Joan W. Bresnan

On the Form and Functioning of Transformations

0. Introduction

Many linguists are interested in the traditional question, What is the relation between the grammatical form and the logical form of sentences of natural language? Let us suppose that the grammatical form of sentences is given by a transformational grammar $G$ that generates each sentence with an associated syntactic structural description $\sigma$. Let $\sigma$ be the sequence of constituent structure trees in a transformational derivation of a sentence $a$ and denote the set of all such pairs $(a, \sigma)$ by $S(G)$. Let us also suppose that the logical form of sentences is represented in an appropriate logic, or disambiguated interpreted language $L$, and denote the well-formed formulas of $L$ by $F(L)$. Let us suppose finally that there is a “translation” mapping $\Phi$ from $S(G)$ to $F(L)$. Then our question is, What are the properties of $\Phi$? Is it one–many, many–one, or one–one, into or onto, partial, recursive, etc.? Obviously, the answers to these questions depend upon finding the appropriate $L$ and $G$, and these are not known in advance of empirical investigation.¹

The basic issue that has divided transformational grammarians for nearly ten years is whether $\Phi$ is trivial. Linguists of the generative semantics school believe that it is, and those of the interpretive semantics school believe that it is not. Behind the disagreement are divergent methodologies and philosophies of science.

On the one side is the belief that it is methodologically preferable to identify deep structure (the initial tree in $\sigma$) with logical form, for then the mapping from logical form to surface structure (the final tree in $\sigma$) would be carried out entirely by a “homogeneous” set of rules, and, other things being equal between two theories, “the theoretically more impoverished [i.e. the one with fewer kinds of rules] is always more highly valued.” (Postal (1972, 136)) This view is sometimes supported by a kind of reductionism that maintains that meaning and sound are the only “observables” of language, hence that semantic and phonetic structures alone are empirically

¹ This point is underscored by Hintikka’s argument that quantifiers in natural language are beyond the range of quantification theory; see Hintikka (1974).

Incidentally, I do not wish to suggest that semantic interpretation for natural language sentences must involve a translation into logical form. For interesting reflections on this question, see Cooper (1975), Fodor (1975), Jackendoff (1976), and Miller and Johnson-Laird (1976).
justifiable or "natural", and therefore that transformations must directly convert natural semantic or "logical" structures into surface structures without employing an independently defined set of syntactic structures. (Cf. Kiparsky (1973, 52) and Postal (1974, xiv).)

On the other side is the view that linguistic theory has an explanatory aim requiring that we discover the limits of natural grammatical form by which human languages are distinguished from countless conceivable but unspeakable formal systems. As Chomsky puts it (1972, 67; 68):

The fundamental problem of linguistic theory is to account for the choice of a particular grammar, given the data available to the language learner. To account for this inductive leap, linguistic theory must try to characterize a fairly narrow class of grammars that are available to the language learner; it must, in other words, specify the notion "human language" in a narrow and restrictive fashion.

If enrichment of theoretical apparatus and elaboration of conceptual structure will restrict the class of sets of derivations generated by admissible grammars, then it will be a step forward. . . . One who takes the more "conservative" stance, maintaining only that a grammar is a set of conditions on derivations, . . . is saying virtually nothing.

The debate over the general relation between grammatical form and logical form reappears when we ask, What are the categories of grammar? The generative semantics program included an attempt to reduce the syntactic categories of grammar to the categories of a variant of the predicate calculus by extending the use of transformations. Quantifiers were treated as "higher verbs" subject to special "lowering" transformations. Syntactic differences between other categories—such as noun and verb—were coded onto individual lexical items in the form of rule features, which "triggered" the applications of differentiating transformations. This program is described by McCawley (1971, 219ff), concluding that

the resulting inventory of categories . . . matches in almost one-to-one fashion the categories of symbolic logic, the only discrepancy being that the category VP has no corresponding logical category. However, Lakoff argued, there is in fact virtually no evidence for a syntactic category of VP. . . .

Chomsky (1970) argued that this reduction of grammatical categories could not be accomplished without loss of linguistically significant generalizations, and proposed a new theory of syntactic categories, known as the X theory, in which their similarities and differences could be represented systematically. Underlying Chomsky's position is "the thesis of the autonomy of syntax", which is essentially that logical form is not given in advance of grammatical form: the categories and rules of grammar—the "formation rules" of natural language, if you like—constitute an independent system whose formal properties must be determined by empirical investigation.

