

Zellig S. Harris, *Mathematical Structures of Language* (= *Interscience Tracts in Pure and Applied Mathematics*, 21) (New York, John Wiley & Sons, 1968), pp. 230, index of terms, by Maurice Gross.

Preceded by various papers that gave only fragments of his views, Zellig Harris's book has been long awaited; now *Mathematical Structures of Language* is a landmark in the field of formalized linguistics. This work appears in a context of discoveries that have entirely transformed linguistics in the past fifteen years, and states with a very rare precision the assumptions that now underlie linguistic theory. The book is compact, it contains a lot of extremely valuable material and ideas; we will only review a small part of it.

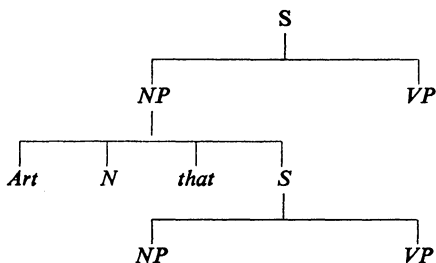
Mathematical linguistics has become popular since about 1960, and Chomsky's study "On Certain Formal Properties of Grammars"¹ has certainly contributed in a decisive way to the development of this field. Previously, some mathematical work had been performed in relation to linguistic observations, but it was mainly of a statistical nature. It consisted in evaluating various numerical parameters relating the size and the uses in texts of a given alphabet or vocabulary (Zipf's law, for example), or in computing certain types of redundancy due to the restricted nature of the sequences of letters or of words occurring in natural languages (Markov's and Shannon's statistics). It is only recently that studies of an algebraic and combinatorial type were introduced in linguistics. They have been motivated by empirical procedures. As a matter of fact, the main activity for a structural linguist consists in discovering the discrete elements of a language, and in studying their combinatorial properties.

¹ This paper contains a mathematical definition and basic properties of the method of syntactic description given by Harris in his fundamental study "From Morpheme to Utterance".

Chomsky's approach focuses on certain essential features of language; some of these features have been studied from an abstract point of view, and they lead to structures that have deep mathematical and linguistic significance. The theoretical views that Harris develops in his book, although based on the same empirical material, are completely new, and the relationship with what is called mathematical structure in Chomsky's framework is quite indirect. In fact, Harris's approach has many facets. While Chomsky made one main hypothesis about language structure, namely, that it is describable by means of certain types of formal systems (in the sense of mathematical logic), Harris looks at various phenomena as being of distinct natures; as such, they require distinct treatments. Thus, for Harris, linguistic theory does not have to be unified for the time being.

Since Chomsky's work has been more widely brought to the knowledge of specialists of linguistics (and of neighboring fields like psychology and philosophy), it is interesting to compare the two points of view.

Chomsky restricts the rules of the grammar to the following two types. Constituent (or formation) rules are noted $A \rightarrow \Psi$, where A is a single symbol holding for a grammatical category, while Ψ represents a sequence of such categories. The arrow roughly means "is constituted by means of". Thus, for example, the rule $S \rightarrow NP VP$ states that any sentence (S) is constituted by a noun phrase (NP) and a verb phrase (VP), occurring in this left-to-right order. Other rules of the same type will state what NP and VP are; for example, we will have the rule $NP \rightarrow Art N that S$ (Article, Noun, the subordinating conjunction *that*, and Sentence, as in *the fact that he did it*). This particular shape of rule allows a tree representation for syntactic structure. The left hand side of each rule is considered as a node, and linked by lines to each symbol (i.e., node) of the corresponding right-hand side. The left-to-right order of the symbols is significant for these trees. The two above examples of rules allow the construction of patterns of structure like:



Such rules are used to display the components of a sentence in a way not obligatorily related to the way they are observed in actual speech or writing. These representations, called deep structures, are defined in such a way that they provide a maximum of simplicity and generality for the overall description of the language. These deep structures are mapped onto actual sentences, namely onto surface structures, by means of rules of the second type, called transformations.

Harris' general picture (in which Chomsky's has its source) presents many similarities. A basic set K of elementary sentences (kernel sentences) corresponds to the deep structures, they are given in lists. A set Φ of transformations acts on them, and on the transforms, yielding subsets of ΦK , $\Phi^2(K)$, ..., $\Phi^n(K)$, etc.

