On the Assignment of Constituent Structures to the Sentences Generated by a Transformational Grammar[1]

D. Terence Langendoen, Brooklyn College of The City University of New York

1. Modes of Representation for Base Phrase-Markers. Within each of the so-called "standard theories" of generative-transformational grammar, it is assumed that there are at least two distinct levels of representation of the structures of sentences, deep structure and surface structure. One aspect of deep-structure representation is the assignment of constituent structure to the strings that are generated by the base component; such structures are called base phrase-markers.[2] Base phrase-markers can be associated with the strings generated by the base component either directly or indirectly. The direct mode of associating base phrase-markers with base-generated strings involves including in the terminal vocabularies of grammars the symbols that are needed to construct base phrase-markers by the base phrase-structure rules themselves. Thus the base component of a direct-mode transformational grammar does not simply generate strings of lexical items, but rather generates phrase-markers that reduce to such strings if the symbols used to represent constituent structures are removed. There are two distinct ways of representing base phrase-markers for sentences generated by grammars in the direct mode. One is in a two-dimensional form, using labeled nodes and directed lines, called tree-diagrams. The other is in a one-dimensional form using labeled brackets, here called bracket-diagrams.[3]

The indirect mode of associating base phrase-markers with base-generated strings does not require the addition of special symbols to the terminal vocabularies of grammars. Rather, base phrase-markers are obtained by construction from the derivations of the base-generated strings with respect to the particular grammars that are used to generate them. Depending on the construction used, the base phrase-markers so obtained may represent the same structural information that is represented by tree-diagrams or by bracket-diagrams, or less information. The construction used by Chomsky (1955) is of the first sort, while that used by Lasnik and Kupin (forthcoming) and Kupin (1978) is of the second sort.

In the following section, I show that the theory of transformational grammar in the direct mode is superior to the theory of transformational grammar in the indirect mode. In the final section, I argue further that bracket-diagrams are superior to tree-diagrams for the direct representation of base phrase-markers.

2. The Superiority of the Direct Mode.

2.1. The Form of Transformations.

2.1.1. Syntactic Transformations. A derivation of a sentence s with respect to a grammar G is conventionally understood of as consisting of a sequence of lines starting with an axiom (usually the string consisting of the designated non-terminal symbol S, possibly flanked by special terminal (boundary) symbols), and ending with s, such that each line is constructed from the lines that precede it by application of a rule of grammar. In all standard theories, syntactic transformations (rules of the transformational component, which apply only after rules of the base component have applied) are applicable not to strings of grammatical formatives, but rather to strings with associated constituent structure. Thus, with respect to grammars in the indirect mode, the lines that appear in a derivation of a sentence are not all drawn from a common vocabulary. From the initial line up to the line that represents the base-generated string, the symbols that appear are drawn exclusively from the terminal and non-terminal vocabularies of the grammar. All subsequent lines consist of symbols that are used to represent phrase-markers (the first of these lines being the deep phrase-marker, the final line being the surface phrase-marker, and all intermediate lines being intermediate phrase-markers).[4] This peculiar property of derivations with respect to a standard-theory transformational grammar has largely gone unnoticed because of Chomsky's subsequent redefinition of the notion of derivation as a sequence of phrase-markers, starting with deep phrase-markers (or possibly base-generated structures that lack lexical items), and ending with surface phrase-markers (see, for example, Chomsky 1970: 63, 113-114, and especially Chomsky 1972:123-124). While derivations conforming to Chomsky's new definition indeed do lack the peculiar property just mentioned, the grammars that construct such derivations have a correspondingly peculiar property that grammars that construct traditionally defined derivations lack: grammars that construct derivations that consist solely of sequences of phrase-markers have infinitely many axioms, and the role of the categorial subcomponent of the base component together with the procedure for constructing base phrase-markers is to generate those axioms. Thus Chomsky secures no real advantage with his redefinition of the notion 'derivation'.

On the other hand, in a derivation of a sentence with respect to a transformational grammar in the direct mode, each line can uniformly be viewed as a phrase-marker, or as a stage in the direct construction of the deep phrase-marker. Moreover, assuming that there are no 'global' rules in any of the grammars under consideration, each line of a derivation follows from the immediately preceding line by application of a rule of grammar, and the syntactic transformations can be represented directly as rewrite-rule schemata that abbreviate sets of rewrite rules of known type on Chomsky's hierarchy. In particular, transformations that simply reorder elements within phrase-markers are represented as type-1 ('non-shrinking') rule schemata, while transformations that delete elements within phrase-markers without any compensatory insertion of elements are represented as a kind of type-0 ('shrinking') rule schema. Since the rules needed to specify the individual rules of grammar given the various schemata are of type-1, it follows from the theorem I proved elsewhere (Langendoen 1976), that transformational grammars that do not make use of deletion operations have a strong generative capacity that does not exceed that of type-1 grammars, while the strong generative capacity of transformational grammars that effect deletion may be greater, depending on the type of deletion operations allowed.[5]

Thus, among the advantages of adopting the theory of transformational grammar in the direct mode are the following. First, the traditional definition of 'derivation' does not need to be altered in order to insure that the lines of a derivation are all the same sort of object. Second, it enables us to represent syntactic transformations directly as rewrite-rule schemata, and so enables us to study the strong generative capacity of transformational grammars by the same tools that have so fruitfully been used in the study of the weak generative capacities of rewrite-rule systems.

Having established that there are definite advantages to be gained by adopting the theory of transformational grammar in the direct mode, we must now inquire whether there are any countervailing disadvantages. To my knowledge, the only discussion of the relative merits of grammars in the direct and indirect modes is to be found in Chomsky (1965: 89). There Chomsky argues that if "we were to include labeled brackets in the strings that constitute a derivation [in the categorial subcomponent of the base] and were to allow the 'rewriting rules' to refer to these symbols [, w]e should have a kind of transformational grammar, and we should have entirely lost the intuition about language structure that motivated the development of phrase structure grammar." However, this objection is completely without force, since the only way that the rewriting rules of the categorial subcomponent could refer to labeled brackets is if they were context-sensitive, and the base categorial rules are assumed to be context-free in any standard-theory transformational grammar. Hence, the 'intuition' that underlay the original development of phrase-structure grammar is preserved in any standard theory of transformational grammar in the direct mode.