What is meant by "the autonomy of syntax" can be better appreciated when we consider recent work in the logical analysis of natural languages. For example, Montague (1973) demonstrated that it is possible to treat quantified noun phrases
FORM AND FUNCTIONING OF TRANSFORMATIONS

and proper nouns in the syntactically natural way—as belonging to the same category NP; he also showed that semantically opaque simple verb phrases (e.g. *seeks a unicorn*) can be treated in the syntactically natural way—as simple VPs. Recently, Cooper and Parsons (1974) have constructed transformational grammars—of both generative semantic and interpretive types—that are provably equivalent to Montague's grammar. What, then, are the "natural" syntactic categories of human languages? Should quantifiers be elements of NPs or Vs? Should VPs be eliminated or retained? Since there is no unique "logical form" given in advance of grammatical form, we must look elsewhere for answers.

However, there are empirical ways of determining the "natural" syntactic categories of human languages. The most valuable tool for this is the syntactic transformation. For the fundamental property of transformations is their *structure dependence*: unlike phrase structure rules, or categorial rules, which specify categories in terms of their immediate constituents, transformations depend upon constituent *structure* and the *structural analyses* of phrases; unlike semantic rules, transformations do not depend upon the semantic composition or lexical contents of phrases. (See Bresnan (to appear) for a fuller discussion of these issues.)

In this study I will define and justify a "relativized" A-over-A Principle that, when applied to cross-categorial transformations formulated within the $X$ theory, reveals natural classes of syntactic categories. In sections 2 and 3 of this article I will develop the necessary concepts for an adequate formulation of this principle. In section 4 I will show that there are certain classes of syntactic categories that behave under different transformations like a single category of type "A" with respect to the Relativized A-over-A Principle. A unified description is given for left-branch, pied piping, and other phenomena previously described by unrelated constraints or unaccounted for.

1. The A-over-A Principle

Chomsky (1973, 235) proposes as a general condition on transformations the following A-over-A Principle:

(1) If a transformation applies to a structure of the form $[\alpha \ldots [\Lambda \ldots ]\Lambda \ldots ]\alpha$, where $\alpha$ is a cyclic node, then it must be so interpreted as to apply to the maximal phrase of the type A.

This principle reduces the possible proper analyses of a given structure with respect to a given transformation. For example, the structure in (3) has two proper analyses with respect to the transformation in (2), as indicated.

(2) $X - NP - Aux - V - (P) - NP - Z$

$$
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\end{array}
\quad \rightarrow \\
\begin{array}{ccccccc}
1 & 6 & 3 & 4 & 5 & \phi & 7 \\
\end{array}
$$

These points should not be obscured by Montague's use of the categorial notation in this paper. Cf. Montague (1970), Gabbay (1973), and especially Partee (1975).
In the proper analysis $\pi_1$, term 6 is a phrase of the type NP; in $\pi_2$, term 6 is also a phrase of the type NP. Since the latter NP properly dominates the former NP, the latter is the maximal phrase of the type NP, and, by (1), the structure must be transformed under the analysis $\pi_2$.

The A-over-A Principle is verified only if the possible applications it permits are empirically justified and those it excludes are empirically unjustified. In the case of (3), transformation under $\pi_2$ yields a grammatical sentence (4), whereas transformation under the excluded proper analysis $\pi_1$ would yield an ungrammatical sentence (5):

(4) Children('s) smoking isn't approved of by doctors.

(5) *Children aren't approved of smoking by doctors.

Various well-known counterexamples to the A-over-A Principle can be resolved when the principle is properly reformulated as a condition relating the form of transformations to their functioning. The main idea in reformulating the A-over-A Principle will be to relativize it to the structural conditions of transformations. In order to define this idea precisely, it is necessary to establish certain limitations on the form of transformations, which is accomplished in section 2.
2. The Form of Transformations

The key concepts in this section are \( |\mathcal{P}|_\pi \) (the value of a predicate \( \mathcal{P} \) under a proper analysis \( \pi \)) and target predicates in a structural condition, which are defined in 2.2 and 2.4, respectively.

