

A Mouton Classic

"I had already decided I wanted to be a linguist when I discovered this book. But it is unlikely that I would have stayed in the field without it. It has been the single most inspiring book on linguistics in my whole career."

Henk van Riemsdijk

ISBN 3-11-017279-8

www.deGruyter.com

Noam Chomsky

Syntactic Structures

W
DE
G

Noam Chomsky
Syntactic
Structures

mouton

de gruyter

Syntactic Structures



Syntactic Structures

by

Noam Chomsky

Second Edition

With an Introduction by David W. Lightfoot

Mouton de Gruyter
Berlin · New York

Mouton de Gruyter
Berlin · New York 2002

- 1995 *The Minimalist Program*. Cambridge, MA: MIT Press.
 2000 *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- Chomsky, N. and M. Halle
 1968 *The sound pattern of English*. New York: Harper and Row.
- Chomsky, N. and H. Lasnik
 1977 Filters and control. *Linguistic Inquiry* 8.3: 425–504.
- Clark, R.
 1992 The selection of syntactic knowledge. *Language Acquisition* 2: 83–149.
- Dresher, B. E.
 1999 Charting the learning path: Cues to parameter setting. *Linguistic Inquiry* 30.1: 27–67.
- Emonds, J.
 1978 The verbal complex V'–V in French. *Linguistic Inquiry* 9: 151–75.
- Fodor, J. D.
 1998 Unambiguous triggers. *Linguistic Inquiry* 29.1: 1–36.
- Harris, Z. S.
 1951 *Methods in structural linguistics*. Chicago: University of Chicago Press.
- Jerne, N. K.
 1967 Antibodies and learning: Selection versus instruction. In G. C. Quarton, T. Melnechuk and F. O. Schmitt, eds., *The neurosciences: A study program*. New York: Rockefeller University Press.
 1985 The generative grammar of the immune system. *Science* 229: 1057–1059.
- Klima, E. S.
 1964 Negation in English. In J. Fodor and J. Katz, eds., *The structure of language*. Englewood Cliffs, NJ: Prentice Hall.
- Lasnik, H.
 2000 *Syntactic Structures revisited: Contemporary lectures on classic transformational theory*. Cambridge, MA: MIT Press.
- Lees, R.
 1957 Review of *Syntactic Structures*. *Language* 33.3: 375–408.
 1960 *The grammar of English nominalizations*. Bloomington: Indiana University Press.
- Lees, R. and E. S. Klima
 1963 Rules for English pronominalization. *Language* 39: 17–28.
- Lenneberg, E. H.
 1967 *The biological foundations of language*. New York: John Wiley.
- Lightfoot, D. W.
 1993 Why UG needs a learning theory: triggering verb movement. In C. Jones, ed., *Historical linguistics: Problems and perspectives*. London: Longman [reprinted in I. Roberts and A. Battye, eds., *Clause structure and language change*. Oxford: Oxford University Press].
 1999 *The development of language: Acquisition, change and evolution*. Oxford: Blackwell.
- Popper, K. R.
 1959 *The logic of scientific discovery*. London: Hutchinson.
- Rizzi, L.
 1990 *Relativized minimality*. Cambridge, MA: MIT Press.

PREFACE

This study deals with syntactic structure both in the broad sense (as opposed to semantics) and the narrow sense (as opposed to phonemics and morphology). It forms part of an attempt to construct a formalized general theory of linguistic structure and to explore the foundations of such a theory. The search for rigorous formulation in linguistics has a much more serious motivation than mere concern for logical niceties or the desire to purify well-established methods of linguistic analysis. Precisely constructed models for linguistic structure can play an important role, both negative and positive, in the process of discovery itself. By pushing a precise but inadequate formulation to an unacceptable conclusion, we can often expose the exact source of this inadequacy and, consequently, gain a deeper understanding of the linguistic data. More positively, a formalized theory may automatically provide solutions for many problems other than those for which it was explicitly designed. Obscure and intuition-bound notions can neither lead to absurd conclusions nor provide new and correct ones, and hence they fail to be useful in two important respects. I think that some of those linguists who have questioned the value of precise and technical development of linguistic theory may have failed to recognize the productive potential in the method of rigorously stating a proposed theory and applying it strictly to linguistic material with no attempt to avoid unacceptable conclusions by *ad hoc* adjustments or loose formulation. The results reported below were obtained by a conscious attempt to follow this course systematically. Since this fact may be obscured by the informality of the presentation, it is important to emphasize it here.