A second possible disadvantage of the theory of transformational grammar in the direct mode has to do with the interpretation of variables in transformational rule schemata. As Chomsky (1965:217) observes, the variables that appear in the transformational rule schemata of grammars in the indirect mode allow for their interpretation as arbitrary strings of grammatical formatives, whereas variables in transformational rule schemata of grammars in the direct mode cannot be so interpreted. Rather, these variables must be interpreted as representing arbitrary sub-phrase-markers of larger phrase-markers. However, this difference does not result in any particular advantage for the theory of transformational grammar in the indirect mode, since the variables that appear in transformational rule schemata can be interpreted as arbitrary sub-phrase-markers simply by convention.[6]

Finally, it might be supposed that it is easier to express genuine linguistic universals in the theory of transformational grammar in the indirect mode than in the corresponding theory in the direct mode. However, no reasons for supposing that this is the case have ever been offered, and it seems extremely unlikely that such reasons can be found, especially since many of the proposed universals have had to do with the constituent structures of the strings undergoing transformation.[7]

2.1.2. Lexical-Insertion Rules. Lexical items are inserted into base-structures by schemata that refer to aspects of the constituent structures into which they are inserted, and properties of other lexical items that have already been inserted. Such schemata must be formulated as transformational rule schemata within the theory of transformational grammar in the indirect mode, and one of the major problems confronting that theory is the formulation of restrictions on the expressive power of those schemata. On the other hand, lexical-insertion schemata can be formulated as context-sensitive rule schemata in the theory of transformational grammar in the direct mode, since only one symbol is rewritten by any such schema, and since all of the information relevant to the insertion of a lexical item is present in the line of a derivation to which a rule that inserts it applies.[8] The fact that lexical-insertion schemata are context-sensitive in form in the theory of transformational grammar in the direct mode allows us to relate the form of syntactic rules to the component to which they belong by a very simple principle that is not expressible as such in the alternative theory. That principle is this. A syntactic rule occurs in the base component of a transformational grammar if and only if it is a constituent-structure rule. The ability to formulate such a simple, yet far-reaching, principle within the theory of transformational grammar in the direct mode constitutes very strong evidence for the preferability of that theory.

2.2. The Treatment of Coordinate Structures. According to the theory of transformational grammar in the indirect mode, the assignment of constituent structure to coordinate structures requires that the base component of a transformational grammar contain "rule schemata that go beyond the range of phrase structure rules in strong generative capacity" (Chomsky 1965: 99). This is an extremely peculiar state of affairs, when one considers that the kind of coordinate structure that necessitates the inclusion of these more powerful schemata in indirect-mode theory can be directly generated by finite-state grammars in the direct mode. To see this, let L be the language $\{(a\ b)^n\ c\ a\ b: n \geq 1\}$, and for each sentence of the form $(a\ b)^n\ c\ a\ b$, let its phrase-marker be of the form $[_S\ ([_S\ a\ b\ ]_S)^n$ - $[_C\ c\ ]_C\ [_S\ a\ b\ ]_S\ ]_S$. That is, the structure of sentences of L is that of a 'flat' coordinate nature, for which rule schemata more powerful in generative capacity than that of phrase-structure rules are required by indirect-mode theory.[9] However, the phrase-markers of the sentences of L can be correctly assigned to those sentences by a finite-state grammar in the direct mode.[10] Thus the direct-mode theory of transformational grammar is better adapted to the description of coordinate structures than is the indirect-mode theory.

3. <u>The Superiority of Bracket-Diagrams</u>. When stating transformational rule schemata in the direct mode, it seems easier to do so using bracket-diagrams than tree-diagrams; similarly the statement of conditions on the form of such schemata seems more easily stated when phrase-markers are given in bracket-diagrammatic form. However, this may reflect not so much the superiority of bracket-diagrams to tree-diagrams for

linguistic purposes, as the greater apparent ease of manipulation of one-dimensional as opposed to two-dimensional formal objects. Perhaps if linguists had greater familiarity with tree automata, they would see more clearly how to manipulate and constrain the manipulation of tree-diagrams by transformational rules.

In this section we consider some psycholinguistic evidence that phrase-markers should be represented in bracket-diagrammatic rather than in tree-diagrammatic form. The evidence has to do with the nature of the on-line parsing of sentences by people who have achieved mastery over a natural language. The discussion of this evidence can proceed on the assumption that the contribution of the transformational component is null; in fact, we need only discuss sentences whose phrase-markers can in principle be assigned to them by finite-state transducers, namely sentences that lack center-embedding. The discussion to follow will proceed in the following stages. First, we consider what consequences follow from the fact that people can assign phrase-markers to the non-center-embedded sentences of the languages they have mastery of, as they listen to them from beginning to end. Second, we define a notion of reduced phrase-marker, which is adequate for purposes of representing the syntactic structures of those sentences, and which can be assigned to those sentences by a finite-state transducer as that transducer 'reads' them from beginning to end. Third, we give a mechanical procedure for constructing such a transducer given an arbitrary non-center-embedding context-free grammar. Fourth, we show that while reduced phrase-markers can be represented in bracket-diagrammatic form, they cannot be represented in tree-diagrammatic form.

3.1. The On-Line Assignment of Phrase-Markers to Non-Center-Embedded Sentences. A procedure for constructing finite-state transducers that parse the sentences generated by certain types of non-center-embedding context-free phrase-structure grammars was first proposed in Chomsky (1959). This procedure is of interest, as Chomsky (1961) points out, since it provides a start toward the solution to the problem of devising a model of speech perception that incorporates a generative grammar. Schematically, how such a finite-state transducer works is as follows. Suppose x is a sentence generated by a non-center-embedding context-free grammar G of the appropriate sort. Then the transducer T, in accepting the string x, produces as output another string $\Phi(y)$, where y is a structural description of x with respect to G, and Φ is an effective one-one mapping onto strings in the output vocabulary of T (Chomsky 1963: 396). As a model of speech perception, however, T is of interest only to the extent to which the mapping Φ can itself be carried out while the sentence is being processed, since it is our impression that we compute the phrase-markers of sentences as we hear them. As Chomsky (1963: 398) points out, the construction of such an on-line mapping Φ was first carried out in my 1961 bachelor's thesis; however, this mapping is not such that it can itself be carried out by a finite-state transducer, and as I showed in my 1975 paper, if the structural descriptions of sentences generated by context-free grammars are considered to be full phrase-markers (proper bracket- or tree-diagrams, or their equivalent), no such mapping exists. This means either that there are sentences generated by non-center-embedding context-free grammars that cannot be parsed on line by finite-state transducers, or that the identification of structural descriptions with full phrase-markers is in error.

In my 1975 paper, I argued for the first of these two alternatives, on the grounds that sentences exhibiting multiple right- or left-embedding beyond a certain degree are unnatural; and that multiple right- and left-embedding structures are replaced by pseudo-coordinate structures by application of readjustment rules. However, the arguments I present there are not convincing, since the intonational contours that such readjustment rules are intended to account for can be predicted without the application of readjustment rules at all (Liberman and Prince 1977: 327-329). But if the assignment of intonational contours can be determined by principles that apply to phrase-markers without readjustment, then the linguistic evidence for readjustment rules of the type I considered vanishes. Thus we are led to the second of the two alternatives, namely that one should not identify the structural descriptions of sentences generated by context-free grammars with the phrase-markers that those grammars conventionally assign to them.

