CONCERNING THE BASE COMPONENT OF A TRANSFORMATIONAL GRAMMAR*

I. INTRODUCTION

The *base component* of a transformational grammar, which specifies what structures are available for the transformations to operate on, has hitherto been regarded by Chomsky and others as involving *rewriting rules*. In Chomsky's most recent treatment (Chomsky, 1965, henceforth *Aspects*), the base component consists of two subcomponents: a *constituent structure* subcomponent, consisting of rewriting rules and rules of another type which will be discussed below, and a *lexicon*. In earlier treatments by Chomsky (Chomsky, 1957, 1962), the lexicon is itself considered to consist of rewriting rules and the base component is regarded as a single system of rewriting rules rather than something consisting of two subcomponents. The term 'rewriting rule' is borrowed from the mathematical theory of rewriting systems. A *rewriting system* is a finite set of rules $\varphi \rightarrow \psi$, where φ and ψ are strings of symbols, and a *derivation* in a rewriting system is a sequence $x_1, x_2, \ldots, x_n$ of strings of symbols such that the first string (or 'line') x_1 of the derivation consists of the single symbol S and each subsequent line can be obtained from the preceding one by substituting a string of symbols ψ for a string φ, where $\varphi \rightarrow \psi$ is a rule of the rewriting system. A derivation is said to be *terminated* if its last line consists only of symbols from a preassigned set called *terminal symbols* (the other symbols which appear in the rules are called *non-terminal symbols*). The last line of a terminated derivation is said to be *generated by* the rewriting system, and the *language generated by* the rewriting system is the set of all strings generated by it.

It should be emphasized that even though the base component of Chomsky's earlier versions of syntactic theory consists entirely of rewriting rules, it is not a rewriting system, since Chomsky imposes an ordering on the rules of the constituent structure component, whereas the term 'rewriting system' refers to an unordered system of rewriting rules (Chomsky, 1963). Thus, while in a rewriting system any derivation is permitted in which each

* Version of January 2, 1968. – I am grateful to Noam Chomsky, Erica Garcia, David Hays, T.R. Hofmann, George Lakoff, Barbara Hall Partee, Paul Postal, John Robert Ross, and William S.-Y. Wang for reading an earlier version of this paper and making valuable suggestions for its improvement. This acknowledgement does not imply that they approve of all that I say here.

Foundations of Language **4** (1968) 243–269.

line is obtained from the preceding one by 'applying a rule', i.e. by replacing the string of symbols on the left half of the rule by that on the right half, in the constituent structure component of a grammar as conceived of by Chomsky, only those derivations are allowed in which the sequence of rule applications accords with the ordering imposed on the rules: if rule A precedes rule B in the ordering, then all applications of rule A must precede all applications of rule B in any derivation. I note in passing that while there is an extensive literature on the mathematical properties of rewriting systems, very little work has been done on the mathematical properties of ordered systems of rewriting rules.[1]

In this paper I will discuss certain inadequacies of the position that the base component of a grammar should involve rewriting rules and will propose an alternative conception of the base component which is free from these inadequacies and which in addition involves neither ordering nor rewriting rules.

II. DERIVATIONS AND TREES

Chomsky's conception of constituent structure component involves rewriting rules of a special type called *phrase structure rules*: rules of the form $\varphi_1 A \varphi_2 \rightarrow \varphi_1 \omega \varphi_2$, where A is a single non-terminal symbol and φ_1, φ_2, and ω are strings of terminal and/or non-terminal symbols, of which at least ω is non-zero. For the remainder of this section I will confine myself to the special case of *context-free* rules, i.e. rules in which the φ_1 and φ_2 are zero and which thus have the form $A \rightarrow \omega$. In the next section I will return to the general case of arbitrary φ_1 and φ_2.

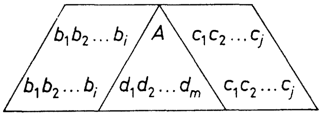
From a derivation that involves only phrase-structure rules one can construct a set of objects called *nodes* with relationships which define the mathematical object known as a *tree* (more precisely, a *rooted, labeled, oriented tree*; the adjectives will be omitted below). Before giving the procedure for constructing a tree from a derivation, it will be necessary to give a definition of the notion 'tree'. A tree is a finite set of objects (called 'nodes') with three relationships ρ 'directly dominates', λ 'is to the left of', and α 'bears the label', satisfying the following axioms: (1) there is a node x_0 such that for no node x does $x \rho x_0$ (x_0 is called the 'root' of the tree); (2) if x is a node distinct from x_0, then $x_0 \rho^* x$, where ρ^* is the relationship which holds between two nodes a and b if there is a chain of nodes $a_1, \ldots, a_n$ such that $a \rho a_1$, $a_1 \rho a_2, \ldots,$ $a_n \rho b$ (ρ^* can be read 'dominates'; this axiom asserts that a tree is 'connected'); (3) if $x \rho y$ and $x' \rho y$, then $x = x'$ (i.e. a tree contains no 'loops'); (4) λ is a partial ordering on the nodes (i.e. if $x \lambda y$ and $y \lambda z$, then $x \lambda z$; if $x \lambda y$, then it is

[1] See, for example, Zwicky (1966), Peters (1966).

false that $y\lambda x$); (5) for any two nodes x and y, if $x \neq y$, then either $x\rho^*y$ or $y\rho^*x$ or $x\lambda y$ or $y\lambda x$; (6) is x is non-terminal (i.e. if there is a z such that $x\rho z$) and $x\lambda y$, then there is an x' with $x\rho x'$ and $x'\lambda y$; if y is non-terminal and $x\lambda y$, then there is a y' with $y\rho y'$ and $x\lambda y'$; (7) every node bears the relation α to exactly one element (its 'label'), the possible labels being a set of objects distinct from the nodes.

Given a derivation involving only phrase structure rules, a tree can be constructed as follows. If the last line of the derivation is $a_1 a_2 \dots a_k$, construct nodes $x_1, x_2, \dots, x_k$ and let $x_1\lambda x_2, x_2\lambda x_3, \dots, x_{k-1}\lambda x_k$ and $x_1\alpha a_1, x_2\alpha a_2, \dots, x_k\alpha a_k$ (N.B. $x_1, \dots, x_k$ must be distinct from each other but $a_1, \dots, a_k$ need not be). This associates to the last line of the derivation a sequence of nodes $x_1, \dots, x_k$. Each higher line of the derivation is made to correspond to a sequence of nodes by comparing it with the line below it. By the assumption that only phrase structure rules are involved, any two consecutive lines can be divided into a common beginning, a common end, and a residue in the middle which for the upper line will consist of a single symbol:

(1)



Suppose that the b's of the lower line correspond to nodes $y_1, \dots, y_i$, the d's to nodes $x_1, \dots, x_m$, and the c's to nodes $z_1, \dots, z_j$. Construct a new node x, let $x\rho x_1, \dots, x\rho x_m$, let $y\lambda x$ for every y which either is one of the y_n's or is dominated by one of them, and let $x\lambda z$ for every z which either is one of the z_n's or is dominated by one of them, and let $x\alpha A$. The upper line of the derivation will correspond to the sequence of nodes $y_1, \dots, y_i, x, z_1, \dots, z_j$. This process of associating each line with a sequence of nodes continues until all the lines of the derivation have been exhausted. It is easy to verify that the set of nodes and the relations thus constructed satisfy axioms (1) to (7). They in addition satisfy the axiom of 'continuity': (8) if $w\rho^*x$, $w\rho^*z$, $x\lambda y$, and $y\lambda z$, then $w\rho^*x$; a tree not possessing this property would be 'discontinuous': there would be a node w which dominated two nodes x and z without dominating all the nodes that were between them.