Basically, transformations reshuffle the components of kernel sentences, and display them in the way they are observed in actual sentences. In both frameworks, transformations have a fundamental property, they leave unchanged the meaning of the structures or sentences on which they operate. Harris has observed another important invariant of transformations. When a set of words fills in a structure, and when the corresponding sentence has a certain acceptability, this acceptability is invariant under transformation. Moreover, when a set of sentences can be ordered with respect to acceptability, the transformations preserve the order. The reader will easily recognize this situation in the following series of examples:

	<i>John ate the cake</i>	
(K)	<i>John ate the book</i>	(structure $N \vee N$)
	<i>John ate liberty</i>	

The acceptability decreases from top to bottom, the same is true when the transformations 'passive', and 'operator adjunction' have applied:

	<i>The cake was eaten by John</i>
passive (K)	<i>The book was eaten by John</i>
	<i>Liberty was eaten by John</i>
	<i>I said that John ate the cake</i>
operator (K)	<i>I said that John ate the book</i>
	<i>I said that John ate liberty</i>

Such observations confirm in a very precise way that transformations are formal features of natural languages that can thus be studied independently of the meaning of the words. Moreover, it cannot be argued that acceptability is only a function of word choice, since a sequence like

John cake the ate with correct choice of words, but without any structure, is utterly unacceptable (i.e., has acceptability zero).

A typical example of transformation is the rule that places tenses on verbs. Tense marks in English may have rather different forms, for example they can be endings like *-ed* in *liked*, *-en* in *eaten*, or auxiliaries like *have*, *is*, *will*. Auxiliaries and endings are sometimes linked, sometimes independent. The auxiliary *have* imposes the ending *-ed* (or *-en*) to its verb, *is* imposes *-ing*, *will* does not impose any ending. The ending *-ed* may occur without any auxiliary, this is not the case for *-en* which is generally preceded by an auxiliary verb. This situation is accounted for in the following way, all such elements are given one single name: *Tense*, by means of the formation rule

$$Tense \rightarrow Aux \text{ } -sfx$$

Aux is an auxiliary verb, and *-sfx* (suffix) the ending which is imposed to the main verb, *Aux* and/or *-sfx* can be zero elements. Thus, formation rules will for example yield deep sequences like:

<i>NP</i>	<i>Aux</i>	<i>-sfx</i>	<i>V</i>
<i>The man</i>	<i>has</i>	<i>-en</i>	<i>eat</i>
<i>The man</i>	<i>is</i>	<i>-ing</i>	<i>eat</i>

Then, a permutation transformation will place the suffix in the correct place, namely to the right of the verb *V*.

In Chomsky's framework, a transformation is defined on a tree, more exactly on an 'analysis' of a tree, namely a sequence of nodes that corresponds in a unique way to the elements of the sentence. The suffix permuting rule will be approximately:

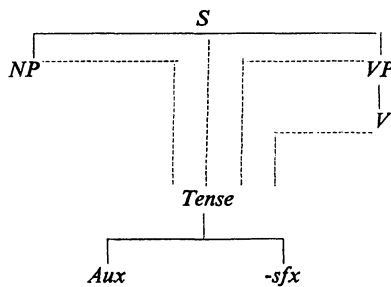
	<i>NP</i>	<i>Aux</i>	<i>sfx</i>	<i>V</i>	<i>X</i>
	1	2	3	4	5
$\rightarrow$	1	2	4	3	5

The elements of the sentence are numbered from left to right, 5 corresponds to the variable *X* that ranges over the rest of the components of the sentence. Here, the interpretation of the arrow is different, it means 'is transformed into'. To the right of the arrow, we find the changes that the rule determines. For Harris, the transformation is a rule that applies to a string, and that exchanges the positions of *-sfx* and *V*, it can be written $-sfx \ V \rightarrow V sfx$.

One of the main differences between the two systems will immediately be noticed by the reader. Harris never uses tree representations for