3.2. The Adequacy of Reduced Phrase-Markers. How much information about the structure of sentences generated by a context-free grammar must be assigned to that sentence? By definition, a structural description of a sentence is a specification of "the elements of which [it] is constructed, the order, arrangement, and interrelations and whatever other grammatical information is needed to determine how [it] is used and understood" (Chomsky and Miller 1963: 285). In particular, from a structural description of a sentence, one must be able to determine what its constituents are, and what grammatical relations hold among them; and if a given sentence is n ways ambiguous, it must have n distinct structural descriptions associated with it, one for each of its interpretations.

Suppose we have a language L that is generated by a non-center-embedding context-free grammar G. It turns out that all of the information needed to determine constituency, grammatical relations, and degree of ambiguity is available in partial phrase-markers of a certain sort that moreover can be associated with the sentences of L by a finite transducer T operating on line. Suppose we identify these partial phrase-markers associated with the sentences of L as the structural descriptions of those sentences with respect to G. It follows, then, that there is a finite transducer T such that if x is a sentence of L, T accepts x and associates with it a structural description y in the course of accepting x. In other words, if we take the structural descriptions of the sentences generated by the non-center-embedding phrase-structure grammar G to be partial phrase-markers, not only is there a finite transducer T that is strongly equivalent to G, but T can also serve as a model of the on-line parsing of sentences of L, since the mapping Φ from the output of the transducer to the structural descriptions of the sentences it accepts is the identity mapping.

We first illustrate what we mean by partial phrase-markers with an artificial example, that also illustrates their adequacy in representing degrees of structural ambiguity. Let G_1 be the non-center-embedding context-free grammar (1). G_1 generates the finite-state language L_1 in (2), and associates with the sentences of L_1 the set of full phrase-markers in (3) (note that (3) is not a finite-state language).[11]

$$(1)\quad G_1 = \left\{\begin{array}{l} N = \{S, A, B, C\} \\ T = \{a, b, c\} \\ A = \{S\} \\ R = \left\{\begin{array}{l} \text{a. } S \rightarrow A\ C \\ \text{b. } A \rightarrow A\ B \\ \text{c. } C \rightarrow B\ C \\ \text{d. } A \rightarrow a \\ \text{e. } B \rightarrow b \\ \text{f. } C \rightarrow c \end{array}\right\} \end{array}\right\}$$

$$(2)\quad L_1 = \{a\ b^n\ c : n \geqq 0\}$$

$$(3)\quad \{[_S([_A)^m[_A\ a\ ]_A([_B\ b\ ]_B]_A)^m[_C([_B\ b\ ]_B[_C)^n\ c\ (]_C)^n]_C]_S : m, n \geqq 0\}$$

It will be observed that each sentence of L_1 of the form $a\ b^n\ c$ is n+1-ways ambiguous; for example, the sentence a b c has the two full phrase-markers in (4), while the sentence a b b c has the three full phrase-markers in (5).[12]

(4) a. $[_S \overbrace{[_A} [_A\ a\ ]_A [_B\ b\ ]_B]_A [_C\ c\ ]_C]_S$

b. $[_S [_A\ a\ ]_A [_C [_B\ b\ ]_B [_C\ c\ \overbrace{]_C}]_C]_S$

(5) a. $[_S \overbrace{[_A [_A} [_A\ a\ ]_A [_B\ b\ ]_B]_A [_B\ b\ ]_B]_A [_C\ c\ ]_C]_S$

b. $[_S \overbrace{[_A} [_A\ a\ ]_A [_B\ b\ ]_B]_A [_C [_B\ b\ ]_B [_C\ c\ \overbrace{]_C}]_C]_S$

c. $[_S [_A\ a\ ]_A [_C [_B\ b\ ]_B [_C [_B\ b\ ]_B [_C\ c\ \overbrace{]_C]_C}]_C]_S$

To obtain the partial phrase-markers of the sentences of L_1 with respect to G_1, we suppress all but the lowest occurrences of the contiguous left-brackets associated with left-embedding categories, and all but the lowest occurrences of the contiguous right-brackets associated with right-embedding categories in the strings in (3). That is, we eliminate the brackets marked with curly braces in (4) and (5), and in general the strings $([_A)^m$ and $(]_C)^n$ in (3). Applying these operations to the strings in (4) and (5) results in (6) and (7); applying them to all of the strings in (3) results in (8).

(6) a. $[_S [_A\ a\ ]_A [_B\ b\ ]_B]_A [_C\ c\ ]_C]_S$

b. $[_S [_A\ a\ ]_A [_C [_B\ b\ ]_B [_C\ c\ ]_C]_S$

(7) a. $[_S [_A\ a\ ]_A [_B\ b\ ]_B]_A [_B\ b\ ]_B]_A [_C\ c\ ]_C]_S$

b. $[_S [_A\ a\ ]_A [_B\ b\ ]_B]_A [_C [_B\ b\ ]_B [_C\ c\ ]_C]_S$

c. $[_S [_A\ a\ ]_A [_C [_B\ b\ ]_B [_C [_B\ b\ ]_B [_C\ c\ ]_C]_S$

(8) $\{[_S [_A\ a\ ]_A ([_B\ b\ ]_B]_A)^m [_C ([_B\ b\ ]_B [_C)^n\ c\ ]_C]_S : m, n \geq 0\}$

It may now be observed that the partial phrase-markers in (6) and (7) are as adequate as the full phrase-markers in (4) and (5) in representing the ambiguity of the sentences a b c and a b b c, and that in general the strings in (8) adequately represent the ambiguity of the sentences of L_1 with respect to G_1. However, the strings in (8), unlike those in (3), can themselves be accepted by a finite-state automaton; hence there must be a finite transducer T_1 that can associate the structural descriptions (in the form of partial phrase-markers) of a sentence x in L_1 with respect to G_1 in the course of accepting x.

To illustrate the adequacy of partial phraee-markers in representing the constituent structures of sentances generated by a non-center-embedding context-free grammar and the grammatical relations that hold among their constituents, consider the grammar G_2 in (9).

(9) G_2 = (

$N = \{S, \bar{S}, N, \bar{N}, V, \bar{V}, C, D, G\}$

T = {boss, doctor, friend, neighbor, student, teacher
believes, knows, understands, that, the, 's}

$A = \{S\}$

R = (

a. $S \rightarrow \bar{N}\ \bar{V}$

b. $\bar{V} \rightarrow V\ \bar{N}$

c. $\bar{V} \rightarrow V\ \bar{S}$

d. $\bar{N} \rightarrow D\ N$

e. $\bar{S} \rightarrow C\ S$

f. $D \rightarrow \bar{N}\ G$

g. C → that

h. D → the

i. N → {boss, doctor, friend, neighbor, student, teacher}

j. V → {believes, knows, understands}

))

G_2 generates the finite-state language L_2 in (10).[13]

(10) L_2 = {the (boss's,...,teacher's)* {boss,...,teacher} {believes,...,understands} (that the (boss's,..., teacher's)* {boss,...,teacher} {believes,...,understands})* the (boss's,...,teacher's)* {boss,..., teacher}}

In particular, G_2 generates the sentence (11).