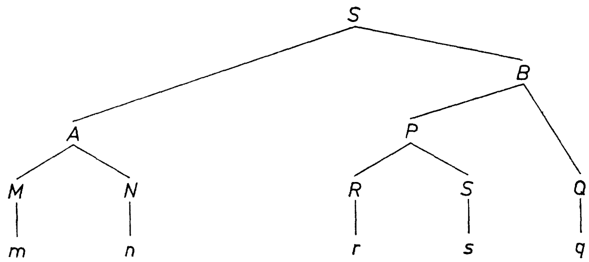
A tree may be represented graphically by a set of points (representing the nodes) arranged so that any point is higher than and connected by a line to any point for which the node corresponding to the former point is in the ρ-relation to the node corresponding to the latter point, with the left-to-right arrangement of the points matching the λ-relation and with labels corresponding to the α-relation. For example, from the derivation

(2)

S
AB
APQ
$MNPQ$
$MNRSQ$
$mNRSQ$
$mnRSq$
$mnRSq$
$mnRsq$
$mnrsq$

one can construct the tree which may be represented graphically as

(3)



A tree whose *terminal nodes* (i.e. nodes which dominate no other nodes) are labeled by morphemes and whose non-terminal nodes are labeled by syntactic category names is called an *immediate constituent structure tree* or *IC-tree*. In a transformational grammar, all syntactic representations of a sentence are in the form of IC-trees: the 'deep syntactic representation' of a sentence is an IC-tree and is converted by the transformational component into the 'surface syntactic representation', which is also an IC-tree, as are all intermediate stages between deep and surface representation; each individual transformation is an operation which converts a class of IC-trees into a class of IC-trees.

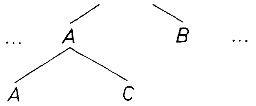
One difficulty with considering the deep IC-tree of a sentence to be constructed from a derivation involving rewriting rules is the fact noted early by Chomsky that two or more IC-trees may correspond to the same derivation. Specifically, to a derivation involving two consecutive lines

(4) $\ldots AB \ldots$
$\ldots ACB \ldots$[2]

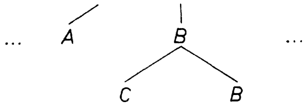
[2] It has been proposed in several works (Postal, 1964; Bach, 1964) that grammars be constrained so as to exclude the rules which would give rise to configurations such as (4), i.e. that rules of the form $A \rightarrow \varphi A$ or $A \rightarrow A\varphi$ be excluded. However, Lakoff and Peters (1966) have recently presented convincing evidence that an adequate grammar of English

can correspond equally well either of the two trees

(5)



(6)



Tree (5) would be appropriate to represent the fact that a rule $A \rightarrow AC$ had applied, tree (6) to represent the fact that a rule $B \rightarrow CB$ had applied. There are three situations under which a pair of consecutive lines such as (4) could arise: (i) the grammar contains the constituent structure rule $A \rightarrow AC$ but not $B \rightarrow CB$; (ii) it contains $B \rightarrow CB$ but not $A \rightarrow AC$; (iii) it contains both. In case (i) only tree (5) should be admitted as an IC-tree, in case (ii) only tree (6), and in case (iii) both trees. Consequently, a derivation (sequence of lines) may not be sufficient to determine what tree should correspond to it, so that one must have recourse to information not in the derivations to determine what trees should be admitted.

This difficulty could be obviated if the base component operated directly in terms of IC-trees rather than through the intermediate stage of a rewriting rule derivation. I know of exactly two proposals for the nature of the base component in which rules operate directly in terms of trees. In the one proposal, which is implicit in such works as Stockwell, Bowen, and Martin 1965 (pp. 8ff.), the base component consists of *tree formation rules*: a rule is interpreted not as an instruction to, say, replace a symbol A by a sequence of symbols BC but rather as an instruction to put two nodes labeled B and C under a hitherto terminal node labeled A. In place of a rewriting rule derivation (sequence of strings of symbols), this proposal substitutes a tree derivation: a sequence of trees, each a part of the following one, the first tree of the sequence consisting of a single node labeled S and the last tree of the sequence being a full IC-tree. I will symbolize a tree formation rule thus: $A > BC$. In the other proposal, to my knowledge first suggested by Richard Stanley (personal communication July 1965), the notion of 'derivation' is dispensed with entirely: the base component is a set of *node admissibility conditions*, for example, the condition that a node is admissible if it is labeled A and directly dominates two nodes, the first labeled B and the second labeled C. I will formulate a node admissibility condition thus: $<A; BC>$. Under the first proposal, a tree is generated by the base component if its terminal

requires rules of those types, specifically the rule NP $\rightarrow$ NP S and the rule schema NP $\rightarrow$ *and* NP^n. The former rule relates to relative clauses, the latter rules to conjoined noun phrases, at least some of which (Lakoff and Peters argue) must be present as such in deep structure rather than derived from conjoined sentences by a transformation of 'conjunction reduction'. The constraint proposed by Postal and Bach in order to insure unique convertability of derivations into IC-trees thus appears to exclude rules which actually must be available in linguistic description.

nodes are labeled by terminal symbols and if there is a derivation which terminates in it. Under the second proposal, a tree is generated by the base component if its root is labeled S, its terminal nodes are all labeled by terminal symbols, and each of its non-terminal nodes meets one or other of the node admissibility conditions of the base component. It will be noted that each of these proposals avoids the difficulty referred to above: a tree formation rule $A > AC$ or a node admissibility condition $<A; AC>$ will give rise to trees such as (5) but not to trees such as (6).

The tree formation rule proposal resembles the rewriting rule proposal in that both involve the notion of 'derivation' and, accordingly, both allow sub-proposals in which only a certain class of derivations is considered to generate admissible deep structures. In particular, both allow the sub-proposal in which the rules have a pre-assigned ordering and only those derivations are admitted in which the steps of the derivation which arise through the application of any particular rule precede the steps which arise through the application of any 'later' rule. They likewise allow the more frequently encountered subproposal that imposes an ordering on the rules but allows one to return to the first rule of the ordering whenever a tree is reached in which S is the only 'unexpanded terminal symbol' present. On the other hand, node admissibility conditions are by nature unordered: the admissibility of a tree is defined in terms of the admissiblity of all of its nodes, i.e. in the form of a condition which has the form of a logical conjunction. Accordingly, the question of the ordering of rules of the base component will be of interest in connection with choosing between these proposals.

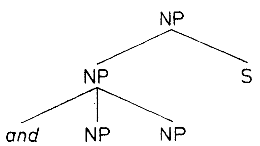
Chomsky, in all the works of his with which I am familiar, treats constituent structure rules as ordered. One fact which casts some doubt on the hypothesis of ordered constituent structure rules is the fact that no examples have been found of dialects differing merely in the ordering of their constituent structure rules, whereas there are numerous examples of adjacent dialects in which the same phonological rules or the same transformational rules apply but in a different order (Klima, 1964; Kiparsky, 1965). There is thus considerable evidence that in acquiring a language not only the transformational and phonological rules of the language must be learned but also the order in which they apply, whereas there is a conspicuous lack of evidence that any ordering of constituent structure rules must be learned. Note further that in the fragments of grammars which have been written using ordered constituent structure rules, little if any work is done by the ordering. The fragments of constituent structure components given in Chomsky (1957) and Chomsky (1962) yield exactly the same IC-trees as they would if interpreted as unordered rather than ordered rules. The ordering of the constituent structure rules in Lees (1960) plays a role only to the extent that it enables one to avoid saying 'elsewhere' in conjunction with certain context-sensitive rules; for example, by placing the first of the following rules before the second:

$$\text{(7)} \qquad N_m + N^0 \rightarrow N_m + Sg$$

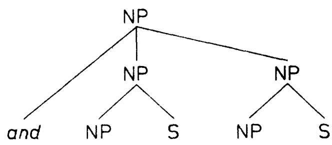
$$\text{(8)} \qquad N^0 \rightarrow \begin{Bmatrix} Sg \\ Pl \end{Bmatrix}.$$

Lees is able to state rule (8) without any indication that N^0 can be rewritten as Pl only when the conditions for rule (7) are not met, i.e. only when it is not preceded by an occurrence of the symbol N_m. However, if one accepts the view of the constituent structure component presented in *Aspects* (p. 139), according to which only context-free constituent structure rules are needed, the context sensitivity of earlier versions of the theory being incorporated into the lexicon rather than the constituent structure component, even this not particularly convincing reason for imposing an ordering on constituent structure rules loses its relevance entirely. In *Aspects*, Chomsky introduces a use of ordering of constituent structure rules which had not appeared in grammars written in the earlier framework, a use which relates to the rules creating 'complex symbols'. I will argue in Section IV that there is no real reason for having such rules in the grammar at all. If this conclusion is accepted, it will mean that here there is likewise no particular reason for imposing an ordering on constituent structure rules.