I will presuppose the general definition of transformations given in Peters and Ritchie’s valuable study (1973). The phrase structure component of the grammar strongly generates an infinite set of well-formed terminal labelled bracketings. Each such labelled bracketing \( \phi \) admits of one or more (standard) factorizations, which are \( n \)-tuples of labelled bracketings \( (\phi_1, \ldots, \phi_n) \) that satisfy the conditions (i) \( \phi = \phi_1 \cdots \phi_n \) and (ii) for each \( i = 1, \ldots, n \), the leftmost symbol of \( \phi_i \) is not a right bracket, nor is the rightmost symbol a left bracket. For example, (6) has the factorizations given in (7)–(10):

\[
\]
\[
\text{(7)} \quad (\phi, \psi, \chi)
\]
\[
\text{(8)} \quad (\phi, \psi \chi)
\]
\[
\text{(9)} \quad (\phi \psi \chi)
\]
\[
\text{(10)} \quad (\phi \psi \chi)
\]

where \( \phi = [s[NP[NMary]N]NP, \psi = [VP[vlikes]v, and } \]
\[
\]

Because we allow \( \epsilon \), the labelled bracketing of length 0, as a possible factor, (6) has infinitely many additional factorizations (e.g. \( (\phi, \epsilon, \psi, \chi) \)).

The labels of the labelled bracketings come from the finite set \( V_N \) of nonterminal symbols, or categories, of the grammar; from \( V_N \) as well are constructed basic predicates, which are the components of the structural conditions of transformations. For every category \( C \in V_N \) and natural numbers \( n, i, j, C^i_{n,j} \) is a basic predicate; such predicates will apply to \( n \)-term factorizations and will mean, intuitively, “the \( i \)th through \( j \)th factors, when concatenated, constitute a phrase of the category \( C \)”.

\( C^i_{n,j} \) is true of a factorization \( (\phi_1, \ldots, \phi_n) \) if \( 1 \leq i \leq j \leq n \) and \( \phi_1 \cdots \phi_j \) is a \( C \) (which means that corresponding labelled brackets \([C \text{ and }]C\), possibly interspersed with other labelled brackets, surround the terminal symbols of \( \phi_1 \cdots \phi_j \)). For example, \( VP^2_{2,2} \) is true of (8) but not (9).

An \( n \)-term structural condition of a transformation is a Boolean combination of such \( n \)-ary basic predicates \( C^i_{n,j} \) (i.e. a compound predicate constructed from the basic predicates together with the connectives “\&”, “\to”, and “\to\to”). Thus the structural condition \( NP^3_{1,1} \) & \( VP^3_{2,3} \) & \( N^3_{3,3} \) is true of (7). In the familiar informal notation for transformations, this structural condition would be written “\( NP \to [VP X \to N]_{VP} \)”; note that “variables” correspond to factors of which nothing is predicated.
An n-term transformation is an ordered pair \((\mathcal{C}, \mathcal{M})\), where \(\mathcal{C}\) is an n-term structural condition and \(\mathcal{M}\) is an n-term transformational mapping. A proper analysis of a well-formed terminal labelled bracketing \(\phi\) for a transformation \(T = (\mathcal{C}, \mathcal{M})\) is a factorization \((\phi_1, \ldots, \phi_n)\) of \(\phi\) of which \(\mathcal{C}\) is true and upon which \(\mathcal{M}\) is defined.

The class of transformations defined in this way is far too broad. The informal notation in which transformations are commonly formulated in empirical work incorporates many implicit limitations, a few of which I will make explicit in the following subsections.

2.1. Structural Conditions

We can dispense with the Boolean connectives "\(\neg\)" and "\(\lor\)" without loss of descriptive adequacy.\(^3\) To describe optional elements, e.g. "\((P)\)", we admit for any category \(C\), new basic predicates of the form \((C)^n_{i\rightarrow j}\); \((C)^n_{i\leftarrow j}\) is true of a factorization \((\phi_1, \ldots, \phi_n)\) if \(C^n_{i\leftarrow j}\) is true of it or if \(\phi_1 \cdots \phi_j = \varepsilon\). (Possibly for all such predicates \(i = j\).) Further, we do not admit "identity" predicates into structural conditions; these are restricted to conditions on transformational mappings (cf. 2.3).

Thus, every structural condition \(\mathcal{C}\) is simply a conjunction of basic n-ary predicates. We further impose the following admissibility condition on all structural conditions \(\mathcal{C}\): if \(\mathcal{P}^n_{i\rightarrow j}, \mathcal{P}^n_{j\leftarrow k}\) are basic predicates occurring in \(\mathcal{C}\), then it is not the case that \(h < j \leq i < k\) or that \(j < h \leq k < i\). This ensures that predicates "overlap" only if one is "included" in the other (e.g. "\([NP\rightarrow VP][NP\rightarrow C]_{VP}\" is not allowed). This condition reflects a fundamental property of constituent structure trees, namely that the dominance and precedence relations are complementary. Hence we may think of it as part of the universal definition of structure-dependence of transformations.\(^4\)

Finally, to eliminate possible redundancy and conflict, we require that if any basic predicate \(\mathcal{P}^n_{i\rightarrow j}\) occurs in \(\mathcal{C}\), no other predicate \(\mathcal{P}^n_{j\leftarrow k}\) (having the same subscripts) occurs in \(\mathcal{C}\).

