*Reprint from*

# FOR ROMAN JAKOBSON

*1956*

*Reprinted from* American Documentation, Vol. VIII, No. 4, Pages 284-291


# *LOGICAL STRUCTURES IN LANGUAGE**

*NOAM CHOMSKY***

The intensive development of techniques of descriptive linguistics during recent years has not been accompanied, at least among linguists, by a parallel development of the theory of meaning. Linguists have in general been relatively unconcerned with the problem of giving an account of meaning, and they have self-consciously attempted to avoid founding their descriptive science on a semantic or partially semantic basis. The motives and the consequences of this attempt have often been misunderstood. As is generally recognized, the vagueness and elusiveness of semantic notions have, of course, been one primary motivation. Naturally, linguists have tried to operate on as firm a basis as possible. But an even more powerful motivation is that semantic notions appear to be useless or misleading, even if we accept them unquestioningly.[1] The linguist avoids semantic foundations in the study of linguistic form for the simple and sufficient reason that this course is apparently the only one that enables him to arrive at a clear insight into grammatical structure. He thus finds himself studying language as an instrument, with almost no concern for the uses to which the instrument is put.

I think there is ample justification for rejecting any appeal to meaning in the study of linguistic form. But it is necessary to make a clear distinction between the appeal to meaning and the study of meaning. While the first has been a constant source of confusion, the second is clearly an integral part of a full-scale description of language. Rejection of the appeal to meaning in grammatical description does not require and should not support a corresponding rejection of the study of meaning. In this paper I would like to suggest that the results of a purely formal study of the structure of language can lead quite naturally to the study of meaning in ways which may be fruitful and revealing.

Before considering this question, it will be necessary to clarify somewhat the object of formal analysis and the nature of some of its results, focusing attention particularly on the notions "structure of a language" and "grammar of a language."

By a *language* in this discussion is meant a set of sentences all constructed from a finite alphabet of phonemes (or letters). These sentences may not be meaningful, in any independent sense of the word, nor need they ever have been used by speakers of the language. It is quite easy to construct sequences of English words which are quite without meaning and outside the linguistic experience of speakers of English who will, nevertheless, immediately recognize them as English sentences. A grammar of a language is essentially a theory of the set of sentences constituting the language. Grammatical research aims at determining the principles of construction underlying this set of sentences, and a grammar is a statement of these principles. More concretely, one can think of a grammar as being a device which generates all of the grammatical sentences of the language under analysis without generating non-sentences. The theory of grammar is concerned with the precise specification of how these sentence-generating devices operate.

Clearly, it would be absurd to try to give rules of sentence construction directly in terms of phonemes or letters. One would not hope to construct a grammar of English on any reasonably limited scale in terms of rules of phoneme succession alone. For this reason a linguist will search for units which function more directly in sentence construction, let us say morphemes or words. It will be much more efficient to give a joint description of the word structure of sentences and the phoneme structure of words, than a direct description of the phoneme structure of sentences. Briefly, this

*Paper presented to Joint Meeting AAAS – ADI – New York. December, 1956.

**Massachusetts Institute of Technology.

[1]For discussion of this point see my "Semantic Considerations in Grammar," *Georgetown Univ. Monograph Series in Linguistics*, vol. 8, pp. 141-150 (1955) and *Syntactic Structures*, The Hague, Mouton and Co., 1957, p. 9.

**Meeting of the Aristotelian Society at 21, Bedford Square, London, W.C.1, on 24th November, 1958, at 7.30 p.m.**

# IV.—WHAT IS SAID TO BE

*By* ISRAEL SCHEFFLER AND NOAM CHOMSKY

(*Read by* I. SCHEFFLER)

## I.—PROBLEM

### A

PROFESSOR QUINE[1] has recently put his ontological criterion[2] thus:

(1) An entity is assumed by a theory if and only if it must be counted among the values of the variables in order that the statements affirmed in the theory be true.[3]

---

[1] We are indebted to Professor Quine for a helpful discussion of an earlier draft of this paper.