I thus conclude that nothing currently in print gives any valid reason for preferring a theory with ordered constituent structure rules to one with unordered constituent structure rules. However, it is of interest to consider not merely the question of whether constituent structure rules *must* be ordered but also the question of whether they *may* be ordered: Chomsky points out (personal communication) that since not all unordered sets of constituent structure rules can be ordered in such a way that the ordered system of rules generates the same language as did the unordered system, an affirmative answer to the question of whether the constituent structure rules of all languages may be ordered would further delimit the class of 'possible natural languages' and thus be of interest to linguistic theory. Lakoff and Peters (1966) provides some rules which bear crucially on the question of whether constituent structure rules may be ordered. If the rules NP→NP S and $NP \rightarrow NP^n$ are part of an unordered system of rules, then both the configurations



and



would appear in deep structure, whereas imposing an ordering on the rules would exclude one or other of these two configurations. I maintain that both of these configurations do in fact appear in deep structures of English, specifically in the deep structures of

(9) A man and a woman who met in Vienna live upstairs.
(10) The man who I saw and the boy who you met are similar.

respectively. If this analysis is correct, it would imply that the constituent structure rules of English not only do not have to be regarded as ordered but indeed must be regarded as unordered.

III. CONTEXT-SENSITIVE RULES

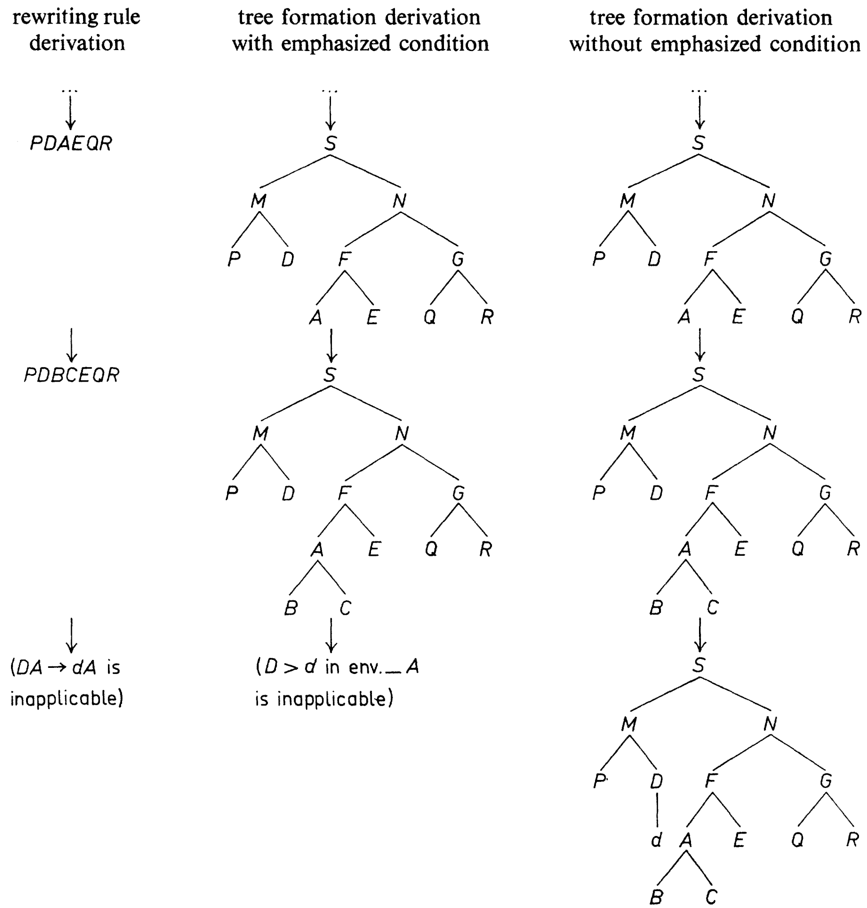
Since the notion of context-sensitive rewriting rule will be crucial to a discussion of the history of certain matters relating to the lexicon of a transformational grammar, I will at this point digress briefly into that topic. Consider first the mathematical object known as a *context-sensitive grammar*, i.e. a rewriting system whose (unordered) rules are of the form $\varphi_1 A \varphi_2 \rightarrow \varphi_1 \omega \varphi_2$, where A is a non-terminal symbol and φ_1, φ_2, and ω are strings of terminal and/or non-terminal symbols, of which at least ω must be non-zero. A rule $\varphi_1 A \varphi_2 \rightarrow \varphi_1 \omega \varphi_2$ may be interpreted as an instruction 'rewrite A as ω when it is in the environment φ_1—φ_2'. However, that interpretation may not be unique: from the formula $AB \rightarrow ACB$ one cannot tell whether he is to rewrite A as AC in the environment —B or to rewrite B as CB in the environment A—. Thus the procedure for constructing an IC-tree fails even more than in the context-free case: in the case of context-free rewriting rules, the derivations did not contain enough information to determine what trees were to be admitted and it was necessary to look at the rules to obtain the additional information needed, whereas in the context-sensitive case, even if one examines the rules he may not find the information needed to characterize the admissible trees. One can again do as before and replace rewriting rules by either tree-formation rules or node admissibility conditions. However, it is important to note that there are several possible interpretations of the terms 'context-sensitive tree formation rule' and 'context-sensitive node admissibility condition', and under most of these interpretations the system of rules will not always generate the same language as the original system of rewriting rules did.[3] The one interpretation of the term 'context-sensitive

[3] A system F of tree formation rules generates a tree if there is a terminated derivation in F whose last 'line' is that tree. The set of trees generated by F may be denoted by T(F). A system G of node admissibility conditions generates a tree if its terminal nodes are labeled with terminal symbols, its root is labeled S, and each of its non-terminal nodes meets a condition of G. T(G) may be used to denote the set of trees generated by G. The language

tree formation rule' under which the same language is generated as by the corresponding rewriting system is the interpretation of a rule '$A > BC$ in env. D_E' as meaning 'put nodes labeled B and C under a hitherto terminal node labeled A if that node is immediately preceded and immediately followed by *hitherto terminal* nodes (i.e. nodes which at this point of the derivation do not dominate any other nodes) labeled D and E respectively. If the emphasized condition were not imposed, the tree formation rules could generate trees which terminated in strings of symbols that were not generated by the original rewriting rules. For example, if the rewriting system also had rules $DA \rightarrow dA$, $B \rightarrow b$, $C \rightarrow c$, $E \rightarrow e$, then a string of symbols of the form $\dots dbce \dots$ could be generated by the tree formation rules without the emphasized condition but would not be generated by the rewriting system or by the tree formation rules with the emphasized condition. The reason is that in the latter two systems, once the symbol A has been rewritten as BC or nodes labeled B and C added under a node labeled A, the definition of derivation would not permit one to (respectively) rewrite D as d or add a node labeled d under the node labeled D, since (respectively) the current line of the derivation would not contain the sequence of symbols DA or in the evolving tree the terminal node immediately to the right of the node labeled D would be labeled B rather than A, as illustrated on p. 252.