sentence structure. All examples of representations are given in terms of sequences of symbols. Even more, the notion of phrase is not explicitly mentioned. Thus, Harris does not make any difference between noun phrases and nouns, noun phrases are nouns to which various adjunctions are made that do not change very much the initial status of the noun. A notion like *NP* is then a notion derived from basic notions like *N* and adjunctions. On the other hand, for Chomsky, *NP* and *N* are notions defined on the same level (auxiliary vocabulary), and their dependence is only indirectly indicated by means of rules where both symbols appear. As a consequence, the symbol *NP* is excluded by Harris in the description of structures and rules. In a related way, Harris considers that there is not enough motivation to use the symbol *VP* (verb phrase or predicate), he instead used the notation $V\Omega$, where Ω ranges over the possible complements of *V*; when necessary Ω is explicited into *N* (direct complement of *V*), or *PN* (indirect complement preceded by the preposition *P*), or into combinations of these types. It is important to notice that from the Chomskyan point of view, the notions *NP* and *VP*, although constantly used in tree representations, raise many unsolved questions. For example, there are ways in which the notion *VP* can be justified. It is true that subjects have very special properties when compared to complements, the constraint of agreement with the verb is such an example. So, to a certain extent, a dichotomy like subject-rest of the sentence is justified. However, there are questions that have to be raised, and that have not yet (to the best of our knowledge) received any answer. An actual sentence, and its surface structure K_s (*s* in an integer) are rather remote from the corresponding deep structure K_0 . The link between them is a chain of *s* transformations. Each transformation applies to a tree structure K_i , and the result is a new tree structure K_{i+1} . The main problem with *VP* is the following, at which levels K_j : $0 \leq j \leq s$ is *VP* necessary? The arguments for having *VP* in K_s , or in K_0 , or in any K_j should be different. One cannot use the existence of an agreement between subject and verb for K_0 ; agreement phenomena are known to be rather superficial. In the same way, the semantic or logical notions of subject and predicate may be used in K_0 , but not in K_s .

The example of tense placement also shows that it can be very difficult to justify precise shapes for tree structures. The way in which the description of tenses is always presented leads us to the following situation in terms of trees.



The subtree headed by *Tense* occurs between the symbols *NP* and *V*; thus, *a priori*, *Tense* could be attached in any of the ways indicated by the dotted lines, without affecting the arguments that we used for placing it between *NP* and *V*.

We can justify very strongly that at a certain level of depth (i.e., for some $K_j : j < s$), the element *tense* (i.e., *Aux -sfx*) has to occur between subject and verb. In terms of trees, we have many questions to answer: for each K_j ($0 \leq j \leq s$), to which node (*NP*, *S*, *VP*, *V*, ...) are we going to attach *Tense*? It is not unreasonable to attach *Tense* to *S*, at least at a deep level. Semantically, the time modality that corresponds to *Tense* modifies the whole real-life situation that corresponds to the sentence *S*, and not a subpart of it. It is also a traditional view to attach tenses to verbs or verb phrases; it is of course quite natural to 'pack' together a verb and its suffix, but such a solution is not obvious for auxiliaries, since the "degree of attachment" between an auxiliary and a verb is not the same as between a verb and a suffix. For example, adverbs can be inserted between *Aux* and *V* but not between *V* and *-sfx*. This fact could justify the attachment of *-sfx* to *V*, and of *Aux* to *VP* in some K_j , but it does not bring any solution for the attachment of *Tense*. One could think at first sight that the node *NP* (subject) is easy to reject as a candidate node for the attachment of *Tense*, however it is not so. The verb suffix *-s* (3rd person singular, present) belongs, from the morphological point of view, to the tense system. But presumably, it has its source in the subject *NP*, and as such is initially attached to this *NP*. Later it is moved to the right of the verb, and in order to have a maximum of generality, one should use for *-s* the same permutation rule as with the other suffixes (*-ed*, *-en*, *-ing*). If *-s* is attached to *NP* it might be tempting to attach other suffixes to *NP*, this would allow the suffix rule to apply in exactly the same conditions for *-s*, *-ed*, etc. Moreover, there are deeper facts that indicate the existence of constraints between subjects and tenses. For example, the sentence:

Some money interests John

is hard to accept (much harder than: *Money interests John*) while the sentence:

Some money would interes John

is quite natural. This pair shows a dependence between the determiner of the subject, and the auxiliary of the verb. Thus, even for such a common phenomenon as tense representation, it is very hard to determine the various patterns of tree structure that are needed between deep and surface structure.

It seems very likely that certain of the questions that are raised are strictly linked to the nature of the formal apparatus, which means that it will be very difficult, if possible at all, to justify empirically the choice of a solution. In certain cases, the feeling is that the questions have a parasitic nature, namely, that no empirical data can be linked to them, which could be due to the fact that the theory is not quite adequate. Harris, by not using the device of tree representations, avoids many problems that have a very artificial flavor, but that are central in Chomsky's theory. With less formal apparatus, Harris forces himself to stay much closer to the empirical data which have so far proved to be the most important ones, nor does he have to work on problems that may be extremely artificial. We can remark that the most controversial part of Chomsky's theory is the one that deals with the form and the role of the deep structure. The problems debated in this context are in most cases (if not in all) very remote from the linguistic data.

