

Linguistics Department
University of Maryland
1401 Marie Mount Hall
College Park, MD 20742-7515

ASPECTS OF THE THEORY OF SYNTAX

Noam Chomsky



THE M.I.T. PRESS

*Massachusetts Institute of Technology
Cambridge, Massachusetts*

I

Methodological Preliminaries

§ 1. *GENERATIVE GRAMMARS AS THEORIES OF LINGUISTIC COMPETENCE*

THIS study will touch on a variety of topics in syntactic theory and English syntax, a few in some detail, several quite superficially, and none exhaustively. It will be concerned with the syntactic component of a generative grammar, that is, with the rules that specify the well-formed strings of minimal syntactically functioning units (*formatives*) and assign structural information of various kinds both to these strings and to strings that deviate from well-formedness in certain respects.

The general framework within which this investigation will proceed has been presented in many places, and some familiarity with the theoretical and descriptive studies listed in the bibliography is presupposed. In this chapter, I shall survey briefly some of the main background assumptions, making no serious attempt here to justify them but only to sketch them clearly.

Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. This seems to me to have been the position of the founders of modern general linguistics, and no cogent reason for

modifying it has been offered. To study actual linguistic performance, we must consider the interaction of a variety of factors, of which the underlying competence of the speaker-hearer is only one. In this respect, study of language is no different from empirical investigation of other complex phenomena.

We thus make a fundamental distinction between *competence* (the speaker-hearer's knowledge of his language) and *performance* (the actual use of language in concrete situations). Only under the idealization set forth in the preceding paragraph is performance a direct reflection of competence. In actual fact, it obviously could not directly reflect competence. A record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on. The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance. Hence, in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior.¹ Observed use of language or hypothesized dispositions to respond, habits, and so on, may provide evidence as to the nature of this mental reality, but surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline. The distinction I am noting here is related to the *langue-parole* distinction of Saussure; but it is necessary to reject his concept of *langue* as merely a systematic inventory of items and to return rather to the Humboldtian conception of underlying competence as a system of generative processes. For discussion, see Chomsky (1964).

A grammar of a language purports to be a description of the ideal speaker-hearer's intrinsic competence. If the grammar is, furthermore, perfectly explicit — in other words, if it does not rely on the intelligence of the understanding reader but rather provides an explicit analysis of his contribution — we may (somewhat redundantly) call it a *generative grammar*.

A fully adequate grammar must assign to each of an infinite range of sentences a structural description indicating how this

sentence is understood by the ideal speaker-hearer. This is the traditional problem of descriptive linguistics, and traditional grammars give a wealth of information concerning structural descriptions of sentences. However, valuable as they obviously are, traditional grammars are deficient in that they leave unexpressed many of the basic regularities of the language with which they are concerned. This fact is particularly clear on the level of syntax, where no traditional or structuralist grammar goes beyond classification of particular examples to the stage of formulation of generative rules on any significant scale. An analysis of the best existing grammars will quickly reveal that this is a defect of principle, not just a matter of empirical detail or logical preciseness. Nevertheless, it seems obvious that the attempt to explore this largely uncharted territory can most profitably begin with a study of the kind of structural information presented by traditional grammars and the kind of linguistic processes that have been exhibited, however informally, in these grammars.²

The limitations of traditional and structuralist grammars should be clearly appreciated. Although such grammars may contain full and explicit lists of exceptions and irregularities, they provide only examples and hints concerning the regular and productive syntactic processes. Traditional linguistic theory was not unaware of this fact. For example, James Beattie (1788) remarks that

Languages, therefore, resemble men in this respect, that, though each has peculiarities, whereby it is distinguished from every other, yet all have certain qualities in common. The peculiarities of individual tongues are explained in their respective grammars and dictionaries. Those things, that all languages have in common, or that are necessary to every language, are treated of in a science, which some have called *Universal* or *Philosophical* grammar.

Somewhat earlier, Du Marsais defines universal and particular grammar in the following way (1729; quoted in Sahlin, 1928, pp. 29-30):

Il y a dans la grammaire des observations qui conviennent à toutes les langues; ces observations forment ce qu'on appelle la grammaire

générale: telles sont les remarques que l'on a faites sur les sons articulés, sur les lettres qui sont les signes de ces sons; sur la nature des mots, et sur les différentes manières dont ils doivent être ou arrangés ou terminés pour faire un sens. Outre ces observations générales, il y en a qui ne sont propres qu'à une langue particulière; et c'est ce qui forme les grammaires particulières de chaque langue.