It should be clear from this example that a context-sensitive rewriting rule whose non-terminal symbols are interpreted as grammatical categories, e.g. $\mathrm{V} \rightarrow \mathrm{V_t}$ in env. __NP, cannot be interpreted as 'a verb is transitive when followed by *a noun phrase*' but only as 'a verb is transitive when the occurrence corresponding to it of the symbol V is followed by *the symbol NP* at some point in the derivation'. The latter interpretation corresponds to tree-formation rules with the emphasized condition and to rewriting systems, the former to node admissibility conditions. The former might seem at first glance to correspond alternatively to tree-formation rules without the emphasized condition. However, it turns out that, unlike the context-free case, a set of tree-formation rules (with or without the emphasized condition) may generate a smaller set of trees than would the same formulas interpreted as node admis-

L(F) or L(G) generated by a system of tree-formation rules F or a system of node admissibility conditions G is the set of all strings of terminal symbols spelled out by the terminal nodes of the trees of T(F) and trees of T(G) respectively. Many of the familiar results about languages (sets of strings) and their grammars are not true of sets of trees and their grammars. For example, while the intersection of two context-free languages may fail to be context-free, the intersection of two context-free sets of trees is context-free (a set of trees is called context-free if it is generated by a system of context-free node admissibility conditions). While the union of two context-free languages is always context-free, the union of two context-free sets of trees may fail to be context-free. While a finite language is always context-free, a finite set of trees (even a one-member set) may fail to be context-free.

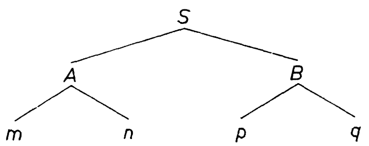


sibility conditions. To cite an obvious example, the set of node admissibility conditions

(11) $< S;\ AB >$
$< A;\ mn \text{ in env. } \_\_p >$
$< B;\ pq \text{ in env. } n\_\_ >$

generates the tree

(12)



whereas the system of tree-formation rules

(13) $S > AB$
$A > mn$ in env. $\_\_ p$
$B > pq$ in env. $n \_\_$

(with or without the emphasized condition) does not.

What this means is that there are at least three different kinds of context-sensitivity, of which any or all or none might be of relevance to the description of natural languages: the 'string context-sensitivity' of rewriting rules and tree formation rules with the emphasized condition, the 'tree context-sensitivity' of node admissibility conditions, and the intermediate type represented by tree formation rules with the emphasized condition. I will argue in the next section that the type of context sensitivity which plays a role in natural language is a restricted type of tree context-sensitivity and moreover that the restricted type of node admissibility conditions which I will consider will generate exactly the same trees as do tree formation rules with a corresponding restriction. I note in passing that it is generally tree context-sensitivity which figures implicitly in most American structuralist grammars of the 1940's[4], a notable exception being Harris (1946), which has a clear interpretation in terms of rewriting rules and string context-sensitivity. Since node admissibility conditions can be used equally well to define classes of discontinuous trees as to define classes of continuous trees, this framework provides a ready way of interpreting the discontinuous structures which figured in these grammars.

IV. THE LEXICON IN A TRANSFORMATIONAL GRAMMAR

The grammatical theory of *Syntactic Structures* treated the lexicon as consisting simply of those constituent structure rules which rewrite a non-terminal symbol as a specific morpheme or string of morphemes. Next to nothing was said in *Syntactic Structures* about the mode of representation of those morphemes; a morpheme appears in the rules either in ordinary orthography or written as a sequence of (presumably atomic) phonological segments or in the form of a special symbol such as *Past* which is given its phonological shape by later rules of the grammar. Soon after the appearance of *Syntactic Structures*, Chomsky adopted Halle's proposal (Halle, 1959) that a morpheme (except for 'grammatical morphemes', which could still be denoted by a special symbol such as *Past*) should appear in the consti-

[4] For example, Bloch (1946). Postal (1964), to my mind incorrectly, interprets Bloch's references to contexts as string context-sensitivity. The one example which Postal cites of context-sensitivity in Bloch's paper cannot be reformulated as a rule of a context-sensitive grammar, since the context is a node which dominates the node in question rather than a node which precedes or follows it.

tuent structure rules in the form of a matrix of +'s, −'s, and blanks, which represent the underlying values of phonological features, the blanks corresponding to feature specifications whose value of + or − can be predicted.

Note that in this conception of the base component, each 'dictionary entry' provides exactly two pieces of information about each lexical item: its underlying phonological shape and the syntactic category to which it belongs, the category being denoted by the symbol which appears to the left of the arrow in the rule which introduces the lexical item in question. With the advent of Katz and Fodor's programmatic sketch of a semantic theory (Katz and Fodor, 1963), the 'dictionary entry' was enlarged to include also a semantic representation. After this (obviously necessary) revision of the theory, the 'dictionary' of a transformational grammar was regarded as consisting of rewriting rules whose left half is the name of a syntactic category and whose right half is a complex entity consisting of a matrix of phonological features and a semantic representation. Alternatively, such a dictionary entry could be viewed simply as a single complex of three types of information (syntactic, semantic, phonological) which characterize the way the lexical item behaves, subject to the constraint that the syntactic information be in the form of a single unanalyzable category symbol.

A large part of *Aspects* is devoted to showing that this last constraint is responsible for several inadequacies in the earlier theory and to exploring the questions of what syntactic information each dictionary entry must contain and what is an appropriate formalism for representing that information. The earlier theory allowed a category to be subdivided only through creating a proliferation of category symbols via constituent structure rules such as

(14)
$$N \rightarrow \begin{Bmatrix} N_{\text{common}} \\ N_{\text{proper}} \end{Bmatrix}$$
$$N_{\text{common}} \rightarrow \begin{Bmatrix} N_{\text{com-human}} \\ N_{\text{com-proper}} \end{Bmatrix}$$
$$N_{\text{proper}} \rightarrow \begin{Bmatrix} N_{\text{prop-human}} \\ N_{\text{prop-nonhuman}} \end{Bmatrix}$$

Even with this proliferation of category symbols, there is often no symbol available for a category implicit in the rules, as will prove the case any time subclassifications crosscut each other, as in (14), where the rules have introduced no symbol for the category 'human noun', which can only be designated by the disjunction $\{N_{\text{com-human}}, N_{\text{prop-human}}\}$; if nouns had first been subdivided into human and non-human and then those categories subdivided into proper and common, then the rules would provide no symbol for the category 'common noun'. What is needed to avoid this defect is a mode of representation involving 'complex symbols' such as $\begin{bmatrix} \text{N} \\ +\text{Human} \\ -\text{Common} \end{bmatrix}$, which represent a syntactic category by a set of simultaneous components rather than by a single unanalyzable symbol. In *Aspects* Chomsky explores such componential representations in detail.

The constituent structure rules which divided syntactic categories into subcategories in works such as Lees (1960) were of three types: (i) rules such as (14), which are generally context-free, and two types of context-sensitive rules; (ii) rules which subclassify members of a category according to what adjuncts they take (e.g. classify 'verb' into 'transitive verb', 'verb with directional complement', etc.), and (iii) rules which subclassify members of a category according to what kinds of lexical items may co-occur with them (e.g. 'verb with human subject', 'verbs with abstract object', etc.). In *Aspects*, these three types of rules are replaced by rules which add three types of features to complex symbols: (i) 'inherent' features such as [+Animate]; (ii) 'strict subcategorization' features such as [+__ NP] (i.e. occurs followed by NP in deep structure, i.e. is transitive); and (iii) 'selectional' features such as [+[−Animate] __ [+Abstract]] (i.e. takes inanimate subject and abstract object). The syntactic information in a dictionary entry will then be in the form of a category symbol such as N or V plus a set of specifications for syntactic features of these three types.[5]

Chomsky's conception of the base component of a grammar, as presented in *Aspects*, is accordingly as follows:

(I) (a) The base component consists of two subcomponents; (i) a constituent structure subcomponent, consisting of an ordered system of rules of two types, context-free rewriting rules and rules which create 'complex symbols'; and (ii) a lexicon subcomponent, which is a set (unordered) of 'dictionary entries', each dictionary entry being a complex of syntactic, phonological, and semantic information, as sketched above.

(b) The constituent structure subcomponent defines a set of 'skeleton'

[5] Lakoff (1965) has shown that a fourth type of syntactic feature must be allowed to appear in dictionary entries, namely features which mark a morpheme as having exceptional behavior with respect to some transformation or other (a morpheme would be exceptional, e.g. by virtue of causing a transformation not to apply even though the conditions for it are otherwise met, by causing a normally inapplicable transformation to apply, or by being allowed to appear only in environments where the conditions for a certain transformation will be met). Lakoff notes that such features are implicit in the categories set up in Lees 1960; for example, Lees' V_{t32} consists of those verbs which are subject to deletion of indefinite object (*eat, steal*, etc.), a transformation which verbs are not normally subject to (a different and much preferable analysis of these verbs is given in Gruber (1965)). In addition, there will have to be features corresponding to grammatical gender, inflectional type, etc. (cf. *Aspects*, pp. 170ff.).

trees, whose terminal nodes are labeled by 'complex symbols' rather than by lexical items.