[2] After we had submitted this paper, our attention was called to Richard L. Cartwright's article, " Ontology and the Theory of Meaning " (*Philosophy of Science*, 21, 1954, pp. 316–325), by Alan Ross Anderson's review in *The Journal of Symbolic Logic*, 22, 1957, pp. 393–4. Cartwright's paper is devoted to an examination of Quine's criterion, and anticipates our arguments in Section I, regarding the universal and null commitments assigned to false existential statements by variants of the criterion, and the unwanted commitments that follow from certain of its applications. Cartwright concludes that any adequate formulation of an ontological criterion will be intensional, and he proposes such a formulation. Our interest, on the other hand, has been to show that extensional alternatives are trivial if adequate, while the philosophical purposes motivating the original ontological criterion are equally served by arguments making no reference to ontological commitment at all. Consequently, though not disagreeing with Cartwright that any adequate, non-trivial criterion is likely to be intensional, we do not see any philosophical point in developing such a criterion, and we feel, further, that any such criterion will be obscure, for reasons similar to those advanced by Quine.

[3] *From A Logical Point of View*, Cambridge, Massachusetts, Harvard University Press, 1953, p. 103. All references to Quine in the present paper concern passages in this book; the title will thus be omitted from future citations.

Reprinted from INFORMATION AND CONTROL, Volume 2, No. 2, June 1959


INFORMATION AND CONTROL **2**, 137–167 (1959)

# On Certain Formal Properties of Grammars*

NOAM CHOMSKY

*Massachusetts Institute of Technology, Cambridge, Massachusetts and The Institute for Advanced Study, Princeton, New Jersey*

A grammar can be regarded as a device that enumerates the sentences of a language. We study a sequence of restrictions that limit grammars first to Turing machines, then to two types of system from which a phrase structure description of the generated language can be drawn, and finally to finite state Markov sources (finite automata). These restrictions are shown to be increasingly heavy in the sense that the languages that can be generated by grammars meeting a given restriction constitute a proper subset of those that can be generated by grammars meeting the preceding restriction. Various formulations of phrase structure description are considered, and the source of their excess generative power over finite state sources is investigated in greater detail.

## SECTION 1

A language is a collection of sentences of finite length all constructed from a finite alphabet (or, where our concern is limited to syntax, a finite vocabulary) of symbols. Since any language $L$ in which we are likely to be interested is an infinite set, we can investigate the structure of $L$ only through the study of the finite devices (grammars) which are capable of enumerating its sentences. A grammar of $L$ can be regarded as a function whose range is exactly $L$. Such devices have been called "sentence-generating grammars."[1] A theory of language will contain, then, a specifica-

* This work was supported in part by the U. S. Army (Signal Corps), the U. S. Air Force (Office of Scientific Research, Air Research and Development Command), and the U. S. Navy (Office of Naval Research). This work was also supported in part by the Transformations Project on Information Retrieval of the University of Pennsylvania. I am indebted to George A. Miller for several important observations about the systems under consideration here, and to R. B. Lees for material improvements in presentation.

[1] Following a familiar technical use of the term "generate," cf. Post (1944). This locution has, however, been misleading, since it has erroneously been interpreted as indicating that such sentence-generating grammars consider language

Reprinted from LANGUAGE
Vol. 31, Nos. 1 & 2, January–March, 1955
*Printed in U.S.A.*

# LOGICAL SYNTAX AND SEMANTICS

## THEIR LINGUISTIC RELEVANCE

NOAM CHOMSKY
*Harvard University*

The relation between linguistics and logic has been discussed in a recent paper by Bar-Hillel,[1] where it is argued that a disregard for work in logical syntax and semantics has caused linguists to limit themselves too narrowly in their inquiries, and to fall into several errors. In particular, Bar-Hillel asserts, they have attempted to derive relations of synonymy and so-called 'rules of transformation', such as the active-passive relation, from distributional studies alone, and they have hesitated to rely on considerations of meaning in linguistic analysis. No one can quarrel with the suggestion that linguists interest themselves in meaning or transformation rules, but the relevance of logical syntax and semantics[2] (at least as we now know them) to this study is very dubious. I think that a closer investigation of the assumptions and concerns of logical syntax and semantics will show that the hope of applying the results which have been achieved in these fields to the solution of linguistic problems is illusory.

Bar-Hillel sums up the major part of his argument in the following statement of a presumed relationship between linguistics and logical syntax (236–7):

> There exists a conception of syntax, due to Carnap, that is purely formal (structural) and adequate in a sense in which the conception prevalent among American structural linguists is not. This conception entails a certain fusion between grammar and logic, with grammar treating approximately the formational part of syntax and logic its transformational part. The relation of COMMUTABILITY may be sufficient as a basis for formational analysis, but other relations, such as that of formal CONSEQUENCE, must be added for transformational analysis.