Within traditional linguistic theory, furthermore, it was clearly understood that one of the qualities that all languages have in common is their "creative" aspect. Thus an essential property of language is that it provides the means for expressing indefinitely many thoughts and for reacting appropriately in an indefinite range of new situations (for references, cf. Chomsky, 1964, forthcoming). The grammar of a particular language, then, is to be supplemented by a universal grammar that accommodates the creative aspect of language use and expresses the deep-seated regularities which, being universal, are omitted from the grammar itself. Therefore it is quite proper for a grammar to discuss only exceptions and irregularities in any detail. It is only when supplemented by a universal grammar that the grammar of a language provides a full account of the speaker-hearer's competence.

Modern linguistics, however, has not explicitly recognized the necessity for supplementing a "particular grammar" of a language by a universal grammar if it is to achieve descriptive adequacy. It has, in fact, characteristically rejected the study of universal grammar as misguided; and, as noted before, it has not attempted to deal with the creative aspect of language use. It thus suggests no way to overcome the fundamental descriptive inadequacy of structuralist grammars.

Another reason for the failure of traditional grammars, particular or universal, to attempt a precise statement of regular processes of sentence formation and sentence interpretation lay in the widely held belief that there is a "natural order of thoughts" that is mirrored by the order of words. Hence, the rules of sentence formation do not really belong to grammar but to some other subject in which the "order of thoughts" is studied. Thus in the *Grammaire générale et raisonnée* (Lancelot

et al., 1660) it is asserted that, aside from figurative speech, the sequence of words follows an "ordre naturel," which conforms "à l'expression naturelle de nos pensées." Consequently, few grammatical rules need be formulated beyond the rules of ellipsis, inversion, and so on, which determine the figurative use of language. The same view appears in many forms and variants. To mention just one additional example, in an interesting essay devoted largely to the question of how the simultaneous and sequential array of ideas is reflected in the order of words, Diderot concludes that French is unique among languages in the degree to which the order of words corresponds to the natural order of thoughts and ideas (Diderot, 1751). Thus "quel que soit l'ordre des termes dans une langue ancienne ou moderne, l'esprit de l'écrivain a suivi l'ordre didactique de la syntaxe française" (p. 390); "Nous disons les choses en français, comme l'esprit est forcé de les considérer en quelque langue qu'on écrive" (p. 371). With admirable consistency he goes on to conclude that "notre langue *pédestre* a sur les autres l'avantage de l'utile sur l'agréable" (p. 372); thus French is appropriate for the sciences, whereas Greek, Latin, Italian, and English "sont plus avantageuses pour les lettres." Moreover,

le bon sens choisirait la langue française; mais . . . l'imagination et les passions donneront la préférence aux langues anciennes et à celles de nos voisins . . . il faut parler français dans la société et dans les écoles de philosophie; et grec, latin, anglais, dans les chaires et sur les théâtres; . . . notre langue sera celle de la vérité, si jamais elle revient sur la terre; et . . . la grecque, la latine et les autres seront les langues de la fable et du mensonge. Le français est fait pour instruire, éclairer et convaincre; le grec, le latin, l'italien, l'anglais, pour persuader, émouvoir et tromper: parlez grec, latin, italien au peuple; mais parlez français au sage. (pp. 371-372)

In any event, insofar as the order of words is determined by factors independent of language, it is not necessary to describe it in a particular or universal grammar, and we therefore have principled grounds for excluding an explicit formulation of syntactic processes from grammar. It is worth noting that this naïve view of language structure persists to modern times in

various forms, for example, in Saussure's image of a sequence of expressions corresponding to an amorphous sequence of concepts or in the common characterization of language use as merely a matter of use of words and phrases (for example, Ryle, 1953).