2.2. Proper Analyses

The limitations we have imposed on the structural conditions of transformations permit us to bring the concept of proper analysis of Peters and Ritchie (1973) much closer to the working practical conception of syntacticians, which is essentially that a proper analysis of a structure \(\phi\) is an assignment of substructures of \(\phi\) to the basic predicates of the structural condition of a transformation. If we dropped these

\(^3\) The use of disjunction in transformations has been explicitly rejected by McCawley (1972) and Lakoff (1971). See also Bresnan (to appear). Certain very restricted cases of disjunction are expressible in the notation provided in section 4.1.

\(^4\) It would be a mistake to suppose that to admit predicates over sequences of factors is to undermine the structure-dependence of transformations by making indirect reference to functional or relational notions possible. Note that by the definition of is a, the condition \(V^n_{1,1} \& NP^n_{2,2} \& VP^n_{3,2}\) will apply to structures \([VP][VP[VPNP]], [VP][VP[VPNP]]\) as well as \([VP][VPNP]\); so there is no unique grammatical relation picked out by this condition, which in the informal notation would be written "\([VPNP]\)."
conditions on the form of transformations, and regarded structural conditions merely as Boolean combinations of predicates, we would lose this desirable result, for a structural condition could be true of a factorization without any of its atomic predicates being true of it.

For example, let \( C = \text{NP}_1^2 \lor \neg \text{VP}_2^2 \) and consider the factorization (11).

\[
(11) \quad ([NL[VP[vgo]_v, [PP[paway]PP]VP]_S)
\]

\( C \) is true of (11), but neither \( \text{NP}_1^2 \) nor \( \text{VP}_2^2 \) is true of (11). Similarly, \( C' = \text{NP}_1^2 \lor \neg \text{NP}_2^2 \) is true of (11) without either of its atomic predicates being true of (11).

Since a structural condition \( C \) is a conjunction of basic predicates \( P \), every \( P \) in \( C \) holds of a factorization if \( C \) holds of it. We can therefore define a notion that will be very useful later, the value of a predicate:

**Definition**

If \( \pi = (\phi_1, \ldots, \phi_n) \) is a proper analysis of \( \phi \) for \( T = (C, M) \), then for any basic predicate \( P \) in \( C \), \( |P|_\pi \), the value of \( P \) under \( \pi \), is defined by

\[
|P|_\pi = \begin{cases} 
I(\phi_1 \cdots \phi_i) & \text{if } \phi_1 \cdots \phi_i \text{ has an interior} \\
e & \text{otherwise.}
\end{cases}
\]

"|P|_\pi" is well-defined since \( I \) is well-defined. \( I(\psi) \), the interior of \( \psi \), is the longest well-formed terminal labelled bracketing in \( \psi \) that contains the terminal string of \( \psi \); see Peters and Ritchie (1973, definition 2.5)."

To illustrate, the structural condition of transformation (2) would be written \( \text{NP}_2^2 \& \text{Aux}_2^3 & \text{V}_4^4 & (P)_2^2 & \text{NP}_6^6 \). Converting the proper analysis of (3) into labelled bracketings, we have \( |\text{NP}_6^6|_\pi = |\text{NP}[\text{Nchildren('s')}]_N|_{NP} \) and \( |\text{NP}_6^6|_\pi = |\text{NP}[\text{S}[\text{NP}[\text{Nchildren('s')}]_N]\text{VP}[\text{vsmoking}]_V]\text{VP}S|_{NP}. \)

### 2.3. Transformational Mappings

We did not admit "identity" predicates into structural conditions, but will follow Peters and Ritchie (1973, note 6) in introducing them as universal conditions on transformational mappings. This will reserve identity for recoverability, in accordance with Chomsky (1965).

An \( n \)-term transformational mapping is defined by a set of elementary operations \( M = \{o_1, \ldots, o_n\} \), in which each \( o_m \) is a pair \([T_D, (h, i)]\) or a triple \([T_S, (h, i), (j, k)]\), \([T_L, (h, i), (j, k)]\), \([T_R, (h, i), (j, k)]\). \( T_D, T_S, T_L, \) and \( T_R \) are respectively the deletion, substitution, left-adjunction, and right-adjunction elements. The second and third components of the triples denote the terms upon which \( o_m \) operates. The operations above respectively amount to "delete \( h \) through \( i'\), "substitute \( j \) through \( k \) for \( h \) through \( i'\), "left (right) adjoin \( j \) through \( k \) to \( h \) through \( i'\)." The operations \( o_m \) must not "overlap" in their second components, if the mapping is to be defined. See Peters and Ritchie (1973) for precise definitions. (Their definitions
may require modification for empirical reasons, but the details are not relevant here.)