Another important feature of Harris' transformations is that they are not *a priori* oriented. If f_1 and f_2 are sentences that present a transformational relation, Harris will write $f_1 \leftrightarrow f_2$. Thus, asking whether one derives f_2 from f_1 , or f_1 from f_2 has not much meaning. There is, however, some ordering between related sentences, transformations usually modify the complexity of the sentences to which they apply. High complexity is sometimes hard to characterize, it may consist in morpheme adjunction, but zeroing transformations, by making sentences less explicit and less regular with respect to their elementary components, also increase some kind of complexity. Thus, transformations could be ordered with respect to a notion of increasing complexity, and when they apply to actual sentences, undoing them into simpler sentences, the final result is a pattern of kernel sentences. Such a notion of ordering is in sharp contrast with the notion of ordering in generative grammar, where linguists have always to decide about the direction of a derivation

(i.e., whether $f_1 \rightarrow f_2$, or else $f_2 \rightarrow f_1$). From the empirical point of view, it turns out that there is not much difference between oriented, and non-oriented transformations. From the formal point of view, the situation is very reminiscent of a certain equivalence status between combinatorial systems. We know that given a terminal and a nonterminal vocabulary ($V_T \cup V_N$), a semi-Thue System is a set of n rules $\{\varphi_i \rightarrow \Psi_i : 1 \leq i \leq n\}$ on this vocabulary that are used to derive a language from a finite set of axioms A . A Thue system is composed of similar elements, but the rules are unordered, they write $\varphi_i \leftrightarrow \Psi_i$, and they are equivalent to two rules ($\varphi_i \rightarrow \Psi_i$ and $\Psi_i \rightarrow \varphi_i$) of the preceding type. However, it can be proved that whatever language can be described by one type of system can also be described by the other. The impression given by both types of transformations is very similar: any describable syntactic phenomenon seems describable by either system. There might be a way in which one could say that nonoriented rules are more abstract than those of the other type. Thue systems (nonoriented rules), are very close to certain abstract algebraic systems (congruence relations over free monoids), while oriented rules are much closer to computational devices, thus more concrete. But such a theoretical difference has no obvious implication for the actual linguistic work.

We said before, when we discussed the use of tree structures, that Harris was staying very close to the empirical data. On the other hand, when he is concerned with the products of transformations for a whole language, he deals with a very abstract level of linguistic description. Harris supposes that a certain number of theoretical and empirical questions are solved, he then studies a system of products of operators. The fact that certain questions are still pending does not affect his general model; the framework and the types of solutions he gives as examples can very plausibly be extended to a complete analysis of English (and of many other languages). Even counterexamples or alternatives to his analyses would not change the general structure of the model. Still, an important amount of work of regularization has to be done. From a mathematical point of view, it is very hard to deal with a set Φ of operators whose products are not all defined. There are different situations that complicate a possible abstract study. There are products like $\varphi_i \varphi_j$ of two elements of Φ that are not defined, but at the same time, products like $\varphi_i \varphi_j \varphi_k$ (with same i and j) are defined. Moreover these restrictions vary according to the nature of the kernel sentence (S) to which these products apply. This situation is limited to products of finite length, the algebra of transformations can thus be presented by means of tables. Few mathe-

mathematical properties can be given for such systems, one of the reasons is that they have been defined by Harris for the first time. Observations like the fact that some products are Abelian, some are not (the corresponding transformations are unordered or not), some are associative, and some are not, indicate a high degree of complexity. It is interesting to note that Harris can motivate certain regularizations from an empirical point of view. By extending the notion of transformation, he is able to show that products like the above undefined $\varphi_i \varphi_j$ may be considered as defined. So far, the main use of these operations is the description of the transformational history of sentences by means of special graphs that show how elementary units (morphemes, kernel sentences) are compounded. These graphs appear to constitute an important invariant with respect to the meaning of sentences, whether they are considered within a particular language like English, or between pairs of unrelated languages like English and Korean. The invariance of the shape of these graphs is exemplified by two conceptions of transformations that Harris presents.