But the fundamental reason for this inadequacy of traditional grammars is a more technical one. Although it was well understood that linguistic processes are in some sense "creative," the technical devices for expressing a system of recursive processes were simply not available until much more recently. In fact, a real understanding of how a language can (in Humboldt's words) "make infinite use of finite means" has developed only within the last thirty years, in the course of studies in the foundations of mathematics. Now that these insights are readily available it is possible to return to the problems that were raised, but not solved, in traditional linguistic theory, and to attempt an explicit formulation of the "creative" processes of language. There is, in short, no longer a technical barrier to the full-scale study of generative grammars.

Returning to the main theme, by a generative grammar I mean simply a system of rules that in some explicit and well-defined way assigns structural descriptions to sentences. Obviously, every speaker of a language has mastered and internalized a generative grammar that expresses his knowledge of his language. This is not to say that he is aware of the rules of the grammar or even that he can become aware of them, or that his statements about his intuitive knowledge of the language are necessarily accurate. Any interesting generative grammar will be dealing, for the most part, with mental processes that are far beyond the level of actual or even potential consciousness; furthermore, it is quite apparent that a speaker's reports and viewpoints about his behavior and his competence may be in error. Thus a generative grammar attempts to specify what the speaker actually knows, not what he may report about his knowledge. Similarly, a theory of visual perception would attempt to account for what a person actually sees and the mechanisms that determine this rather than his statements about what he sees and why, though these state-

ments may provide useful, in fact, compelling evidence for such a theory.

To avoid what has been a continuing misunderstanding, it is perhaps worth while to reiterate that a generative grammar is not a model for a speaker or a hearer. It attempts to characterize in the most neutral possible terms the knowledge of the language that provides the basis for actual use of language by a speaker-hearer. When we speak of a grammar as generating a sentence with a certain structural description, we mean simply that the grammar assigns this structural description to the sentence. When we say that a sentence has a certain derivation with respect to a particular generative grammar, we say nothing about how the speaker or hearer might proceed, in some practical or efficient way, to construct such a derivation. These questions belong to the theory of language use—the theory of performance. No doubt, a reasonable model of language use will incorporate, as a basic component, the generative grammar that expresses the speaker-hearer's knowledge of the language; but this generative grammar does not, in itself, prescribe the character or functioning of a perceptual model or a model of speech production. For various attempts to clarify this point, see Chomsky (1957), Gleason (1961), Miller and Chomsky (1963), and many other publications.

Confusion over this matter has been sufficiently persistent to suggest that a terminological change might be in order. Nevertheless, I think that the term "generative grammar" is completely appropriate, and have therefore continued to use it. The term "generate" is familiar in the sense intended here in logic, particularly in Post's theory of combinatorial systems. Furthermore, "generate" seems to be the most appropriate translation for Humboldt's term *erzeugen*, which he frequently uses, it seems, in essentially the sense here intended. Since this use of the term "generate" is well established both in logic and in the tradition of linguistic theory, I can see no reason for a revision of terminology.

§ 2. TOWARD A THEORY OF PERFORMANCE

There seems to be little reason to question the traditional view that investigation of performance will proceed only so far as understanding of underlying competence permits. Furthermore, recent work on performance seems to give new support to this assumption. To my knowledge, the only concrete results that have been achieved and the only clear suggestions that have been put forth concerning the theory of performance, outside of phonetics, have come from studies of performance models that incorporate generative grammars of specific kinds — that is, from studies that have been based on assumptions about underlying competence.³ In particular, there are some suggestive observations concerning limitations on performance imposed by organization of memory and bounds on memory, and concerning the exploitation of grammatical devices to form deviant sentences of various types. The latter question is one to which we shall return in Chapters 2 and 4. To clarify further the distinction between competence and performance, it may be useful to summarize briefly some of the suggestions and results that have appeared in the last few years in the study of performance models with limitations of memory, time, and access.

For the purposes of this discussion, let us use the term “acceptable” to refer to utterances that are perfectly natural and immediately comprehensible without paper-and-pencil analysis, and in no way bizarre or outlandish. Obviously, acceptability will be a matter of degree, along various dimensions. One could go on to propose various operational tests to specify the notion more precisely (for example, rapidity, correctness, and uniformity of recall and recognition, normalcy of intonation).⁴ For present purposes, it is unnecessary to delimit it more carefully. To illustrate, the sentences of (1) are somewhat more acceptable, in the intended sense, than those of (2):

- (1) (i) I called up the man who wrote the book that you told me about
(ii) quite a few of the students who you met who come from New York are friends of mine

- (iii) John, Bill, Tom, and several of their friends visited us last night
- (2) (i) I called the man who wrote the book that you told me about up
- (ii) the man who the boy who the students recognized pointed out is a friend of mine

The more acceptable sentences are those that are more likely to be produced, more easily understood, less clumsy, and in some sense more natural.⁵ The unacceptable sentences one would tend to avoid and replace by more acceptable variants, wherever possible, in actual discourse.