(c) The deep IC-trees generated by the base component are those which can be obtained by applying the *lexical insertion rule* to the above 'skeleton' trees: under each complex symbol one may insert a morpheme (i.e. a copy of its dictionary entry) if the feature specifications in the dictionary entry are consistent with the feature specifications in the complex symbol (e.g. under a complex symbol with the specifications V and [+[+Animate] __] one may insert a verb which is specified as taking an animate subject or a verb which is unspecified as to the animateness of permitted subjects but not a verb which is specified as not allowing an animate subject).

I turn my attention now to the rules of the constituent structure component which create complex symbols. Consider first the rules which add 'inherent' features to a complex symbol. The following are taken from *Aspects* (p. 83):

(15)
(i) $N \rightarrow [+N, \pm \text{Animate}, \pm \text{Common}]$
(ii) $[+\text{Common}] \rightarrow [\pm \text{Count}]$
(iii) $[-\text{Count}] \rightarrow [\pm \text{Abstract}, - \text{Animate}]$
(iv) $[+\text{Animate}] \rightarrow [\pm \text{Human}]$

These rules are suspicious, since they duplicate information about the language which is already present in the lexicon: the content of rule (i) is that nouns may be animate or inanimate, and common or proper, a fact which the grammar already indicates by containing dictionary entries such as those for *boy, book, John, Egypt* which contain the four combinations of values for the features [Animate] and [Common]. Chomsky points out (*Aspects*, p. 165) that rules such as (15) may be reformulated as rules predicting certain values of features which need not be specified in dictionary entries (*redundancy rules*). For example, if the rules (15) only allow + for the coefficient of [Animate] in complex symbols which contain the specification [+Human], then animateness need not be specified in the dictionary entries of [+Human] nouns and is predictable by a rule [+Human]→[+Animate]. However, it is highly questionable that rules such as (15) rather than the corresponding redundancy rules should be considered part of the grammar. Note that such a treatment of redundancy would be the direct opposite of that normally adopted in phonology (see Stanley, 1967), where only the impossibility of a feature specification need be the subject of a rule (namely a rule predicting the opposite value for that feature) and the possible segment combinations are simply those which are not specifically excluded by such rules. Moreover, a grammar with rules such as (15) has the peculiar property of making the sentence

(16) My neighbor is tall.

appear ambiguous: due to the male/female distinction in personal pronouns, Chomsky's approach would require the base component to contain a rule

(17) $[+\text{Human}] \rightarrow [\pm\text{Female}]$

and (16) would be derivable in two different ways: one with the specification [+Female] and one with [−Female].[6] Furthermore, while semantically distinct lexical items count as different for the purposes of the 'conjunction collapsing' transformation, the occurrence of *neighbor* which Chomsky would derive with specifications of [+Female] and [−Female] do not:

(18) *The balls were well-attended and wooden respectively.
(19) John's neighbors are a man and a woman respectively.

Consider now the rules for introducing the other two types of features (*Aspects*, pp. 94, 97; Chomsky later shows the environment of rules (20) to

[6] The rules which I ridicule here are not totally unreasonable, since they account for the two possible reflexivizations in

(a) My neighbor hurt himself.
(b) My neighbor hurt herself.

The description of such sentences as these raises the question of whether the choice between himself/herself/itself is made on the basis of linguistic properties of the antecedent noun phrase or on the basis of one's knowledge about the intended referent of that noun phrase. Chomsky's treatment demands the former conclusion, mine the latter. According to Chomsky's treatment,

(c) *The waitress hurt himself.

is ungrammatical, according to mine merely semantically anomalous. Some support for the latter view is given by Postal's observation (personal communication, July 1966) that in cases of semantic anomaly, either of the two incompatible elements may be interpreted metaphorically; this test is passed here, since either *the waitress* or *himself* may be interpreted metaphorically, indicating respectively the effeminacy of a waiter or the masculinity of a waitress. In discussing sentences such as these, it is worthwhile to draw a distinction between 'meaning' and 'presupposition'. The information that my neighbor is a woman would be classified as part of the presupposition rather than the meaning of (b), corresponding to the fact that one would not utter such a sentence in order to convey the information that the neighbor is a female but only to convey the information that that individual has suffered an injury. This distinction is supported by the fact that (b) is more correctly paraphrased by

(d) My neighbor suffered an injury.

than by

(e) My neighbor, who is a woman, suffered an injury.

One does not use a sentence like (b) to inform his listener of his neighbor's sex unless he is being rather devious.

be fully predictable):

$$(20)\qquad V \rightarrow CS \text{ in env. } \_\_ \begin{Bmatrix} \text{NP} \\ \# \\ \text{Adjective} \\ \text{Predicate-Nominal} \\ \textit{like}\ \text{Predicate-Nominal} \\ \text{Prepositional-Phrase} \\ \textit{that}\ \text{S}' \\ \text{NP } (\textit{of}\ \text{Det N})\ \text{S}' \end{Bmatrix}$$

$$(21)\qquad V \rightarrow CS \text{ in env. } \begin{Bmatrix} \alpha \text{ Aux } \_\_ \\ \_\_ \text{ Det } \alpha \end{Bmatrix}, \quad \text{where} \quad \alpha \text{ is a N.}$$

Rule (20) adds the feature [+__ NP] to an occurrence of the symbol V which is in the environment __ NP, the feature [+__#] to an occurrence of the symbol V which is in the environment __#, etc.; rule (21) adds the feature [+[+Anim]__] to a V complex symbol which is in the environment $\begin{bmatrix} \text{N} \\ +\text{Anim} \end{bmatrix}$ Aux __, etc. These rules are also suspect since the information which they mark in the evolving structure is completely redundant, and since the lexical insertion rule, which is the only device in the grammar which makes any use of this redundant information, could perfectly well be formulated in such a way as to avoid reference to these two kinds of features: to determine whether a morpheme specified as $\begin{bmatrix} \text{V} \\ +[+\text{Anim}]\_\_ \\ +\_\_\text{NP} \end{bmatrix}$ may be inserted under a V node in a skeleton tree, one need not look at the complex symbol which Chomsky sets up there but could instead simply examine the preceding NP to see whether its head is [+Anim] and examine the 'right sister' of the V node to see whether it is labeled NP. This loosely described procedure of examining adjacent pieces of a tree can be made precise by saying that it amounts to treating each dictionary entry as a context-sensitive node admissibility condition, the context-sensitivity being expressed by the selectional and strict subcategorization features.

I accordingly propose the following alternative to Chomsky's conception of the base component:

(II) (a) The base component is a set (unordered) of rules of two types, constituent structure and lexical.

(b) Both types of rules are node admissibility conditions, the former being context-free and the latter context-sensitive.

(c) The form of the rule is $<A;\ \omega>$ (ω being a non-zero string of non-terminal symbols) for constituent structure rules and $<A;\ x$ in env. $y>$

(where x is a complex of phonological and semantic information[7] and y is expressed in terms of selectional and strict subcategorization features) for lexical rules; the rule asserts that a node in a tree is admissible if it bears the label to the left of the semicolon, directly dominates nodes labeled as indicated to the right of the semicolon, and (in the case of lexical rules) meets the environment condition.