(11) The boss believes that the doctor's friend's neighbor knows that the student understands the teacher.

The full phrase-marker that G_2 associates with (11) is given in (12).[14]

(12) $[_S[_{\bar{N}}[_D \text{ the }]_D[_N \text{ boss }]_N]_{\bar{N}}[_{\bar{V}}[_V \text{ believes }]_V[_{\bar{S}}[_C \text{ that }]_C$ -

$[_S[_{\bar{N}}[_D[_{\bar{N}}[_D[_{\bar{N}}[_D \text{ the }]_D[_N \text{ doctor }]_N]_{\bar{N}}[_G \text{ 's }]_G]_D[_N \text{ friend }]_N]_{\bar{N}}$ -

$[_G \text{ 's }]_G]_D[_N \text{ neighbor }]_N]_{\bar{N}}[_{\bar{V}}[_V \text{ knows }]_V[_{\bar{S}}[_C \text{ that }]_C$ -

$[_S[_{\bar{N}}[_D \text{ the }]_D[_N \text{ student }]_N]_{\bar{N}}[_{\bar{V}}[_V \text{ understands }]_V[_{\bar{N}}[_D \text{ the }]_D$ -

$[_N \text{ teacher }]_N]_{\bar{N}}]_{\bar{V}}]_S]_{\bar{S}}]_{\bar{V}}]_S]_{\bar{S}}]_{\bar{V}}]_S$

To obtain the partial phrase-markers of (11) with respect to G_2, we omit all but the lowest of the contiguous labeled brackets that correspond to left- and right-embedding categories in (12)--those marked with curly braces--yielding (13).

(13) $[_S[_{\bar{N}}[_D \text{ the }]_D[_N \text{ boss }]_N]_{\bar{N}}[_{\bar{V}}[_V \text{ believes }]_V[_{\bar{S}}[_C \text{ that }]_C$ -

$[_S[_{\bar{N}}[_D \text{ the }]_D[_N \text{ doctor }]_N]_{\bar{N}}[_G \text{ 's }]_G]_D[_N \text{ friend }]_N]_{\bar{N}}$ -

$[_G \text{ 's }]_G]_D[_N \text{ neighbor }]_N]_{\bar{N}}[_{\bar{V}}[_V \text{ knows }]_V[_{\bar{S}}[_C \text{ that }]_C$ -

$[_S[_{\bar{N}}[_D \text{ the }]_D[_N \text{ student }]_N]_{\bar{N}}[_{\bar{V}}[_V \text{ understands }]_V[_{\bar{N}}[_D \text{ the }]_D$ -

$[_N \text{ teacher }]_N]_{\bar{N}}]_{\bar{V}}]_S$

Despite the elimination of certain labeled brackets, enough information remains in (13) for us to determine completely what constituents occur in (11) with respect to G_2, and what grammatical relations they bear to one another. To determine, for example, that the string the doctor's is a D in (11) with respect to G_2, we note first that it is flanked on the right by $]_D$ in (13) (i.e., that no lexical material intervenes between the lexical string in question and an occurrence of $]_D$), and that the $[_D$ to the immediate left of the string is the first such bracket encountered to the left of this particular right bracket. The fact that this left bracket is also the 'mate' of other right brackets in (13) is immaterial (if you like, you may call this occurrence of $[_D$ a 'superbracket'). To determine that the string that the student understands the teacher is an $\bar{S}$ in (11) with respect to G_2, we note first that it is flanked on the left by $[_{\bar{S}}$, and on the right by the first occurrence of $]_{\bar{V}}$ that is to the right of this occurrence of $[_{\bar{S}}$. Since all occurrences of $]_{\bar{S}}$ in (12) are eliminated in forming the partial phrase-marker (13), we cannot characterize a constituent as an $\bar{S}$ in (11) by its occurrence between $\bar{S}$-brackets in partial phrase-markers of that sentence with respect to G_2. However, $\bar{S}$-constituents are always correctly determined by the procedure just mentioned, and in general it is always possible to find an appropriate 'mate' for those brackets whose 'true mates' have been eliminated by the procedure for forming partial phrase-markers. We can, moreover, establish all of the other constituents of (11) with respect to G_2, given the partial phrase-marker (13).

Not only can we establish the constituents of sentences of L_2 with respect to G_2 on the basis of the partial phrase-markers of those sentences, we can also establish what grammatical relations those constituents have in those sentences. To do so, we define first the notion of an unmatched bracket, as in (14).

(14) A bracket $[_A$ or $]_A$ is unmatched in a substring x of a partial phrase-marker of a sentence with respect to a non-center-embedding context-free grammar if there is no occurrence of $]_A$ or $[_A$ respectively in x.

The definitions of the grammatical relations subject-of and complement-of with respect to G_2 in terms of partial phrase-markers of the sentences generated by that grammar are given in (15).

(15) a. The string $[_{\bar{N}}\ x_1\ ]_{\bar{N}}$ is the subject of the string $[_S [_{\bar{N}}\ x_1\ ]_{\bar{N}}\ x_2\ ]_S$ in a partial phrase-marker Q with respect to G_2 if and only if x_2 does not contain unmatched occurrences of $]_{\bar{N}}$.

b. The string $[_{\bar{S}}\ x_2\ ]_{\bar{V}}$ is the complement of the string $[_{\bar{V}}\ x_1\ [_{\bar{S}}\ x_2\ ]_{\bar{V}}$ in a partial phrase-marker Q with respect to G_2 if and only if x_1 does not contain unmatched occurrences of $[_{\bar{S}}$.

According to these definitions, the string corresponding to *the doctor's friend's neighbor* is analyzed as the subject of the string corresponding to *the doctor's friend's neighbor knows that the student understands the teacher* in (13), but not the one corresponding to *the doctor's friend*; and that the string corresponding to *that the doctor's* friend's neighbor knows that the student understands the teacher is the complement of the string corresponding to believes that the doctor's friend's neighbor knows that the student understands the teacher in (13), but not the one corresponding to that the student understands the teacher, etc. By means of similar definitions, we can establish all of the linguistically significant grammatical relations among the constituents of (11) with respect to G_2 using the partial phrase-marker (13). In conclusion, then, it appears that partial phrase-markers provide completely adequate representations of the structural descriptions of the sentences generated by non-center-embedding phrase-structure grammars.