The notion "acceptable" is not to be confused with "grammatical." Acceptability is a concept that belongs to the study of performance, whereas grammaticalness belongs to the study of competence. The sentences of (2) are low on the scale of acceptability but high on the scale of grammaticalness, in the technical sense of this term. That is, the generative rules of the language assign an interpretation to them in exactly the way in which they assign an interpretation to the somewhat more acceptable sentences of (1). Like acceptability, grammaticalness is, no doubt, a matter of degree (cf. Chomsky, 1955, 1957, 1961), but the scales of grammaticalness and acceptability do not coincide. Grammaticalness is only one of many factors that interact to determine acceptability. Correspondingly, although one might propose various operational tests for acceptability, it is unlikely that a necessary and sufficient operational criterion might be invented for the much more abstract and far more important notion of grammaticalness. The unacceptable grammatical sentences often cannot be used, for reasons having to do, not with grammar, but rather with memory limitations, intonational and stylistic factors, "iconic" elements of discourse (for example, a tendency to place logical subject and object early rather than late; cf. note 32, Chapter 2, and note 9, Chapter 3), and so on. Note that it would be quite impossible to characterize the unacceptable sentences in grammatical terms. For example, we cannot formulate particular rules of the grammar in such a way as

to exclude them. Nor, obviously, can we exclude them by limiting the number of reapplications of grammatical rules in the generation of a sentence, since unacceptability can just as well arise from application of distinct rules, each being applied only once. In fact, it is clear that we can characterize unacceptable sentences only in terms of some "global" property of derivations and the structures they define — a property that is attributable, not to a particular rule, but rather to the way in which the rules interrelate in a derivation.

This observation suggests that the study of performance could profitably begin with an investigation of the acceptability of the simplest formal structures in grammatical sentences. The most obvious formal property of utterances is their bracketing into constituents of various types, that is, the "tree structure" associated with them. Among such structures we can distinguish various kinds — for example, those to which we give the following conventional technical names, for the purposes of this discussion:

- (3) (i) nested constructions
- (ii) self-embedded constructions
- (iii) multiple-branching constructions
- (iv) left-branching constructions
- (v) right-branching constructions

The phrases *A* and *B* form a nested construction if *A* falls totally within *B*, with some nonnull element to its left within *B* and some nonnull element to its right within *B*. Thus the phrase "the man who wrote the book that you told me about" is nested in the phrase "called the man who wrote the book that you told me about up," in (2i). The phrase *A* is self-embedded in *B* if *A* is nested in *B* and, furthermore, *A* is a phrase of the same type as *B*. Thus "who the students recognized" is self-embedded in "who the boy who the students recognized pointed out," in (2ii), since both are relative clauses. Thus nesting has to do with bracketing, and self-embedding with labeling of brackets as well. A multiple-branching construction is one with no internal structure. In (1iii), the Subject Noun Phrase is multiple-branch-

ing, since "John," "Bill," "Tom," and "several of their friends" are its immediate constituents, and have no further association among themselves. In terms of bracketing, a multiple-branching construction has the form $[[A][B]\cdots[M]]$. A left-branching structure is of the form $[[[\cdots]\cdots]\cdots]$ — for example, in English, such indefinitely iterable structures as $[[[[[John]'s\ brother]'s\ father]'s\ uncle]$ or $[[[the\ man\ who\ you\ met]\ from\ Boston]\ who\ was\ on\ the\ train]$, or (iii), which combines several kinds of left-branching. Right-branching structures are those with the opposite property — for example, the Direct-Object of (ii) or $[this\ is\ [the\ cat\ that\ caught\ [the\ rat\ that\ stole\ the\ cheese]]]$.