Specifically, we shall investigate three models for linguistic structure and seek to determine their limitations. We shall find that a certain very simple communication theoretic model of language and a more powerful model that incorporates a large part of what is now generally known as "immediate constituent analysis" cannot properly serve the purposes of grammatical description. The investigation and application of these models brings to light certain facts about linguistic structure and exposes several gaps in linguistic theory; in particular, a failure to account for such relations between sentences as the active-passive relation. We develop a third, *transformational* model for linguistic structure which is more powerful than the immediate constituent model in certain important respects and which does account for such relations in a natural way. When we formulate the theory of transformations carefully and apply it freely to English, we find that it provides a good deal of insight into a wide range of phenomena beyond those for which it was specifically designed. In short, we find that formalization can, in fact, perform both the negative and the positive service commented on above.

During the entire period of this research I have had the benefit of very frequent and lengthy discussions with Zellig S. Harris. So many of his ideas and suggestions are incorporated in the text below and in the research on which it is based that I will make no attempt to indicate them by special reference. Harris' work on transformational structure, which proceeds from a somewhat different point of view from that taken below, is developed in items 15, 16, and 19 of the bibliography (p. 115). In less obvious ways, perhaps, the course of this research has been influenced strongly by the work of Nelson Goodman and W. V. Quine. I have discussed most of this material at length with Morris Halle, and have benefited very greatly from his comments and suggestions. Eric Lenneberg, Israel Scheffler, and Yehoshua Bar-Hillel have read earlier versions of this manuscript and have made many valuable criticisms and suggestions on presentation and content.

The work on the theory of transformations and the transformational structure of English which, though only briefly sketched

below, serves as the basis for much of the discussion, was largely carried out in 1951 – 55 while I was a Junior Fellow of the Society of Fellows, Harvard University. I would like to express my gratitude to the Society of Fellows for having provided me with the freedom to carry on this research.

This work was supported in part by the U.S.A. Army (Signal Corps), the Air Force (Office of Scientific Research, Air Research and Development Command), and the Navy (Office of Naval Research); and in part by the National Science Foundation and the Eastman Kodak Corporation.

Massachusetts Institute of Technology,
Department of Modern Languages and
Research Laboratory of Electronics,
Cambridge, Mass.

NOAM CHOMSKY

August 1, 1956.

TABLE OF CONTENTS

Introduction to Second Edition by David W. Lightfoot	v
Preface	5
1. Introduction	11
2. The Independence of Grammar	13
3. An Elementary Linguistic Theory	18
4. Phrase Structure	26
5. Limitations of Phrase Structure Description	34
6. On the Goals of Linguistic Theory	49
7. Some Transformations in English	61
8. The Explanatory Power of Linguistic Theory	85
9. Syntax and Semantics	92
10. Summary	106
11. Appendix I: Notations and Terminology	109
12. Appendix II: Examples of English Phrase Structure and Transformational Rules	111
Bibliography	115

INTRODUCTION

Syntax is the study of the principles and processes by which sentences are constructed in particular languages. Syntactic investigation of a given language has as its goal the construction of a grammar that can be viewed as a device of some sort for producing the sentences of the language under analysis. More generally, linguists must be concerned with the problem of determining the fundamental underlying properties of successful grammars. The ultimate outcome of these investigations should be a theory of linguistic structure in which the descriptive devices utilized in particular grammars are presented and studied abstractly, with no specific reference to particular languages. One function of this theory is to provide a general method for selecting a grammar for each language, given a corpus of sentences of this language.

The central notion in linguistic theory is that of "linguistic level." A linguistic level, such as phonemics, morphology, phrase structure, is essentially a set of descriptive devices that are made available for the construction of grammars; it constitutes a certain method for representing utterances. We can determine the adequacy of a linguistic theory by developing rigorously and precisely the form of grammar corresponding to the set of levels contained within this theory, and then investigating the possibility of constructing simple and revealing grammars of this form for natural languages. We shall study several different conceptions of linguistic structure in this manner, considering a succession of linguistic levels of increasing complexity which correspond to more and more powerful modes of grammatical description; and we shall attempt to show that linguistic theory must contain at least these levels if it is to

provide, in particular, a satisfactory grammar of English. Finally, we shall suggest that this purely formal investigation of the structure of language has certain interesting implications for semantic studies.¹

¹ The motivation for the particular orientation of the research reported here is discussed below in § 6.