3.3. Construction of an On-Line Finite-State Transducer Strongly Equivalent to a Non-Center-Embedding Context-Free Grammar. Partial phrase-markers are obtained from full phrase-markers by eliminating the strings of left-brackets that correspond to left-embedding, and the strings of right-brackets that correspond to right embedding. Given a non-center-embedding context-free grammar G, meeting the conditions in Langendoen (1975: 537), there is a mechanical procedure for constructing a strongly equivalent finite-state transducer P that parses the sentences of L(G) from beginning to end as they are accepted. The procedure can be obtained by adapting the construction for 'minimally augmented finite parsers' given in Langendoen (1975: 538-539), as in (16).

(16) Construction of an on-line finite-state transducer $P = (\Sigma, T_1, T_2, S_I, S_F, \Pi)$ that is strongly equivalent to a non-center-embedding context-free grammar $G = (N, T, \{S\}, R)$ meeting the conditions in Langendoen (1975: 537).

I. $\Sigma = \{S_I \cup S_F\} \cup \{(A_1 \ldots A_m)_L, (A_1 \ldots A_m)_R : m \geqq 1, A_1 = S,$ and for all i, j, $1 \leqq i < j \leqq m$, $A_i, A_j \in N$, $A_i \neq A_j$, and there are strings X, $Y \in N^*$ such that $X A_i Y \rightarrow X A_{i+1} Y$ is in R$\}$ are the states of P.

II. $T_1 = T \cup \{\#\}$ is the input vocabulary of P.

III. $T_2 = T_1 \cup \{[_A,]_A : \text{all } A \in N\}$ is the output vocabulary of P.

IV. S_I is the initial state, and S_F is the final state, of P.

V. The members of Π (productions of P) are of the form $\sigma_1\ x \rightarrow y\ \sigma_2$, where $\sigma_1, \sigma_2 \in \Sigma$, $x \in T_1^*$, $y \in T_2^*$, and e is the null string in both T_1^* and T_2^*. In particular:

A. If $B_n \rightarrow a \in R$, then for all $(B_1 \ldots B_n)_L$ in Σ, Π contains:

$$(B_1 \ldots B_n)_L\ a \rightarrow [_{B_n}\ a\ ]_{B_n}\ (B_1 \ldots B_n)_R.$$

B. If $B_n \rightarrow C_1 \ldots C_p \in R$, then for all $(B_1 \ldots B_n) \in \Sigma$, Π contains:

1. if $C_1 \neq B_k$ $(1 \leqq k \leqq n)$:

$$(B_1 \ldots B_n)_L\ e \rightarrow [_{B_n}\ (B_1 \ldots B_n C_1)_L;$$

2. a. if $C_i, C_{i+1} \neq B_k$ $(1 \leqq i \leqq p-1;\ 1 \leqq k \leqq n)$:

$$(B_1 \ldots B_n C_i)_R\ e \rightarrow e\ (B_1 \ldots B_n C_{i+1})_L$$

b. if $C_1 = B_j$, $C_2 \neq B_k$ $(1 \leqq j, k \leqq n)$:

$$(B_1 \ldots B_j)_R\ e \rightarrow e\ (B_1 \ldots B_n C_2)_L$$

c. if $C_{p-1} \neq B_j$, $C_p = B_k$ $(1 \leqq j, k \leqq n)$:

$$(B_1 \ldots B_n C_{p-1})_R\ e \rightarrow e\ (B_1 \ldots B_k)_L;$$

3. if $C_p \neq B_k$ $(1 \leqq k \leqq n)$:

$$(B_1 \ldots B_n C_p)_R\ e \rightarrow\ ]_{B_n}\ (B_1 \ldots B_n)_R.$$

C. 1. $S_I\ \# \rightarrow \#\ (S)_L$ is in Π;

2. $(S)_R\ \# \rightarrow \#\ S_F$ is in Π.

D. Nothing else is in Π.

The nature of the construction in (16) is best revealed by an example. Let $G = G_2$ in (9). By (16), a strongly equivalent, on-line finite-state transducer P_2 is given in (17).[15]

(17) $P_2 = (\Sigma = \{S_I, S_F, (S)_L, (S)_R, (S\bar{N})_L, (S\bar{N})_R, (S\bar{V})_L, (S\bar{V})_R, (S\bar{N}D)_L, (S\bar{N}D)_R, (S\bar{N}N)_L, (S\bar{N}N)_R, (S\bar{V}V)_L, (S\bar{V}V)_R, (S\bar{V}\bar{N})_L, (S\bar{V}\bar{N})_R, (S\bar{V}\bar{S})_L, (S\bar{V}\bar{S})_R, (S\bar{N}DG)_L, (S\bar{N}DG)_R, (S\bar{V}\bar{N}D)_L, (S\bar{V}\bar{N}D)_R, (S\bar{V}\bar{N}N)_L, (S\bar{V}\bar{N}N)_R, (S\bar{V}\bar{S}C)_L, (S\bar{V}\bar{S}C)_R, (S\bar{V}\bar{N}DG)_L, (S\bar{V}\bar{N}DG)_R\}$

$T_1 = \{$boss, doctor, friend, neighbor, student, teacher, believes, knows, understands, that, the, 's, #$\}$

$T_2 = T_1 \cup \{[_S,]_S, [_{\bar{S}},]_{\bar{S}}, [_N,]_N, [_{\bar{N}},]_{\bar{N}}, [_V,]_V, [_{\bar{V}},]_{\bar{V}}, [_C,]_C, [_D,]_D, [_G,]_G\}$

$\Pi = \{$

a. $S_I \; \# \longrightarrow \# \; (S)_L$

b. $(S)_L \; e \longrightarrow [_S \; (S\bar{N})_L$

c. $(S\bar{N})_L \; e \longrightarrow [_{\bar{N}} \; (S\bar{N}D)_L$

d. $(S\bar{V}\bar{N})_L \; e \longrightarrow [_{\bar{N}} \; (S\bar{V}\bar{N}D)_L$

e. $(S\bar{N}D)_L$ the $\longrightarrow [_D$ the $]_D \; (S\bar{N}D)_R$

f. $(S\bar{V}\bar{N}D)_L$ the $\longrightarrow [_D$ the $]_D \; (S\bar{V}\bar{N}D)_R$

g. $(S\bar{N})_R \; e \longrightarrow e \; (S\bar{N}DG)_L$

h. $(S\bar{V}\bar{N})_R \; e \longrightarrow e \; (S\bar{V}\bar{N}DG)_L$

i. $(S\bar{N}DG)_L$'s $\longrightarrow [_G$'s $]_G \; (S\bar{N}DG)_R$

j. $(S\bar{V}\bar{N}DG)_L$'s $\longrightarrow [_G$'s $]_G \; (S\bar{V}\bar{N}DG)_R$