The effect of these superficial aspects of sentence structure on performance has been a topic of study since almost the very inception of recent work on generative grammar, and there are some suggestive observations concerning their role in determining acceptability (that is, their role in limiting performance). Summarizing this work briefly, the following observations seem plausible:

- (4) (i) repeated nesting contributes to unacceptability
- (ii) self-embedding contributes still more radically to unacceptability
- (iii) multiple-branching constructions are optimal in acceptability
- (iv) nesting of a long and complex element reduces acceptability
- (v) there are no clear examples of unacceptability involving only left-branching or only right-branching, although these constructions are unnatural in other ways — thus, for example, in reading the right-branching construction "this is the cat that caught the rat that stole the cheese," the intonation breaks are ordinarily inserted in the wrong places (that is, after "cat" and "rat," instead of where the main brackets appear)

In some measure, these phenomena are easily explained. Thus it is known (cf. Chomsky, 1959*a*; and for discussion, Chomsky, 1961, and Miller and Chomsky, 1963) that an optimal perceptual

device, even with a bounded memory, can accept unbounded left-branching and right-branching structures, though nested (hence ultimately self-embedded) structures go beyond its memory capacity. Thus case (4i) is simply a consequence of finiteness of memory, and the unacceptability of such examples as (2ii) raises no problem.

If (4ii) is correct,⁶ then we have evidence for a conclusion about organization of memory that goes beyond the triviality that it must be finite in size. An optimal finite perceptual device of the type discussed in Chomsky (1959a) need have no more difficulty with self-embedding than with other kinds of nesting (see Bar-Hillel, Kasher, and Shamir, 1963, for a discussion of this point). To account for the greater unacceptability of self-embedding (assuming this to be a fact), we must add other conditions on the perceptual device beyond mere limitation of memory. We might assume, for example, that the perceptual device has a stock of analytic procedures available to it, one corresponding to each kind of phrase, and that it is organized in such a way that it is unable (or finds it difficult) to utilize a procedure φ while it is in the course of executing φ . This is not a necessary feature of a perceptual model, but it is a rather plausible one, and it would account for (4ii). See, in this connection, Miller and Isard (1964).

The high acceptability of multiple-branching, as in case (4iii), is easily explained on the rather plausible assumption that the ratio of number of phrases to number of formatives (the node-to-terminal node ratio, in a tree-diagram of a sentence) is a rough measure of the amount of computation that has to be performed in analysis. Thus multiple coordination would be the simplest kind of construction for an analytic device — it would impose the least strain on memory.⁷ For discussion, see Miller and Chomsky (1963).

Case (4iv) suggests decay of memory, perhaps, but raises unsolved problems (see Chomsky, 1961, note 19).

Case (4v) follows from the result about optimal perceptual models mentioned earlier. But it is unclear why left- and right-branching structures should become unnatural after a certain point, if they actually do.⁸

One might ask whether attention to less superficial aspects of grammatical structure than those of (3) could lead to somewhat deeper conclusions about performance models. This seems entirely possible. For example, in Miller and Chomsky (1963) some syntactic and perceptual considerations are adduced in support of a suggestion (which is, to be sure, highly speculative) as to the somewhat more detailed organization of a perceptual device. In general, it seems that the study of performance models incorporating generative grammars may be a fruitful study; furthermore, it is difficult to imagine any other basis on which a theory of performance might develop.

There has been a fair amount of criticism of work in generative grammar on the grounds that it slights study of performance in favor of study of underlying competence. The facts, however, seem to be that the only studies of performance, outside of phonetics (but see note 3), are those carried out as a by-product of work in generative grammar. In particular, the study of memory limitations just summarized and the study of deviation from rules, as a stylistic device, to which we return in Chapters 2 and 4, have developed in this way. Furthermore, it seems that these lines of investigation can provide some insight into performance. Consequently, this criticism is unwarranted, and, furthermore, completely misdirected. It is the descriptivist limitation-in-principle to classification and organization of data, to "extracting patterns" from a corpus of observed speech, to describing "speech habits" or "habit structures," insofar as these may exist, etc., that precludes the development of a theory of actual performance.

§ 3. *THE ORGANIZATION OF A GENERATIVE GRAMMAR*

Returning now to the question of competence and the generative grammars that purport to describe it, we stress again that knowledge of a language involves the implicit ability to understand indefinitely many sentences.⁹ Hence, a generative grammar must be a system of rules that can iterate to generate an in-