Later he urges that linguists 'follow Carnap's lead' even further, introducing semantic notions for, presumably, an even more adequate theory.[3]

In particular, he claims that the prevalent conception of syntax among linguists cannot lead to the definition and fruitful analysis of such relations as that between *oculist* and *eye-doctor*, and that between the active and the passive,

[1] Logical syntax and semantics, *Lg.* 30.230–7 (1954).

[2] I want to make it clear that the remarks which follow are critical not of logical syntax and semantics as such, but of the claim that these disciplines furnish solutions to linguistic problems. I have borrowed freely from various critical accounts of the theory of meaning, including W. V. Quine, *From a logical point of view*, esp. Ch. 2, 7, 8 (Cambridge, Mass., 1953); M. G. White, The analytic and the synthetic: An untenable dualism, *John Dewey: Philosopher of science and freedom* 316–30 (New York, 1950), reprinted in Linsky, *Semantics and the philosophy of language* 272–86 (Urbana, 1952).

[3] I will not discuss these proposals separately, since, as I will try to show, they fail for the same reasons.

Reprinted from INFORMATION AND CONTROL, Volume 2, No. 4, December 1959


INFORMATION AND CONTROL **2**, 393–395 (1959)

# A Note on Phrase Structure Grammars

NOAM CHOMSKY

*Massachusetts Institute of Technology, Cambridge, Massachusetts*

In Chomsky (1959)[1] a class of grammars is studied each of which contains a finite number of "rules" of the form $A \to \varphi$, where $A$ is a single symbol and $\varphi$ is not null. Such grammars (there called type 2 grammars) we will now call *context-free* (CF) phrase structure grammars. A sequence $(\varphi_1, \cdots, \varphi_n)$ of strings is called a *$\varphi$-derivation* of the CF grammar G if $\varphi = \varphi_1$ and for each $i < n$, there are strings $A, \psi_1, \psi_2, \psi_3$ such that $\varphi_i = \psi_1 A \psi_2$, $\varphi_{i+1} = \psi_1 \psi_3 \psi_2$ and $A \to \psi_3$ is a rule of G. $\varphi$ is a *terminal string* if it contains no $A$ such that for some $\psi$, $A \to \psi$ is a rule of G. The language $L_G$ *generated by* G is the set of terminal strings that appear in (and thus conclude) $S$-derivations of G, where $S$ is a designated initial symbol. A string is *derivable* if it is a step in an $S$-derivation. G is a *self-embedding* (s.e.) grammar if it contains strings $A, \varphi_1, \varphi_2$ such that $\varphi_1$ and $\varphi_2$ are not null and there is an $A$-derivation of $\varphi_1 A \varphi_2$. $\varphi \Rightarrow \psi$ if and only if $\psi$ is a line of a $\varphi$-derivation.

Given a finite state Markov source $\Sigma$ with a designated initial state $S_0$ and a symbol emitted with each interstate transition, we define the *language generated by* $\Sigma$ as the set of strings produced as the system moves from $S_0$ to a first recurrence of $S_0$. The set of languages that can be generated in this way we call *finite state languages*.[2] Clearly finite state languages constitute a proper subset of the languages that can be generated by CF grammars (cf. Chomsky (1959), §5). It is an interesting and important problem to characterize precisely the set of nonfinite state languages that can be generated by CF grammars. As a step towards this, it was proven in Chomsky (1959) that a set of strings is

[1] The notations and terminology of that paper will be used in this note. In particular, we use the following notational convention: capital letters will be used for nonterminal strings (see below); small Latin letters for terminal strings; Greek letters for arbitrary strings; early letters of all alphabets for single symbols; late letters for arbitrary strings.

[2] Finite state languages are what are called "regular events" in Kleene (1956).

Reprinted from INFORMATION AND CONTROL, Volume 1, No. 2, May 1958
Academic Press Inc. *Printed in U.S.A.*

INFORMATION AND CONTROL **1**, 91–112 (1958)

# Finite State Languages*

NOAM CHOMSKY

*Massachusetts Institute of Technology, Cambridge, Massachusetts*

AND

GEORGE A. MILLER

*Harvard University, Cambridge, Massachusetts*

A finite state language is a finite or infinite set of strings (sentences) of symbols (words) generated by a finite set of rules (the grammar), where each rule specifies the state of the system in which it can be applied, the symbol which is generated, and the state of the system after the rule is applied. A number of equivalent descriptions of finite state languages are explored. A simple structural characterization theorem for finite state languages is established, based on the cyclical structure of the grammar. It is shown that the complement of any finite state language formed on a given vocabulary of symbols is also a finite state language, and that the union of any two finite state languages formed on a given vocabulary is a finite state language; i.e., the set of all finite state languages that can be formed on a given vocabulary is a Boolean algebra. Procedures for calculating the number of grammatical strings of any given length are also described.