k. $(S\bar{N}DG)_R \; e \longrightarrow]_D \; (S\bar{N}D)_R$

l. $(S\bar{V}\bar{N}DG)_R \; e \longrightarrow]_D \; (S\bar{V}\bar{N}DG)_R$

m. $(S\bar{N}D)_R \; e \longrightarrow e \; (S\bar{N}N)_L$

n. $(S\bar{V}\bar{N}D)_R \; e \longrightarrow e \; (S\bar{V}\bar{N}N)_L$

o. $(S\bar{N}N)_L \{$boss,...,teacher$\} \longrightarrow [_N \{$boss,...,teacher$\}]_N \; (S\bar{N}N)_R$

p. $(S\bar{V}\bar{N}N)_L \{$boss,...,teacher$\} \longrightarrow [_N \{$boss,...,teacher$\}]_N \; (S\bar{V}\bar{N}N)_R$

q. $(S\bar{N}N)_R \; e \longrightarrow]_{\bar{N}} \; (S\bar{N})_R$

r. $(S\bar{V}\bar{N}N)_R \; e \longrightarrow]_{\bar{N}} \; (S\bar{V}\bar{N})_R$

s. $(S\bar{N})_R \; e \longrightarrow e \; (S\bar{V})_L$

t. $(S\bar{V})_L \; e \longrightarrow [_{\bar{V}} \; (S\bar{V}V)_L$

u. $(S\bar{V}V)_L \{$believes,..., understands$\} \longrightarrow [_V \{$believes, ...,understands$\}]_V \; (S\bar{V}V)_R$

v. $(S\bar{V}V)_R$ e $\rightarrow$ e $(S\bar{V}\bar{N})_L$

w. $(S\bar{V}V)_R$ e $\rightarrow$ e $(S\bar{V}\bar{S})_L$

x. $(S\bar{V}\bar{S})_L$ e $\rightarrow [_{\bar{S}}$ $(S\bar{V}\bar{S}C)_L$

y. $(SVSC)_L$ that $\rightarrow$

$[_C$ that $]_C$ $(SVSC)_R$

z. $(S\bar{V}\bar{S}C)_R$ e $\rightarrow$ e $(S)_L$

aa. $(S\bar{V}\bar{N})_R$ e $\rightarrow]_{\bar{V}}$ $(S\bar{V})_R$

ab. $(S\bar{V})_R$ e $\rightarrow]_S$ $(S)_R$

ac. $(S)_R$ # $\rightarrow$ # S_F })

P_2, as it stands, is non-deterministic and therefore highly inefficient as an on-line parser of the sentences of (10). We can, however, easily convert P_2 into a deterministic finite-state transducer P_2' that is also strongly equivalent to G_2. Such a parser is presented in (18).[16]

P_2' = ($\Sigma = \{S_I, S_F, (S)_L, (S\bar{V})_L, (S\bar{N}D)_R, (S\bar{V}\bar{N})_R, (S\bar{V}V)_R, (S\bar{V}\bar{N}D)_R\}$

$T_1 = T_1$ of P_2

$T_2 = T_2$ of P_2

Π = { a. S_I # $\rightarrow$ # $(S)_L$

b. $(S)_L$ the $\rightarrow [_S [_{\bar{N}} [_D$ the $]_D$ $(S\bar{N}D)_R$

c. $(S\bar{N}D)_R$ {boss's,...,teacher's} $\rightarrow$

$[_N$ {boss,...,teacher} $]_N]_{\bar{N}} [_G$'s $]_G]_D$ $(S\bar{N}D)_R$

d. $(S\bar{V}\bar{N}D)_R$ {boss's,...,teacher's} $\rightarrow$

$[_N$ {boss,...,teacher} $]_N]_{\bar{N}} [_G$'s $]_G]_D$ $(S\bar{V}\bar{N}D)_R$

e. $(S\bar{N}D)_R$ {boss,...,teacher} $\rightarrow [_N$ {boss,...,teacher} $]_N]_{\bar{N}}$ $(S\bar{V})_L$

f. $(S\bar{V}\bar{N}D)_R$ {boss,...,teacher} $\rightarrow$

$[_N$ {boss,...,teacher} $]_N]_{\bar{N}}$ $(S\bar{V}\bar{N})_R$

g. $(S\bar{V})_L$ {believes,...,understands} $\rightarrow$

$[_{\bar{V}} [_V$ {believes,...,understands} $]_V$ $(S\bar{V}V)_R$

h. $(S\bar{V}V)_R$ that $\rightarrow [_{\bar{S}}[_C$ that $]_C$ $(S)_L$

i. $(S\bar{V}V)_R$ the $\rightarrow [_{\bar{N}}[_D$ the $]_D$ $(S\bar{V}\bar{N}D)_R$

j. $(S\bar{V}\bar{N})_R$ # $\rightarrow]_{\bar{V}}]_S$ # S_F

3.4. Reduced Phrase-Markers Can Only Be Represented in Bracket-Diagrammatic Form. In section 3.2, we showed how reduced phrase-markers in bracket-diagrammatic form could be constructed from full phrase-markers in that form. If we attempt by similar means to construct reduced phrase-markers in tree-diagrammatic form from full phrase-markers in that form, we find that it cannot be done. For example, if we eliminate the node-connecting lines covered by the curly braces in the tree-diagrams in (4') and (5') in fn. 12, the result is a set of disconnected tree-diagrams that are simply uninterpretable as phrase-markers. If, on the other hand, we remove the node labels, retaining the node-connecting lines, we lose information that it is necessary to represent in order to give a complete account of the constituents that are present. Since no other means of reducing tree-diagrams appears to be available, we conclude that tree-diagrams do not provide an adequate basis on which to construct reduced phrase-markers.

Thus the representations of constituent structure that are constructed on line for sentences that lack center-embedding are reduced phrase-markers in bracket-diagrammatic form. From this observation, what conclusion follows? The strongest conclusion that we could draw is that the representation of phrase structure at every level of linguistic representation is in the form of reduced phrase-markers in bracket-diagrammatic form. Thus we may suppose that the phrase-structure rules of the base component do not introduce left- and right-brackets in left- and right-embedding constructions respectively, but introduce full bracketing elsewhere (i.e., in non-embedding and in center-embedding constructions). Furthermore, we may suppose that syntactic transformations that convert a center-embedding construction into a left- or right-embedding construction eliminate the appropriate left- or right-brackets.[17] Whether or not such a conclusion can be drawn must await the development of the necessary formal apparatus for constructing such grammars. In the meantime, however, we can endorse the weaker conclusion that the appropriate means of representing phrase structure at the various levels of linguistic representation in a transformational grammar in the direct mode is in the form of bracket-diagrams, since these, and only these, form the basis from which we can most directly construct the objects that result from the on-line parsing of sentences.