On the one hand, Harris uses global transformations like passive, nominalizations, conjunctions. On the other hand, such transformations are in fact types of transformations, and each of them can sometimes take slightly different forms. This is the case, for example, with certain nominalizations (p. 64) that are on the whole very similar, but that may differ with respect to lexical insertion of constants like suffixes (*-ing*, *-er*, etc.). In such a situation, the related transformations are broken into several operators: one is a common part shared by the whole family, the others correspond to the differences observed between them. Whether one uses in a graph the overall transformations or elementary types of operators does not affect the general shape of the graph of a sentence.

Other mathematical devices are used by Harris for the description of very general properties of natural languages. He defines a stochastic process for the segmentation of utterances into morphemes that is based on the fact that at the juncture between two morphemes the occurrences of the two phonemes appearing on each side of the boundary are independent, while two adjacent phonemes within a morpheme may be strongly dependent. Harris also uses for superficial syntactic structure a cancellation system, where elements corresponding to grammatical categories have right and/or left inverses, according to the various types of adjunctions these categories may have. For example, a noun is represented by ' $TN(N$, and right inverse of T ', category of the article). The sequence $T''T N$ (article, noun) cancels to N . The system is different

from Bar-Hillel's and Lambeck's, and is worth further mathematical investigation; in particular the notions of "delay", and "conjugate elements" (p. 43) make the family of languages defined by this mechanism different from the context-free languages.²

The book may surprise certain readers. Traditional linguists may not find in it the kind of material they are used to dealing with. Even the specialists of generative grammar may not, at first sight, recognize its interest. There are various reasons. For example, in many linguistic analyses Harris gives a central role (and has been giving it for a long time) to special words: verbs like *to equal*, *to exceed* in comparative sentences, *used for*, *made for* in nominalizations, and many terms like *occur*, *process*, *quantity*, *measure*, and other noun classifiers. It may seem that there is no other use for these words than allowing longer or more explicit paraphrases of certain sentences in order to analyze them by deletion. We know that in each case the choice made is very precise, and largely motivated by reasons that do not appear in this book. Harris's intention is to publish elsewhere these results, which, although very important, have no direct relevance to the theoretical framework, which is the matter of the present book.

As for mathematicians, they will not find axiomatized theories with proved theorems, or even precise definitions that promise immediate new results, but they will be given a general framework for mathematical studies that have linguistic significance. This book will affect many generations of linguists, it is the sign that linguistics has become an autonomous scientific domain, with its own methods, where formal features have been extracted that already indicate many important directions of investigations, both empirical and theoretical.

REFERENCES

- Bar-Hillel, Yehoshua,
 1953 "A Quasi Arithmetical Notation for Syntactic Description", *Language* 29,
 47-58.
- Chomsky, Noam,
 1959 On Certain Formal Properties of Grammars", *Information and Control* 2,
 137-167.
- 1965 *Aspects of the Theory of Syntax* (Cambridge, Mass.: M.I.T. Press).

² An efficient method for comparing these various devices consists in showing their equivalence with Schützenberger's algebraic characterization of context-free languages.

- Harris, Zellig S.,
 1946 "From Morpheme to Utterance", *Language* 22, 161-183.
 1951 *Methods in Structural Linguistics* (Chicago: University of Chicago Press).
 1957 "Co-occurrence and Transformations in Linguistic Structure", *Language* 33, 283-230.
 1964 "Elementary Transformations" (*Transformations and Discourse Analysis Projects*, 54) (University of Pennsylvania).
 1966 "A Cycling Cancellation Automaton for Sentence Well-Formedness", *I.C.C. Bulletin* 5, 2.
- Lambek, Joachim,
 1960 "On the Calculus of Syntactic Types", *Structure of Language and its Mathematical Aspects* (= *Proceedings of Symposia in Applied Mathematics* 12), ed. by Roman Jakobson (Providence: American Mathematical Society).
- Markov, A. A.,
 1913 "Essai d'une recherche statistique sur le texte du roman 'Eugène Onéguine'", *Bulletin de l'Académie Impériale des Sciences de Saint-Petersbourg* 7.
- Schützenberger, M. P.,
 1963 "On Context-Free Languages and Pushdown Automata", *Information and Control* 6, 3.
- Shannon, C. E.,
 1951 "Prediction and Entropy of Printed English", *Bell System Tech. Journal* 30, 50-64.

Maurice Gross, born 1934, is Head of the Laboratoire d'Automatique Documentaire et Linguistique (CNRS), and teaches at the Centre Universitaire de Vincennes (Paris). His fields of research are general and French linguistics. Among his latest publications are: (with A. Lentin) *Notions sur les grammaires formelles* (1966); *Grammaire transformationnelle du français* (1968).