In the vast majority of communication situations the messages that are exchanged consist of strings of symbols. It is possible to imagine a code that uses only one symbol per message, but few situations are so

* This work was supported in part by the Army (Signal Corps), the Navy (Office of Naval Research), the Air Force (Office of Scientific Research and Operational Applications Laboratory, Air Research and Development Command), the National Science Foundation, and the Social Science Research Council (Committee on Mathematical Training of Social Scientists), and appears as report number AFCRC-TR-58-50, ASTIA Document Number AD 146781.

MORRIS HALLE

# IN DEFENSE OF THE NUMBER TWO*

THE SUBJECT of this essay is the Jakobsonian distinctive feature system, in particular its most controversial proposition concerning the binary structure of all features.[1] The distinctive feature system is a framework for the description of the phonetic facts of language. In the history of phonetics other such frameworks have been known: e.g. Alexander Melville Bell's Visible Speech and its direct descendant, the well-known phonetic alphabet of the International Phonetic Association, Jespersen's antalphabetic notation, or Pike's phonetic system.

These frameworks are fundamentally questionnaires. In using a particular phonetic system, just as in using a particular questionnaire, certain information will be obtained and this information will have a certain structure. And as in the case of a questionnaire, the choice of one system over another is determined by the investigator's belief that the particular set of questions and the manner in which they are phrased are the most appropriate to the research he is interested in. Thus, for example, in the phonetic frameworks just mentioned the position of the epiglottis in the articulation of the different sounds is not considered, in sharp contrast to the position of the major tongue constriction which is of primary concern in all frameworks. The disregard of the epiglottis and the great attention paid to the position of the tongue constriction would normally be justified on the grounds that the former information does not seem to do much for our systematization of the facts of speech, whereas the latter does. If one wanted to force a change one would have to show that this change actually was

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

[1] For a detailed presentation of the distinctive feature system see R. Jakobson, M. Halle, C. G. M. Fant, *Preliminaries to speech analysis, M.I.T. Acoustics Laboratory Technical Report 13* (1st printing: January 1952; 2nd printing with additions and corrections: May 1952; 3rd printing: June 1955), and R. Jakobson and M. Halle, *Fundamentals of language I. Phonology and phonetics* (The Hague, 1956).

# ON ACCENT AND JUNCTURE IN ENGLISH*

BY NOAM CHOMSKY, MORRIS HALLE, FRED LUKOFF

IN RECENT YEARS, A CONSIDERable amount of study has been devoted to the 'suprasegmental' features of pitch, juncture, and stress. In particular, English stress has been investigated in great detail, and a variety of interesting phenomena have been noted. These investigations have led to the widespread adoption of a phonemic notation in which four degrees of stress are marked.[1]

In this paper we explore the adequacy of a more economical phonemic transcription in which only the opposition *accented-unaccented* is marked. We find that for a well-defined and independently significant class of utterances such a transcription suffices to determine the full range of variation in stress which has been brought to light and clarified by recent investigations, provided that we take into consideration the hierarchical organization of the utterance.

1. By a 'transcription' we mean a system of symbols and an associated system of rules which assign a value to each sequence of these symbols. We call each sequence of these symbols a 'representation' of the utterances having the assigned value. Here we are interested in a phonemic transcription whose rules assign a phonetic value to every sequence of its symbols. The phonemic transcription that we construct below will meet a number of requirements that we shall discuss in detail. Conditions different from ours have been proposed for phonemic transcriptions. In the course of our discussion we advance reasons for preferring this set of conditions.

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

[1] Cf. G. L. Trager & H. L. Smith, *Outline of English Structure* (Norman, Oklahoma, 1951). Most of our data is taken from Trager & Smith, and from S. S. Newman, "On the stress system of English," *Word*, 2. 171–187 (1946). We have, on the whole, not attempted to discover new facts or to challenge the accuracy of available data, even though we were in some cases not entirely convinced of their validity. Instead we have designed our rules to fit the stress patterns described in the literature. The present paper offers a different phonemic analysis based on published data. Studies of stress, in general, are somewhat unconvincing because the data are impressionistic. Since the task of finding physical, objective (acoustic and/or articulatory) correlates of stress will require very extensive and time-consuming investigation, the best that can be done at present is to accept the available information on stress, subject to revision in the light of later discoveries.