Footnotes

1. This paper is a conflation of the paper "On Generating the Structural Descriptions of Sentences" that I read at the NELS IX Conference, together with the paper "Finite-State Parsing" that I read at the University of Connecticut on December 8, 1978. I thank Ed Battistella, Janet Fodor, and Mike Stevens for helpful comments and discussion on several of the topics discussed here.

2. We ignore here the problems raised by the rules that insert lexical items into base phrase-markers, which according to the standard theory are not constituent-structure rules, but transformational rules of a certain sort (Chomsky 1965: 88, 98). We take up the matter of lexical-insertion rules below in section 2.1.2.

3. A more common term for these diagrams is 'labeled bracketings'. I use the term 'bracket-diagrams' instead to highlight their similarity to tree-diagrams.

4. According to Chomsky (1965: 89), the lines representing surface phrase-markers must then be converted into bracket-diagrams, which serve as input to the phonological components of grammars. This extra step would not be necessary if surface phrase-markers were already in the form of bracket-diagrams.

5. Since we are now dealing with systems of grammars that directly generate and manipulate phrase-markers, the study of their generative capacities is necessarily of their strong generative capacities.

6. Certain technical difficulties arise, however, whenever a transformation relates two positions, one of which is dominated by a category that does not dominate the other. To take a specific example, consider a transformation that moves a constituent of category B to the beginning of the construction that is the domain of that transformation, and that if B initially occurs within a constituent of type A, it is removed from the domination of A by the transformation. Such a transformation can be expressed in the indirect mode with the variables X_1 and X_2 as in (i).

$$\text{(i)}\quad \begin{array}{ccc} X_1, & B, & X_2 \\ 1 & 2 & 3 \;\longrightarrow \\ 2{+}1 & \emptyset & 3 \end{array}$$

When expressed in the direct mode, however, four variables corresponding to the two in (i) are required (in addition, a fifth, corresponding to the sub-phrase-marker dominated by B must be used). The transformation may be formulated as the type-1 rewrite-rule schema (ii).

$$\text{(ii)}\quad X_1 \langle [_A \rangle\, X_2\, [_B\, X_3\,]_B\, X_4 \langle]_A \rangle\, X_5 \longrightarrow [_B\, X_3\,]_B\, X_1 \langle [_A \rangle\, X_2\, X_4 \langle]_A \rangle\, X_5$$

The added complexity of (ii), however, far from constituting a defect of the theory of transformational grammar in the direct mode, is an advantage, because it reflects directly in the notation in which grammars are written the relative complexity of the application of transformational rule schemata to phrase-markers. Moreover, conditions on transformations can be expressed in the direct-mode theory as conditions on allowable configurations of labeled brackets in the schemata themselves, thus removing from linguistic theory the artificial distinction between conditions on form and conditions on applicability that is required by the indirect-mode theory of transformational grammar.

7. Thus, consider the 'A-over-A' principle proposed by Chomsky (1973: 235-236). As formulated, it would require the phrase of type B that undergoes a transformational rule satisfying the schema (i) in footnote 6 to be the maximal such phrase in the phrase-marker to which the rule applies. The principle is a condition on applicability that must be imposed on rules of the form (i) to eliminate potential ambiguity in the interpretation of the schema. On the other hand, there is no ambiguity in the interpretation of the schema (ii). The phrase of type B that undergoes a rule of the schema must be the maximal such phrase, since otherwise an unmatched occurrence of $[_B$ will appear in the string corresponding to X_2 and an unmatched occurrence of $]_B$ will appear in the string corresponding to X_4, contrary to the assumption that these variables represent well-formed sub-phrase-markers, in which each left-bracket is matched by a corresponding right-bracket. Thus the A-over-A principle is a condition on the form of transformational rules that is an automatic consequence of the theory of transformational grammar in the direct mode.

8. Strictly speaking, this observation holds only for those schemata that insert uninflected lexical items. Schemata that insert inflected lexical items are not context-sensitive, since more than one symbol is rewritten by any such schema (at least one uninflected lexical item together with one symbol representing an inflectional category must appear on the left-hand side of such schemata). Since the schemata

that insert inflected lexical items are not phrase-structure rule schemata, we expect by the principle enunciated directly below that they would appear not in the base component of a standard-theory transformational grammar, but rather in the transformational component. And that is precisely what we find.

9. In fact, it appears not even to be the case that all and only all of the structural descriptions of the sentences of L can be assigned to them by <u>any</u> grammar in the indirect mode, since no sentence of L is structurally ambiguous, and the rule schemata that are conventionally used to generate coordinate structures in the indirect mode all predict that coordinate structures with three or more conjuncts are structurally ambiguous. Since coordinate structures in natural languages in which an overt coordinating conjunction appears only between the last two conjuncts typically are structurally unambiguous, this observation in itself may be sufficient to refute the theory of transformational grammar in the indirect mode.

10. One such grammar contains the following three rules.

(i) $S \rightarrow [_S [_S\ a\ b\ ]_S\ A$

(ii) $A \rightarrow [_S\ a\ b\ ]_S\ A$

(iii) $A \rightarrow [_C\ c\ ]_C [_S\ a\ b\ ]_S]_S$

More perspicuous for linguistic purposes, however, is a context-sensitive grammar with the following four rules:

(iv) $S \rightarrow [_S\ S'\ C\ S'\ ]_S$

(v) $S' \rightarrow S'\ S'\ /\ \_\_\ C$

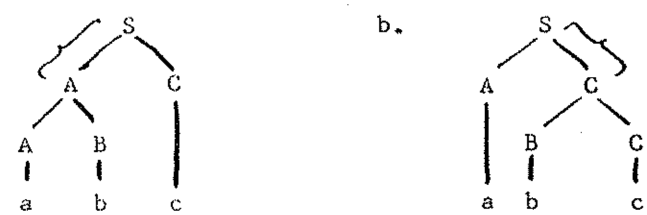
(vi) $C \rightarrow [_C\ c\ ]_C$

(vii) $S' \rightarrow [_S\ a\ b\ ]_S$

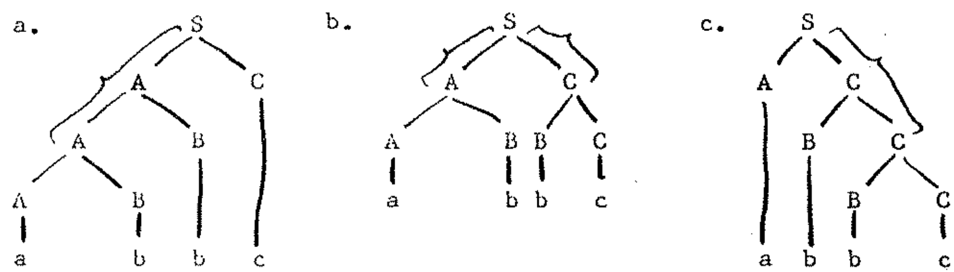
11. Here and elsewhere in this section, we represent grammars as 4-tuples, consisting of a non-terminal vocabulary N, a terminal vocabulary T, axioms A, and rules R.