THE INDEPENDENCE OF GRAMMAR

2.1 From now on I will consider a *language* to be a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements. All natural languages in their spoken or written form are languages in this sense, since each natural language has a finite number of phonemes (or letters in its alphabet) and each sentence is representable as a finite sequence of these phonemes (or letters), though there are infinitely many sentences. Similarly, the set of 'sentences' of some formalized system of mathematics can be considered a language. The fundamental aim in the linguistic analysis of a language L is to separate the *grammatical* sequences which are the sentences of L from the *ungrammatical* sequences which are not sentences of L and to study the structure of the grammatical sequences. The grammar of L will thus be a device that generates all of the grammatical sequences of L and none of the ungrammatical ones. One way to test the adequacy of a grammar proposed for L is to determine whether or not the sequences that it generates are actually grammatical, i.e., acceptable to a native speaker, etc. We can take certain steps towards providing a behavioral criterion for grammaticalness so that this test of adequacy can be carried out. For the purposes of this discussion, however, suppose that we assume intuitive knowledge of the grammatical sentences of English and ask what sort of grammar will be able to do the job of producing these in some effective and illuminating way. We thus face a familiar task of explication of some intuitive concept—in this case, the concept "grammatical in English," and more generally, the concept "grammatical."

Notice that in order to set the aims of grammar significantly it is sufficient to assume a partial knowledge of sentences and non-

sentences. That is, we may assume for this discussion that certain sequences of phonemes are definitely sentences, and that certain other sequences are definitely non-sentences. In many intermediate cases we shall be prepared to let the grammar itself decide, when the grammar is set up in the simplest way so that it includes the clear sentences and excludes the clear non-sentences. This is a familiar feature of explication.¹ A certain number of clear cases, then, will provide us with a criterion of adequacy for any particular grammar. For a single language, taken in isolation, this provides only a weak test of adequacy, since many different grammars may handle the clear cases properly. This can be generalized to a very strong condition, however, if we insist that the clear cases be handled properly for *each* language by grammars all of which are constructed by the same method. That is, each grammar is related to the corpus of sentences in the language it describes in a way fixed in advance for all grammars by a given linguistic theory. We then have a very strong test of adequacy for a linguistic theory that attempts to give a general explanation for the notion "grammatical sentence" in terms of "observed sentence," and for the set of grammars constructed in accordance with such a theory. It is furthermore a reasonable requirement, since we are interested not only in particular languages, but also in the general nature of Language. There is a great deal more that can be said about this crucial topic, but this would take us too far afield. Cf. § 6.

2.2 On what basis do we actually go about separating grammatical sequences from ungrammatical sequences? I shall not attempt to

¹ Cf., for example, N. Goodman, *The structure of appearance* (Cambridge, 1951), pp. 5-6. Notice that to meet the aims of grammar, given a linguistic theory, it is sufficient to have a partial knowledge of the sentences (i.e., a corpus) of the language, since a linguistic theory will state the relation between the set of observed sentences and the set of grammatical sentences; i.e., it will define "grammatical sentence" in terms of "observed sentence," certain properties of the observed sentences, and certain properties of grammars. To use Quine's formulation, a linguistic theory will give a general explanation for what 'could' be in language on the basis of "what *is* plus *simplicity* of the laws whereby we describe and extrapolate what is". (W. V. Quine, *From a logical point of view* [Cambridge, 1953], p. 54). Cf. § 6.1.

give a complete answer to this question here (cf. §§ 6.7), but I would like to point out that several answers that immediately suggest themselves could not be correct. First, it is obvious that the set of grammatical sentences cannot be identified with any particular corpus of utterances obtained by the linguist in his field work. Any grammar of a language will *project* the finite and somewhat accidental corpus of observed utterances to a set (presumably infinite) of grammatical utterances. In this respect, a grammar mirrors the behavior of the speaker who, on the basis of a finite and accidental experience with language, can produce or understand an indefinite number of new sentences. Indeed, any explication of the notion "grammatical in L" (i.e., any characterization of "grammatical in L" in terms of "observed utterance of L") can be thought of as offering an explanation for this fundamental aspect of linguistic behavior.

2.3 Second, the notion "grammatical" cannot be identified with "meaningful" or "significant" in any semantic sense. Sentences (1) and (2) are equally nonsensical, but any speaker of English will recognize that only the former is grammatical.

- (1) Colorless green ideas sleep furiously.
- (2) Furiously sleep ideas green colorless.

Similarly, there is no semantic reason to prefer (3) to (5) or (4) to (6), but only (3) and (4) are grammatical sentences of English.

- (3) have you a book on modern music?
- (4) the book seems interesting.
- (5) read you a book on modern music?
- (6) the child seems sleeping.

