has some relation to language, as being a generalization or a subset of it, but formulating as a mathematical system all the properties and relations necessary and sufficient for the whole of natural language" (1). The mathematical system he presents in chapters 3 to 6 could not be reasonably summarized here, and is in any case summarized to a large extent in chapter 7 (and see also the review by Skalowski, 1969). This brief review will concentrate on the questions of philosophy and methodology of linguistics that are raised by Harris's early chapters and his occasional remarks throughout on the empirical basis, methodological assumptions, and theoretical import of his work.

A word should be said first, however, about the curious position of Harris's work in present-day linguistics. Harris introduced the term transformation into the literature of linguistics, yet today his work is remarkable for the complete lack of contact with what is generally called transformational grammar. This book ignores the Chomskian school of linguistics almost totally, although there are token references to a few of Chomsky's writings at the end of the introduction (4-5), and a footnote on page 30 mentions that category-expanding phrase structure rules are to be found in *Syntactic Structures* (a rather misleading reference, incidentally, to a book which aimed to establish the inadequacy of phrase structure grammars, not to provide
examples of their use). The warning in the preface that the book "reports only the writer's work" must be borne in mind, of course, but in fact it does not mean exactly what it says; an analysis of a Korean sentence by Haeng-Sung Lee is presented on pp.111-113, and there are frequent references to the work of other members and former members of the Pennsylvania school of transformational and discourse analysis such as Gross, Hű, Joshi, Robbins, Sager, and Yamada. The real point is that there is no hint in the book that Harris even reads any work by linguists working elsewhere than in his own department. What is lacking in particular is any trace of comparison, analogy, agreement, or dispute with the MIT approach to transformational generative grammar (with the possible exception of a one-sentence footnote on p.85 distinguishing the sense of the term 'derivation' as used there from the sense in which Chomsky employs it). Even on the minor points of notation and vocabulary, Harris is doggedly different from everybody else: one finds in this book 'J' used for the more usual '*' to mark nonsentences; 'zeroing' for 'deletion'; 'pronoun' for 'pronominalization'; and so on. It is, in fact, hard not to see in Harris's stubborn indifference to MIT linguistics a certain disapproval of Chomsky's work in the philosophy of linguistics.

The key difference between the views of Harris and Chomsky as regards the linguists task is one that does not really separate them at all as regards the substantive problems in grammar which both have to grapple. Harris views the language under study as an enormous corpus of recorded utterances (which he calls sentences most of the time, since he appears not to draw any distinction between the two), and the linguist's job as one of charting the terrain thus delimited. Chomsky, on the other hand, sees the linguist as engaged in constructing a theory which is a formal analogue of the knowledge of language possessed by the native speaker under ideal conditions. These opposing views tend to meet because on the one hand the Chomskian idealization is admitted to be inaccessible, and data consisting of recorded utterances and speakers' judgements about what utterances they would use to express certain ideas are therefore admitted as evidence, and on the other the amassing of a Harriscan super-corpus is admitted to be impracticable, and speakers' intuitions about what utterances would be eventually found in the corpus if the data-gathering went on long enough are therefore admitted within Harris's scheme. Both end up analysing English by reference to what their experience tells them English speakers usually say, or most likely would say.

However, the differences in formal terms between the descriptions they arrive at are considerable, and these differences can frequently be traced back to the procedural basis for Harris's analyses vs. the generative theory underlying Chomsky's. Harris's mathematical structures are to be found in the data and studied for their intrinsic interest by the linguist or the mathematician; Chomsky's are postulated as part of an explanatory theory that belongs ultimately to psychology. Thus Harris's conception of a transformation, for instance, is of an equivalence relation on the set of 'sentences' (strictly, in Harris's terms, the set of propositions, a proposition being a 'sentence' paired with an acceptability grading). Harris's transformations thus impose a partition on the set of utterances in the corpus. "Transformational analysis is thus not primarily an indicator of the structure of each sentence separately, but rather a pairing of sets \{A\}, \{B\}, of sentences,
and so of the corresponding sentences $A_i$, $B_i$ in each set, preserving sentencehood (approximate acceptability as sentence)" (60). Chomskian transformations are nothing like this. They are not relations defined on the set of surface structures, and they are certainly not equivalence relations. They may be, for example, nonreversible: there may be a rule $A \Rightarrow B$ in some language such that the rule $B \Rightarrow A$ is an impossible rule for any language. (WH-REL MOVEMENT in English is such a rule; it transforms strings with the analysis $X [+ \text{wh}] Y$ into strings of the form $[+ \text{wh}] X Y$. Yet the reverse of this is unimaginable. Surely no language has a syntactic rule of the form $aXY \Rightarrow XaY$ where $a$ is some specified element and $X$, $Y$ are variables.)