12. The bracket-diagrams in (4) and (5) correspond to the tree-diagrams in (4') and (5').

(4') a.

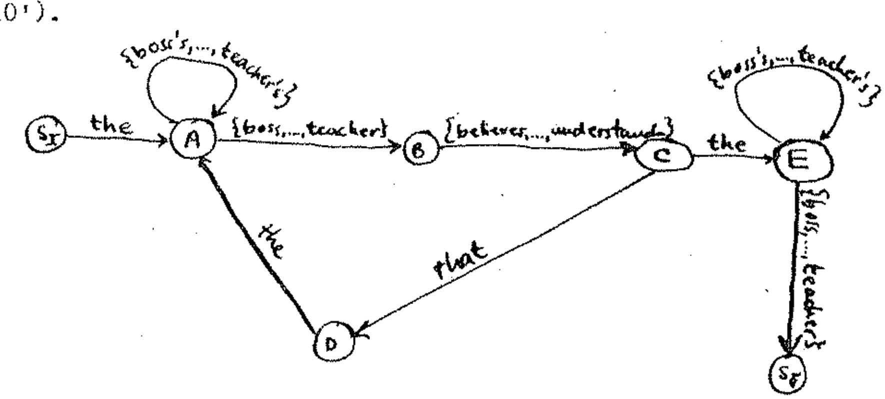


(5') a.



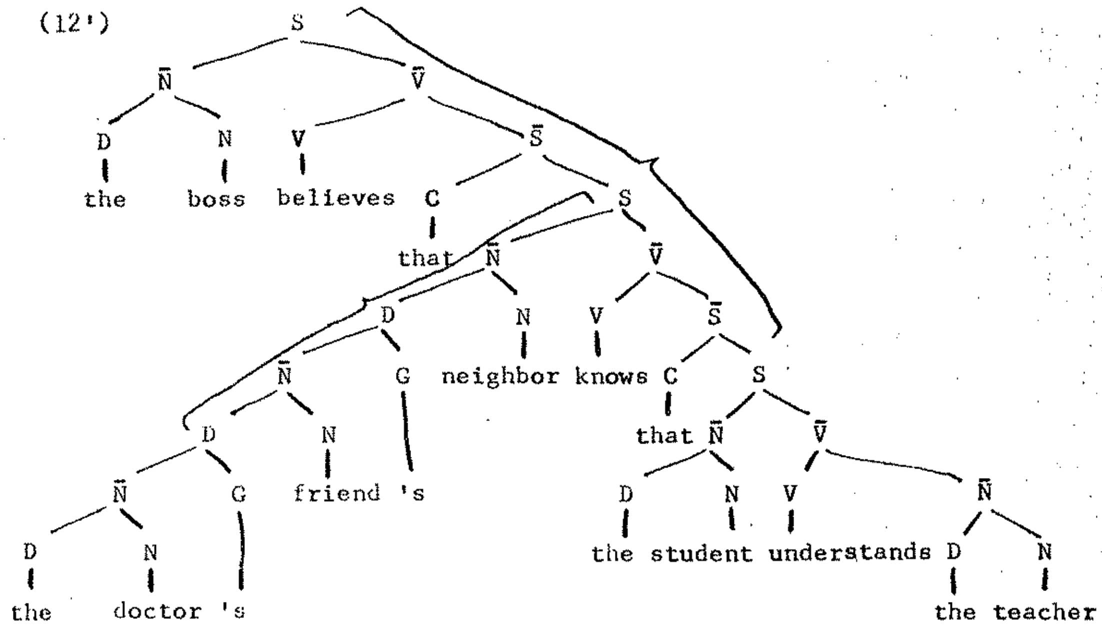
13. A state-diagram of a finite-state automaton that accepts (10) is given in (10').

(10')



14. The tree-diagram that is equivalent to (12) is given in (12').

(12')



15. The construction specifies that $]_{\bar{S}}$ is a member of T_2, even though $]_{\bar{S}}$ does not occur in any rule of Π.

16. The state-diagram for P_2' is very similar to (10') in fn. 13.

17. The intriguing possibility then arises that we could block the transformational introduction of center-embedding by a restriction on the form of transformations that prevents them from introducing the necessary left- or right-brackets.

References

Chomsky, N. 1955. The Logical Structure of Linguistic Theory. Mimeographed. Published 1975. New York: Pantheon.

Chomsky, N. 1959. "On Certain Formal Properties of Grammars." Information and Control 2.137-167.

Chomsky, N. 1961. "On the Notion 'Rule of Grammar'." In R. Jakobson, ed., The Structure of Language and Its Mathematical Aspects. Providence, R.I.: American Mathematical Society, 6-24.

Chomsky, N. 1963. "Formal Properties of Grammars." In R. D. Luce, R. R. Bush, and E. Galanter, eds., Handbook of Mathematical Psychology, Volume 2. New York: Wiley, 323-418.

Chomsky, N. 1965. Aspects of the Theory of Syntax. Cambridge, Mass.: MIT Press.

Chomsky, N. 1970. "Deep Structure, Surface Structure, and Semantic Interpretation." Reprinted in Studies on Semantics in Generative Grammar. The Hague: Mouton, 62-119.

Chomsky, N. 1972. "Some Empirical Issues in the Theory of Transformational Grammar." In Studies on Semantics in Generative Grammar. The Hague: Mouton, 120-202.

Chomsky, N. 1973. "Conditions on Transformations." In S. Anderson and P. Kiparsky, eds., A Festschrift for Morris Halle. New York: Holt, Rinehart and Winston, 232-286.

Chomsky, N. and G. A. Miller 1963. "Introduction to the Formal Analysis of Natural Languages." In R. D. Luce, R. R. Bush, and E. Galanter, eds., Handbook of Mathematical Psychology, Volume 2. New York: Wiley, 279-321.

Kupin, J. J. 1978. "A Motivated Alternative to Phrase Markers." Linguistic Inquiry 9.303-308.

Langendoen, D. T. 1961. Structural Descriptions for Sentences Generated by Non-Self-Embedding Constituent-Structure Grammars. S.B. thesis, M.I.T.

Langendoen, D. T. 1975. "Finite-State Parsing of Phrase-Structure Languages and the Status of Readjustment Rules in Grammar." Linguistic Inquiry 6.533-554.

Langendoen, D. T. 1976. "On the Weak Generative Capacity of Infinite Grammars." CUNYForum 1.13-24.

Lasnik, H. and J. J. Kupin forthcoming. "A Restricted Theory of Transformational Grammar." Theoretical Linguistics 4.

Liberman, M. and A. Prince 1977. "On Stress and Linguistic Rhythm." Linguistic Inquiry 8.249-336.