Such examples suggest that any search for a semantically based definition of "grammaticalness" will be futile. We shall see, in fact, in § 7, that there are deep structural reasons for distinguishing (3) and (4) from (5) and (6); but before we are able to find an explanation for such facts as these we shall have to carry the theory of syntactic structure a good deal beyond its familiar limits.

2.4 Third, the notion "grammatical in English" cannot be identi-

fied in any way with the notion "high order of statistical approximation to English." It is fair to assume that neither sentence (1) nor (2) (nor indeed any part of these sentences) has ever occurred in an English discourse. Hence, in any statistical model for grammaticalness, these sentences will be ruled out on identical grounds as equally 'remote' from English. Yet (1), though nonsensical, is grammatical, while (2) is not. Presented with these sentences, a speaker of English will read (1) with a normal sentence intonation, but he will read (2) with a falling intonation on each word; in fact, with just the intonation pattern given to any sequence of unrelated words. He treats each word in (2) as a separate phrase. Similarly, he will be able to recall (1) much more easily than (2), to learn it much more quickly, etc. Yet he may never have heard or seen any pair of words from these sentences joined in actual discourse. To choose another example, in the context "I saw a fragile—," the words "whale" and "of" may have equal (i.e., zero) frequency in the past linguistic experience of a speaker who will immediately recognize that one of these substitutions, but not the other, gives a grammatical sentence. We cannot, of course, appeal to the fact that sentences such as (1) 'might' be uttered in some sufficiently far-fetched context, while (2) would never be, since the basis for this differentiation between (1) and (2) is precisely what we are interested in determining.

Evidently, one's ability to produce and recognize grammatical utterances is not based on notions of statistical approximation and the like. The custom of calling grammatical sentences those that "can occur", or those that are "possible", has been responsible for some confusion here. It is natural to understand "possible" as meaning "highly probable" and to assume that the linguist's sharp distinction between grammatical and ungrammatical² is motivated by a feeling that since the 'reality' of language is too complex to be described completely, he must content himself with a schematized

² Below we shall suggest that this sharp distinction may be modified in favor of a notion of levels of grammaticalness. But this has no bearing on the point at issue here. Thus (1) and (2) will be at different levels of grammaticalness even if (1) is assigned a lower degree of grammaticalness than, say, (3) and (4); but they will be at the same level of statistical remoteness from English. The same is true of an indefinite number of similar pairs.

version replacing "zero probability, and all extremely low probabilities, by *impossible*, and all higher probabilities by *possible*."³ We see, however, that this idea is quite incorrect, and that a structural analysis cannot be understood as a schematic summary developed by sharpening the blurred edges in the full statistical picture. If we rank the sequences of a given length in order of statistical approximation to English, we will find both grammatical and ungrammatical sequences scattered throughout the list; there appears to be no particular relation between order of approximation and grammaticalness. Despite the undeniable interest and importance of semantic and statistical studies of language, they appear to have no direct relevance to the problem of determining or characterizing the set of grammatical utterances. I think that we are forced to conclude that grammar is autonomous and independent of meaning, and that probabilistic models give no particular insight into some of the basic problems of syntactic structure.⁴

³ C. F. Hockett, *A manual of phonology* (Baltimore, 1955), p. 10.

⁴ We return to the question of the relation between semantics and syntax in §§ 8, 9, where we argue that this relation can only be studied after the syntactic structure has been determined on independent grounds. I think that much the same thing is true of the relation between syntactic and statistical studies of language. Given the grammar of a language, one can study the use of the language statistically in various ways; and the development of probabilistic models for the use of language (as distinct from the syntactic structure of language) can be quite rewarding. Cf. B. Mandelbrot, "Structure formelle des textes et communication: deux études," *Word* 10.1-27 (1954); H. A. Simon, "On a class of skew distribution functions," *Biometrika* 42.425-40 (1955).

One might seek to develop a more elaborate relation between statistical and syntactic structure than the simple order of approximation model we have rejected. I would certainly not care to argue that any such relation is unthinkable, but I know of no suggestion to this effect that does not have obvious flaws. Notice, in particular, that for any n , we can find a string whose first n words may occur as the beginning of a grammatical sentence S_1 and whose last n words may occur as the ending of some grammatical sentence S_2 , but where S_1 must be distinct from S_2 . For example, consider the sequences of the form "the man who ... are here," where ... may be a verb phrase of arbitrary length. Notice also that we can have new but perfectly grammatical sequences of word classes, e.g., a sequence of adjectives longer than any ever before produced in the context "I saw a — house." Various attempts to explain the grammatical-ungrammatical distinction, as in the case of (1), (2), on the basis of frequency of sentence type, order of approximation of word class sequences, etc., will run afoul of numerous facts like these.