Before discussing the relative merits of proposals (I) and (II), it would be worthwhile to take up Chomsky's comparison of proposal (I) with a third alternative (henceforth, proposal (III)), which is virtually identical with proposal (II) except that it treats trees as constructed from rewriting rule derivations rather than described directly. Specifically, proposal (III) treats the constituent structure component as consisting entirely of rewriting rules (i.e. there are no rules creating complex symbols); lexical items are inserted into an evolving deep IC-tree by means of a revised version of the lexical insertion rule, according to which a lexical item having a strict subcategorization feature $[+A_1A_2\ldots A_r\_\_B_1B_2\ldots B_s]$ may be inserted for a category symbol which is preceded by a string of symbols $\alpha_1\alpha_2\ldots\alpha_r$, where $\alpha_1, \alpha_2, \ldots, \alpha_r$ are strings of symbols belonging respectively to the categories $A_1, A_2, \ldots, A_r$, and followed by a string of symbols $\beta_1\beta_2\ldots\beta_s$, where $\beta_1, \beta_2, \ldots, \beta_s$ are strings of symbols belonging respectively to the categories $B_1, B_2, \ldots, B_s$. Chomsky suggests that (I) may be superior to (III) by virtue of the fact that (I) would impose on any language sharp restrictions on the set of strict subcategorization and selectional features available, whereas (III) would allow in any language the full gamut of conceivable strict subcategorization and selectional features: "It is an interesting question whether the greater flexibility by [proposal (III)] is ever needed. If so, this must be the preferable formulation of the theory of the base. If not, then the other formulation... is to be preferred" (*Aspects*, p. 123).

The two cases of 'greater flexibility' in (III) which Chomsky discusses relate to the place of strict subcategorization rules in the ordering of the constituent structure rules and the question of 'single' and 'double' selectional features. On pp. 99–100, Chomsky asserts that the environment of rules such as (20) is completely predictable, so that the only thing which must be specified in a rule A→CS is its position in the ordering: the environment can only be '$\alpha\_\_\beta$, where $\alpha A\beta$ is a σ, where, furthermore, σ is the category symbol that appears on the left in the rule $\sigma\rightarrow\varphi A\psi$ that introduces A... If this condition is adopted as a general condition on the form of a grammar, then the strict subcategorization rules can simply be given in the form A→CS, the rest being supplied automatically by a convention. In other

[7] x will also have to contain the other types of information mentioned in footnote 5.

words, the only characteristic of these rules that must be explicitly indicated in the grammar is their position in the sequence of rules. This position fixes the set of frames that determine subcategorization.[8] Consider the various possible places where a rule V→CS could appear in a grammar of English which conformed to proposal (I). The rule could directly follow the rule which expands VP, in which case the verb node would be subcategorized in terms of the labels on its *sisters* (two nodes are said to be sisters if they are directly dominated by the same node); another possibility consistent with proposal (I) would be for V→CS to follow the rules which expand some of the constituents of the verb phrase, in which case the verb would be subcategorized not in terms of its sisters but in terms of various nodes dominated by those sisters. Thus, in the first case V→ CS would create features such as [+__ NP], whereas if V→CS followed the rule NP→(Det) N(S), features such as [+__ Det N] would be created instead. Proposal (I) presumably would also allow V→CS to appear in both positions, in which case both sets of strict subcategorization features would be created.

However, it is not clear that strict subcategorization of a node other than in terms of its sisters is ever actually necessary. Chomsky (*Aspects*, p. 102) suggests one possible case of such subcategorization: "a verb could be strictly subcategorized with respect to particular types of PrepP's introduced by PrepP expansion rules". But this case is dubious since there is grave doubt that 'PrepP' exists as a deep structure category in English. According to attractive proposals by Postal and Fillmore[9], all English noun phrases have a preposition, which in certain circumstances is deleted (any preposition is deleted when it is in the subject position; *of* is deleted when it follows an unnominalized verb (with a few exceptions such as *approve of*). The preposition can be considered to originate as a feature of the verb and become attached to the noun phrase by a transformation. Under this proposal, deep structures no longer involve the category PrepP. If, as I conjecture, there are no real cases of strict subcategorization of a node in terms of nodes dominated by its sisters, then not only the environment of the rule but also its place in the ordering is completely predictable: it must directly follow the rule which introduces the category being subcategorized. But if both the content of the rule and its place in the ordering are completely predictable, then everything about the rule is predictable, which means that there is no reason for it to appear at all in a grammar.

[8] *Aspects*, p. 99. The A would have to be constrained to be the head of the item dominated by the σ, so that, e.g. verbs would be subcategorized according to whether they may precede two noun phrases but noun phrases would not be subcategorized according to whether they may occur between a verb and another noun phrase.
[9] Fillmore (1966); Postal's proposal is summarized in Lakoff (1965).

Suppose that indeed a node need only be subcategorized in terms of its sisters. Then the constraint quoted above on the environment of strict subcategorization rules can be sharpened to an assertion that a strict subcategorization feature can only be a specification of the labels on the sisters of a node. If the constituent structure rules were interpreted as tree formation rules rather than rewriting rules, other things being the same as in proposal (I) (call this soon to be dismissed proposal (IA)), there would no longer be any need for an ordering of the constituent structure rules: regardless of what sequence the rules applied in, the sisterhood relation would remain the same throughout the derivation, so that the same strict subcategorization features would be added to each complex symbol by V→CS, etc. even if the rules were treated as unordered. Consequently, even if the grammar is to contain rules creating complex symbols, no ordering need be imposed on the constituent structure rules. Of course, as indicated above, there is no need to have rules such as V→CS at all: proposal (II) allows one to do away with such rules without introducing any extra complexity elsewhere in the rules. The sharpened version of the constraint on strict subcategorization features can be incorporated into proposal (II) by making specific the kind of context-sensitivity referred to in clause (c) of proposal (II): a condition $[+A_1 \ldots A_r \_\_ B_1 \ldots B_s]$ is met by a node if its left sisters are labeled A_1, A_2, ..., A_r (in that order) and its right sisters B_1, B_2, ..., B_s (in that order). This, of course, creates a different kind of node admissibility than that thus far considered, since sisterhood rather than mere adjacency is the relationship on which it is based.

Consider now the matter of 'single' and 'double' selectional features. Chomsky compares two alternatives within proposal (I) for the rule adding selectional features to the verb complex symbol in English[10]:

(22) $$[+\mathrm{V}] \rightarrow \mathrm{CS} \text{ in env. } \begin{Bmatrix} \alpha \text{ Aux } \_\_ \\ \_\_ \text{ Det } \beta \end{Bmatrix}$$

which adds two features such as [+[+Human] __] and [+__ [+Abstract]] to a transitive verb, one marking a selectional restriction between verb and subject and the other between verb and object, and

(23) $$[+\mathrm{V}] \rightarrow \mathrm{CS} \text{ in env. } \alpha \text{ Aux } \_\_ \text{ (Det } \beta),$$

which adds one feature such as [+[+Human] __ [+Abstract], which marks a single restriction between subject, verb, and object. Chomsky notes that he has no clear cases upon which to base a preference for one of these alter-

[10] *Aspects*, p. 118. Presumably (23) is to be interpreted as adding to the [+V] complex symbol all features of the form $[+[x] \_\_ [y]]$, where x is a feature of the α and y a feature of the β, although Chomsky does not explicitly say so.

natives over the other but suggests that the latter type of features may be necessary to account for the occurrence of the verb *command* in sentences (i), (ii), and (iii), but not (iv):

(26) (i) John commanded our respect.
(ii) John commanded the platoon.
(iii) John's resignation commanded our respect.
(iv) *John's resignation commanded the platoon.

Chomsky notes that *command* admits both human and abstract[11] subjects ((i) and (iii)) and admits both human and abstract objects ((i) and (ii)), so that if the dictionary entry simply listed what kinds of subjects it allows and what kinds of objects it allows, it would fail to indicate that the combination of abstract subject with human object (iv) may not occur. Chomsky accordingly interprets the sentences (24) as giving some evidence that the double selectional features of (23) rather than the single selectional features of (22) are required. Chomsky does not take up the possibility that there might be languages which lacked items such as *command* and for which rules such as (22) would be admissible. If there are such languages, then a child in acquiring its native language would have to learn whether the language categorized verbs in terms of 'double' or 'single' selectional features. Either way, proposal (I) would force the constituent structure component to have rules corresponding to only one type of selectional feature, whereas, Chomsky argues, proposal (III) would allow both kinds of features to appear side by side in the dictionary of a single language, thus allowing each language a much wider class of features than are needed to describe it.

