4

PHRASE STRUCTURE

4.1 Customarily, linguistic description on the syntactic level is formulated in terms of constituent analysis (parsing). We now ask what form of grammar is presupposed by description of this sort. We find that the new form of grammar is *essentially* more powerful than the finite state model rejected above, and that the associated concept of "linguistic level" is different in fundamental respects.

As a simple example of the new form for grammars associated with constituent analysis, consider the following:

(13) (i) *Sentence* → *NP* + *VP*
(ii) *NP* → *T* + *N*
(iii) *VP* → *Verb* + *NP*
(iv) *T* → *the*
(v) *N* → *man*, *ball*, etc.
(vi) *Verb* → *hit*, *took*, etc.

Suppose that we interpret each rule $X \rightarrow Y$ of (13) as the instruction "rewrite X as Y". We shall call (14) a *derivation* of the sentence "the man hit the ball." where the numbers at the right of each line of the derivation refer to the rule of the "grammar" (13) used in constructing that line from the preceding line.[1]

[1] The numbered rules of English grammar to which reference will constantly be made in the following pages are collected and properly ordered in § 12, *Appendix II*. The notational conventions that we shall use throughout the discussion of English structure are stated in § 11, *Appendix I*.

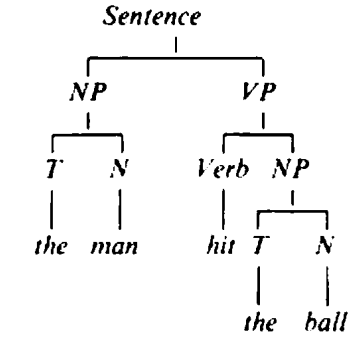
In his "Axiomatic syntax: the construction and evaluation of a syntactic calculus," *Language* 31.409–14 (1955), Harwood describes a system of word class analysis similar in form to the system developed below for phrase structure. The system he describes would be concerned only with the relation between *T* + *N* + *Verb* + *T* + *N* and *the* + *man* + *hit* + *the* + *ball* in the example discussed

(14) *Sentence*

NP + *VP*	(i)
T + *N* + *VP*	(ii)
T + *N* + *Verb* + *NP*	(iii)
the + *N* + *Verb* + *NP*	(iv)
the + *man* + *Verb* + *NP*	(v)
the + *man* + *hit* + *NP*	(vi)
the + *man* + *hit* + *T* + *N*	(ii)
the + *man* + *hit* + *the* + *N*	(iv)
the + *man* + *hit* + *the* + *ball*	(v)

Thus the second line of (14) is formed from the first line by rewriting *Sentence* as *NP* + *VP* in accordance with rule (i) of (13); the third line is formed from the second by rewriting *NP* as *T* + *N* in accordance with rule (ii) of (13); etc. We can represent the derivation (14) in an obvious way by means of the following diagram:

(15)



The diagram (15) conveys less information than the derivation (14), since it does not tell us in what order the rules were applied in (14).

in (13)–(15); i.e., the grammar would contain the "initial string" *T* + *N* + *Verb* + *T* + *N* and such rules as (13iv–vi). It would thus be a weaker system than the elementary theory discussed in § 3, since it could not generate an infinite language with a finite grammar. While Harwood's formal account (pp. 409–11) deals only with word class analysis, the linguistic application (p. 412) is a case of immediate constituent analysis, with the classes $C_{i \ldots m}$ presumably taken to be classes of word sequences. This extended application is not quite compatible with the formal account, however. For example, none of the proposed measures of goodness of fit can stand without revision under this reinterpretation of the formalism.