A further point is that Harris's "transformational analysis" is just that: analysis, not generation of sentence structure. Every sentence in a language is to be decomposable in Harris's framework into "elementary sentences" or "kernel sentences" via the transformations or "operators", and this decomposition must be unique, in the way that every integer is uniquely factorizable into primes (106-111). It is hard to see how such a requirement could even be formulated within current generative theory. There is no provision made for an 'upwards' analysis procedure in a TG grammar (cf. Fisher, 1971, 372-3).

Harris maintains in general a steadfast allegiance to the notion of a discovery procedure that will erect the mathematical structure of a linguistic description directly from observationally accessible data. Thus he outlines (24-8) in brief his recurrent dependence procedure for arriving at morphemic segmentations directly from the facts about what phoneme strings are observed. Roughly, this involves segmenting utterances at the peaks of a graph obtained by plotting (with the utterance spelled out along the abscissa), for each phoneme, the number of phonemes that can succeed that phoneme in other utterances that begin with the same string. Wherever a high peak shows that at this point almost any phoneme could, on statistical grounds, be the immediate successor, there will be a morpheme boundary. This procedure dates, of course, from 1955. Strangely, in the first footnote of Harris (1955), which introduced the procedure to the literature, we read: "I have had the advantage of discussing the subject of this paper with Noam Chomsky" (190); yet in 1958 we find Chomsky stating flatly, "There is no procedure — not a trace of a procedure — for arriving at morphemes from the bottom. And where we can formulate such a procedure even tentatively, the results are unacceptable in the case of languages that we know something about" (Hill, 1962, 172). Why should a procedure that seemed useless to Chomsky in 1958 still seem acceptable to Harris ten years later, if he had been discussing the matter profitably with Chomsky prior to 1955? The answer must surely be that what satisfies Harris as a 'procedure' will not satisfy Chomsky. Chomsky was viewing the goal of structuralist linguistics as the construction of an algorithm for going to morphemes direct from phonemes; Harris's procedure is not algorithmic but probabilistic. (Cf. the frequent locutions of the type "almost always", "proves useful", and so on throughout the book.) The recurrent dependence procedure for morpheme analysis will not segment took into take + PAST, for example; in other words, it has to be supplemented by knowledge about morphological trouble spots gathered by other means anyway, and this was Chomsky's point about the nonexistence of a procedure for finding morpheme boundaries.

It should be noted, incidentally, that the morpheme analysis procedure depends crucially on the fact that not all combinations of phonemes occur within morphemes — or, more generally, at any level of
linguistic description, not all combinations of units occur. This is a principle upon which Harris lays explicit emphasis without ever paying attention to its logical status. It is no doubt true of all languages, of course; but that is a contingent matter, and it does not necessarily follow that, for example, it is this property of language that is the essential one in making language acquisition possible. Harris seems to suggest this when he says (11): "The words and grammatical forms may be taught, in the sense of being singled out, or they may be noticed and finally recognized by the child or other language-learner; but in either case, the possibility of distinguishing the elements requires that not all combinations of elements occur." It is not a necessary truth that a child could not learn a language that contained every mathematically possible combination of elements, merely that he could not learn it using Harris's procedures. By his own admission, Harris is merely exhibiting various mathematical operations that may be performed on phonetically transcribed data. Any leap from this to making claims about necessary conditions for language learning is illegitimate unless it takes the form of proposing a specific, falsifiable hypothesis about language acquisition — one that could in principle at least be disconfimed empirically.

Besides his 'phoneme-to-morpheme' procedure, Harris outlines (21-3) the more familiar procedure of phonemic analysis. Here he is entirely explicit in claiming that a sufficient basis for analysis of phonetic data into phonemes is the relation "is a repetition of", defined on the set of utterances by reference to the results of a straightforward experimental procedure, the pair test. This leads to a statement of equivalence that may be regarded as a rule of correspondence for the theory of phonemics: "utterances A and B are repetitions of each other if and only if they are represented by the same sequence of phonemes" (23). The second half of this statement makes reference to the relation "is the same string of phonemes as", which is clearly just the relation of identity between strings of phonemes, and is an equivalence relation (i.e., is reflexive, symmetric, and transitive). It is easy to show that if we have \( xRy \) and we know that \( R \) is an equivalence relation, then \( R \) is also an equivalence relation. Harris's statement therefore makes the claim that the relation "is a repetition of" is an equivalence relation. But the facts of natural language show that this claim is false. Postal (1968, 216-228) gives examples to show that the condition of transitivity is not met by the relation of repetition or "free variation" (though he mistakenly speaks of the nontransitivity of "contrast" in parts of his argument), and Pullum (1972) provides evidence that the condition of symmetricity is not met either. This grave flaw in the foundations of phonemic theory is not even alluded to in Harris's book although the evidence against phonemics has been piling up in the writings of Chomsky and of Morris Halle since 1960 or earlier.