However, Chomsky's analysis of *command* can not be accepted. Lakoff (1965, p. E–2) argues that since *command* has a different meaning in (ii) than it does in (i) and (iii), the above problem would not arise: the one meaning (assuming for sake of argument Chomsky's incorrect formulation of the restriction) would have a restriction that subject and object both be human, the other a restriction that the object be abstract. Since the lexicon must in any case indicate which meanings go with what phonological and syntactic information, the one meaning could simply be associated with the 'single' selectional features [+[+Human] __] and [+__ [+Human]] and the other

[11] The inadequacy of this formulation is discussed in section V. I suspect that (26 iii) must actually be derived from the structure which also underlies

(a) John commanded our respect by the fact that he resigned.

or perhaps

(b) John commanded our respect by the way in which he resigned.,

in which case the problem in question here would not arise.

with [+__ [+Abstr]]; no need for 'double' selectional features would arise, since (i), (ii), and (iii) would be generated and correlated with the proper meaning and (iv) would be excluded. Until a clear case of something requiring double selectional features can be found[12], it would seem worthwhile to propose as a linguistic universal that only single selectional features are needed to characterize co-occurrence restrictions between lexical items. This proposed universal would have to be imposed on grammars regardless of whether the base component is regarded as conforming to proposal (I), (II), or (III). Once this constraint is imposed, exactly the same selectional features are available in all three cases, which would mean that the rules which attach selectional features would no longer be a way in which languages might differ from each other according to the predictions from any of the three theories.

Thus, the imposition of the highly plausible universal constraints on grammars that strict subcategorizations be only in terms of sisters and that there be only 'single' selectional features eliminates the 'greater flexibility' of proposal (III) which led Chomsky to conjecture that proposal (III) (and by implication, proposal (II)) would allow a wider class of grammars than corresponded to the possible diversity of natural languages. Rather, with these constraints imposed, proposal (I) requires grammars to contain a large number of rules whose content is completely predictable and a rule ordering on the constituent structure component which is predictable to the extent that it plays any role in the operation of the rules. Proposal (II) does without these superfluous rules and this superfluous dimension of organization, as well as doing away with the *ad hoc* devices which proposal (I) requires to predict the predictable rules and predictable ordering.

Since the extra machinery present in proposal (I) would be needed only if

[12] A possible case of double selectional restrictions is given by verbs such as *marry*, which one might analyze, following Katz (1967, p. 168) as requiring subject and object to have opposite values for the feature [Female]. However, a consideration of this selectional restriction leads one to doubt that it is in any way a linguistic constraint. A treatment such as that of Katz (1967) would require every lexical item having to do with marriage (e.g. *divorce, adultery, engaged, spouse*) to be marked with such a selectional restriction. However, these lexical items would all have semantic representations which made reference to marriage, and the supposed selectional restriction would be predictable from that reference to marriage and thus (contrary to Katz's treatment) would not have to be part of the dictionary entry. However, the prediction of this selectional restriction from the semantic representation of these morphemes would actually be an application of factual information, namely that in our culture marriage is entered into only by two persons of opposite sex. In a culture in which a woman of high status was allowed to assume the male role in a marriage and take women as wives (as Oswald Werner informs me is the case in parts of Dahomey and Nigeria), the lexical items relating to marriage would undoubtedly exhibit 'selectional restrictions' which accorded with this different factual situation. However, that is merely to say that sentences such as *My sister is that woman's spouse* are factually rather than linguistically odd.

one or other of the two universal constraints proposed above is incorrect, it is proposal (II) and not proposal (I) which makes the stronger claim about language. As a result of the specific way in which I have formulated clause (c) of proposal (II), there is yet another way in which it makes a stronger claim about language than proposal (I) does. Clause (c) asserts that lexical rules are of the form $<A; x \text{ in env. } y>$, interpreted as meaning that a node labeled A is admissible if it dominates the complex of information x and satisfies the environmental condition y. This specific form of the proposal implies that the environmental conditions y are not part of the complex of information by which the morpheme is represented in the trees generated by the base component. On the other hand, proposal (I) gives rise to trees which contain selectional and strict subcategorization features. Accordingly, while proposal (I) does not exclude a transformation from making reference to these types of features, proposal (II) does. Proposal (II) thus entails a host of predictions about natural languages such as, for example, the prediction that whenever a transformation deletes the object of a transitive verb leaving no trace such as an agreement marker, then all subsequent transformations will treat the verb exactly the same way as they would treat an intransitive verb.[13]

It should be noted that if the environments in context-sensitive rules are restricted to specifications of the sisters of the node in question, then an unordered set of tree formation rules generates exactly the same set of trees as does the corresponding set of node admissibility conditions. As was noted above, this is not true if a wider class of environments is admitted. Proposal (II) with the above restriction on environments is thus fully equivalent to the proposal obtained by replacing node admissibility conditions by unordered tree formation rules.

V. SELECTIONAL RESTRICTIONS

In this section I will dispute a point which I assumed to be true in the last section: that 'selectional restrictions' have to do with the base component of a grammar.

[13] It has occasionally been claimed (e.g. Lees; 1960, p. 33) that the English adjective-preposing transformation must make reference to strict subcategorization features on the grounds that the *-ing* form of a transitive verb may not be preposed even if its object has been deleted:

(a) There are sleeping children upstairs.
(b) *There are eating men upstairs.

But this generalization does not cover the facts, as is shown by examples such as

(c) Visiting relatives are a nuisance.

Most of the discussion of selectional restrictions which has appeared in print is misleading in that it suggests to the reader that selection has to do with only a small set of very general features such as 'Animate', 'Human', 'Male', 'Physical object', etc. While the selectional restrictions stated in works such as *Aspects* are all in terms of this small range of features, it is important to note that very few of those restrictions have been stated correctly. For example, Chomsky's description of *command* in sentence (24iii) as taking an abstract subject and abstract object does not correctly capture the restriction, since a wide variety of combinations of 'abstract' noun phrases violate the selectional restriction in question:

(27) *Our respect commanded John's decision to resign.
(28) *The fact that $2+2=4$ commanded our respect.

In reality, an enormous range of features would be needed to express the full range of selectional restrictions to be found in English, as is clear from a consideration of the selectional violations in

(29) *That verb is in the indicative tense.
*Bernstein's theorem is non-denumerable.
*John diagonalized that differentiable manifold.
*That electron is green.
*I ate three phonemes for breakfast.
*He pronounces diffuseness too loud.
*My hair is bleeding.
*That unicorn's left horn is black.

I maintain that selectional restrictions are actually semantic rather than syntactic in nature, that the full range of properties which figure in semantic representations can figure in selectional restrictions and that only semantic properties figure in selectional restrictions, and that it is the semantic representation of an entire syntactic constituent such as a noun phrase rather than (as implied by the proposals of *Aspects*) merely properties of the lexical item which constitutes its 'head' that determines whether a selectional restriction is met or violated. Lakoff and Ross (1967) have observed that paraphrases satisfy the same selectional restrictions, e.g. that there is no verb with a selectional restriction which would be met if *a bachelor* were used as its subject but violated if *an unmarried man* were used as its subject. There are also clear cases of pairs of sentences in which the same selectional restriction is violated by (in the one case) material introduced by the head noun and (in the other case) by material introduced by a modifier, for example,

(30) a. *My sister is the father of two.
b. *My buxom neighbor is the father of two.

and in which different adjuncts to a given head produce selectional violation and non-violation respectively:

(31) a. *The arm of the statue is bleeding.
b. My arm is bleeding.

Moreover, despite Chomsky's assertion that "*every* [original emphasis] syntactic feature of the Subject and Object imposes a corresponding classification on the verb" (*Aspects*, p. 97), no clear case has been adduced of a selectional restriction which involves a non-semantic feature. For example, while one might suggest that *name* has a selectional restriction involving the feature 'proper' (vs. 'common'):

(32) a. They named their son John.
b. *They named their son the redheaded boy over there.

or that *count* has a selectional restriction involving the feature 'countable' (vs. 'mass'):

(33) a. I counted the pigs.
b. *I counted the sand. (in the sense of 'enumerate', as opposed to the sense 'include in an enumeration'),

the restrictions actually involve the semantic properties of designating a proper name:

(34) They named their son something absurd.

and designating a set:

(35) *I counted the pig.
(36) I counted the crowd.

In McCawley (forthcoming), I argue that the selectional restrictions imposed by a lexical item can be predicted from its meaning and that the supposed counterexamples to this assertion, i.e. items which supposedly have the same meaning but different selectional restrictions, actually have different meanings. For example, the various Japanese verbs which may be glossed 'put on (said of headwear)' (*kaburu*), 'put on (said of gloves)' (*hameru*), 'put on (said of coats, shirts, etc.)' (*kiru*), etc. actually mean the specific manner of putting on, as is shown by the fact that when one puts on an article of clothing in an unnatural manner (e.g. puts a pair of socks on his hands, uses a necktie to hold up his trousers), it is not the garment but the manner in which it is put on that determines the choice of verb: putting socks on one's hands would demand the use of *hameru* rather than *haku* and putting a shirt on top of one's head would demand the use of *kaburu* rather than *kiru*. Note that according to this analysis, the sentence

(37) kutu o kabutta. (*kutu*='shoes', *o*=acc. case, *kabutta*=past tense of *kaburu*)

is not linguistically odd but will almost always be token-odd: there is nothing wrong with the sentence per se but only with using it to refer to putting on shoes in the way that people normally put them on. Thus *kaburu, hameru,* etc. probably have identical selectional restrictions.

Fillmore (personal communication) has proposed that selectional restrictions are not restrictions imposed by a lexical item on other syntactic constituents but rather presuppositions about the intended referents of those constituents, e.g. that the selectional restriction imposed by *diagonalize* is not that its object have a semantic representation consistent with the semantic features that characterize matrixhood but rather the presupposition that the intended referent of the object be a matrix, and that selectional violation consists not in a semantic representation violating a condition imposed by some lexical item but rather in a contradiction between the assertions and presuppositions made about the various entities to which the sentence refers. I regard this as the most worthwhile proposal which has yet been made about selectional restrictions. First of all, it explains why under the earlier proposals a selectional restriction could require the absence but could not require the presence of a 'semantic marker'[14], e.g. that a restriction 'requires female subject' was met equally well by a noun phrase unspecified as to sex (e.g. *my neighbor*) as by one specified as female (e.g. *my sister*). Secondly, Fillmore's proposal (unlike the earlier proposals) requires no modification to make it consistent with the conclusion (McCawley, forthcoming) that the lexical material of a noun phrase may originate in a 'higher' sentence than the one in which it appears, e.g.

(38) John denies that he kissed the girl who he kissed.,

which is most normally interpreted with *the girl who he kissed* not being part of John's denial but being the speaker's description of the girl who John was talking about. Third, as pointed out by Fillmore, this proposal allows an item to impose a selectional restriction 'on itself' in the sense that *bachelor* may be regarded not as meaning 'human, male, adult, unmarried' but rather as having the meaning 'unmarried' and the 'selectional restriction' (presup-

[14] Katz (1966, p. 160) asserts that a restriction may require the presence of a marker: "the reading for 'burn up' will have the selectional restriction ⟨(Physical Object)⟩ which permits a reading for the nominal subject of an occurrence of 'burn up' to combine with it just in case that reading has the semantic marker (Physical Object)". However, Katz gives no justification for this formulation.

position concerning intended referent) 'human, male, adult', which fits well the fact noted by Fillmore that one may apply the word *bachelor* to someone known to be a male adult human in order to express that he is unmarried but may not apply *bachelor* to someone known to be an unmarried adult human in order to express that he is male.

The reader will have undoubtedly noticed that the last sentence is inconsistent with what I said earlier about the selectional restrictions imposed by an item being predictable from its meaning: 'male' surely is not predictable from 'unmarried'. I can maintain my earlier assertion only by revising the representation of *bachelor* to the extent of replacing 'unmarried' by 'not having a wife'. If I assert that the selectional restriction imposed by a property is that the item to which the property is applied be a 'candidate' for having the property[15] and assert that candidacy for not having a property is the same as candidacy for having the property, then 'human', 'male', and 'adult', which are necessary to make one a candidate for having a wife, will be imposed as 'selectional restrictions' by *bachelor*. This approach is confirmed by the fact that the other conditions which are involved in candidacy for having a wife also play a role in the applicability of *bachelor*; for example, one would not call a Roman Catholic priest a bachelor, even though he is human, male, and adult, and has no wife.

Chicago

BIBLIOGRAPHY

Bach, Emmon: 1964, *An Introduction to Writing Transformational Grammars*, Holt, Rinehart, and Winston, New York.

Bloch, Bernard: 1946, 'Studies in Colloquial Japanese II: Syntax', *Language* **22**, 138–143.

Chomsky, Noam A.: 1957, *Syntactic Structures*, Mouton, The Hague.

Chomsky, Noam A.: 1962, 'A Transformational Approach to Syntax', in *Third Texas Conference on Problems of Linguistic Analysis of English* (ed. by A.A. Hill), pp. 124–158. University of Texas.

Chomsky, Noam A.: 1963, 'Formal Properties of Grammars', in *Handbook of Mathematical Psychology* (ed. by R. Duncan Luce, Robert R. Bush, and Eugene Galanter), vol. 2, pp. 323–418. Wiley, New York.

Chomsky, Noam A.: 1965, *Aspects of the Theory of Syntax*, M.I.T. Press, Cambridge, Mass.

Fillmore, Charles J.: 1966, 'A Proposal Concerning English Prepositions', *Georgetown Monograph Series on Languages and Linguistics* 19, 19–34.

Gruber, Jeffrey: 1965, *Studies in Lexical Relations*, M.I.T. Dissertation.

Halle, Morris: 1959, *The Sound Pattern of Russian*, Mouton, The Hague.

Harris, Zellig: 1946, 'From Morpheme to Utterance', *Language* **22**, 161–183.

[15] While this discussion is confined to 'properties', i.e. one-place predicates, it can be extended in an obvious fashion to multiple-place predicates.

Katz, Jerrold J.: 1966, *The Philosophy of Language*, Harper & Row New York.
Katz, Jerrold J.: 1967, 'Recent Issues in Semantic Theory', *Foundations of Language* **3,** 124–194.
Katz, Jerrold J. and Fodor Jerry A.: 1963, 'The Structure of a Semantic Theory', *Language* **39,** 140–210.
Kiparsky, R.P.V.: 1965, *Linguistic Change,* M.I.T. Dissertation.
Klima, Edward S.: 1964, 'Relatedness Between Grammatical Systems', *Language* **40**, 1–20.
Lakoff, George P.: 1965, *On the Nature of Syntactic Irregularity*, Indiana University Dissertation.
Lakoff, George P. and Peters, P.S.: 1966, 'Phrasal Conjunction and Symmetric Predicates', Report NSF-17 of Harvard University Computation Laboratory.
Lakoff, George P. and Ross, John R.: 1967, 'Is Deep Structure Necessary?'. Duplicated.
Lees, Robert B.: 1960, *The Grammar of English Nominalizations*, Supplement to vol. 26 of *International Journal of American Linguistics.*
McCawley, James D.: (forthcoming), 'Where do Noun Phrases Come From?', to appear in Roderick Jacobs and Peter S. Rosenbaum (eds.), *Readings in English Transformational Grammar.*
Peters, P.S.: 1966, 'A Note on Ordered Phrase Structure Grammars', in Report NSF-17 of Harvard University Computation Laboratory.
Postal, Paul M.: 1964, *Constituent Structure*, Supplement to vol. 30 of *International Journal of American Linguistics.*
Stanley, Richard: 1967, 'Phonological Redundancy Rules', *Language*, **43**, 393–436.
Stockwell, Robert P., Bowen, Donald J. and Martin, John W.: 1965, *The Grammatical Structures of English and Spanish*, University of Chicago.
Zwicky, Arnold M.: 1966, 'On Context-sensitive Recognition and Related Matters', Mimeographed (Urbana, Ill.).