However, in one sense it is hardly relevant that Harris's procedures for phonemic and morphemic analysis lead to unacceptable results, for they are rather clear cases of what Abercrombie (1953) calls pseudo-procedures anyway. That is, they are stated with a view to providing a rigorous programme of analytical methods that linguists could in principle follow, but with the tacit admission that no linguist will ever in fact follow through the procedures without availing himself of shortcuts. The phoneme-to-morpheme procedure has been given a computer testing (on word-length strings of written English, with the results checked of course against traditional, intuitively established morphemic segmentations) which involved the feeding into a computer of the unabridged Webster as well as several technical dictionaries,
alphabetized forwards and backwards; but to what end is not clear, for no one would attempt the analysis of an undescribed language by such methods, even if some funding agency offered to provide the many years of computer time it would take. We already know so much more about morphological, phonological, and syntactic structure than such an exercise in approximation would teach us that the air of quantified rigour possessed by the description of the procedures is nothing more than a smokescreen for the real, intuition-based business of investigating language structure.

Adherence to a discovery-procedure-based view of the linguist's task is not the only way Harris finds of putting obstacles in his path. Perhaps the most striking example of his strange, maverick empiricism is his willingness to relinquish even the denumerability of the set of utterances. This comes about through his tolerance of 'sentences' like "[ ] is a sound", "[ ] is his name", etc., where [ ] represents any sound whatever, linguistic or not. The set of all sounds that may be made with the vocal apparatus is clearly not only nonfinite but nondenumerable, and thus for Harris so is the set of all sentences in a language; in fact the set of sentences is not even a subset of the set of all word sequences (11). A much more normal view of the delimitation of the set of sentences in a language would be that accepted by, for example, Linsky (1952, 4) when he draws the distinction between objects and the names of objects, and between use and mention of the latter. Of the sentence John is tall Linsky says:

This sentence mentions the person John, but what appears in the sentence is not John but John's name. The sentence mentions John but does not use him; it uses John's name but it does not mention it. Now we might attempt to use John himself in this sentence instead of using his name. We might attempt to do this by placing John himself on this page in approximately the same position now occupied by his name. The result would be a physical object consisting of John followed by the words "is tall." But nobody, I believe, would be inclined to call this a sentence of any language.

I have the same feeling about the sequence of noises consisting of the sound of someone retching followed by the sound of a voice saying "is the sound of someone being sick." But apparently Harris would disagree.

Harris's singular attitude to the use/mention distinction relates to another curious feature of his approach which is encapsulated in the slogan "The metalanguage is in the language" (17, 125-8, et passim). This is not an easy notion to grasp, but Harris appears to be suggesting that all sentences about sentences be included in the enormous corpus that is imagined as the object of study, so that sentences with grammatical peculiarities can be viewed as substrings of larger ones commenting on their peculiarity and containing them. Harris considers this move to be a necessary one if an infinite regress of description is to be avoided, for there would never be an end to the specification of metalanguages if for each language L there was another, L', which was not contained in L but was available for talking about the sentences of L. The language under study is thereby immediately extended to include a denumerably infinite subset of strings such as X is a word where X is any sequence of phonemes (17), or Y is a sentence where Y is any word sequence (125). The use/mention distinction does in fact appear on p.125 in the unlikely guise of "a matter of morphophonemic variation." ('He went' is a sentence is related to the equally possible He went is a sentence by a morphophonemic rule referring to intonation in Harris's system; a more usual approach would be to regard the second string as simply deviant.)
This book is difficult to read and highly technical. The linguist will find in it much that seems to him mystifying, unnecessary, and infuriatingly oblivious to basic objections that have long been raised. But it is also a very important addition to the literature of structural linguistics in the broad sense, in which the linguist will find much that is stimulating, provoking, and insightful. It ought not to be ignored.

G.K. Pullum

REFERENCES


Harris, Z.S. (1955) 'From phoneme to morpheme.' Language 31, 190-222.


Pullum, G.K. (1972) 'Free variation isn't symmetric, either.' IJAL 38, 268-270.