Now we drop the deletion and substitution operations above, and add operations of the form \([\mathcal{T}_D, (h, i), (t, u)]\) and \([\mathcal{T}_S, (h, i), (j, k), (t, u)]\). These are defined in the same way as \([\mathcal{T}_D, (h, i)]\) and \([\mathcal{T}_S, (h, i), (j, k)]\), but will be subject to the conditions “delete \(h\) through \(i\) under identity to \(t\) through \(u\)” and “substitute \(j\) through \(k\) for \(h\) through \(i\) only if \(h\) through \(i\) is identical to \(t\) through \(u\)”, respectively. These conditions are given below.

For natural numbers \(h, i, j, k, n\), the \(n\)-ary identity predicate “\(h-i \equiv j-k\)” is defined by Peters and Ritchie to hold of an \(n\)-term factorization \((\phi_1, \ldots, \phi_n)\) if (i) \(1 \leq h \leq i \leq n\) and \(1 \leq j \leq k \leq n\), and (ii) \(C(\phi_h \cdot \phi_i) = C(\phi_j \cdot \phi_k)\), where “\(C(\psi)\)” denotes the contents of \(\psi\). (The contents of \(\psi\) is the concatenation of the interiors of the factors of the coarsest factorization of \(\psi\) of which every factor has an interior; see Peters and Ritchie (1973, definition 2.7).) For natural numbers \(h, i, n\) and terminal string \(x\), the \(n\)-ary predicate “\(h-i \equiv x\)” is defined by Peters and Ritchie to hold of an \(n\)-term factorization \((\phi_1, \ldots, \phi_n)\) if (i) \(1 \leq h \leq i \leq n\) and (ii) \(d(\phi_h \cdot \phi_i) = x\), where \(d(\psi)\) is the debracketization of \(\psi\); see Peters and Ritchie (1973, definition 2.2).

Now let \(T = (\mathcal{C}, \mathcal{M})\) be any \(n\)-term transformation and \(\pi = (\phi_1, \ldots, \phi_n)\) be a factorization. \(\mathcal{M}\) is defined on \(\pi\) only if the following condition holds: for all \(h, i, j, k, t, u\), if \([\mathcal{T}_D, (h, i), (t, u)] \in \mathcal{M}\) or \([\mathcal{T}_S, (h, i), (j, k), (t, u)] \in \mathcal{M}\), then \((\mathcal{C} \& h-i \equiv t-u)\) is true of \(\pi\) or \(h = t, i = u\), and there is an \(x\) such that \((\mathcal{C} \& t-u \equiv x)\) is true of \(\pi\). This condition is necessary but not sufficient for \(\mathcal{M}\) to be defined on \(\pi\). \(T\) must also meet the condition on recoverability of deletions, which guarantees that if a copy of \(C(\phi_h \cdot \phi_i)\) is not preserved in the output of \(T\), \(d(\phi_h \cdot \phi_i)\) must be a designated element.\(^5\)

To illustrate, the transformational mapping of (2) is defined by \(\mathcal{M} = \{[\mathcal{T}_S, (2, 2), (6, 6), (2, 2)], [\mathcal{T}_D, (6, 6), (6, 6)]\}\). \(\mathcal{M}\) is defined on the factorization \(\pi_2\) in (3) (as well as \(\pi_1\)) because the structural condition holds of \(\pi_2\), and so do \(6-6 \equiv 6-6\) and \(2-2 \equiv \Delta\), where “\(\Delta\)” is a designated terminal symbol; \(\mathcal{M}\) meets the recoverability condition because a copy of the deleted sixth term is preserved in the output.

Finally, we require that each \(\mathcal{M}\) have at most one deletion operation \([\mathcal{T}_D, (h, i), (j, k)]\). This restriction is linguistically natural: in fact, it follows from more general restrictions on permissible combinations of operations. Although I will not consider it further here, it appears that “natural”—that is, empirically justified—transformational mappings each perform no more than one basic set of operations, such as deletion, movement (substitution and deletion, or adjunction and deletion), or possibly copying (adjunction or substitution). In none of the natural transformational

---

\(^5\) Peters and Ritchie (1973) prove that this recoverability condition is not sufficient to guarantee that all transformational languages are recursive. However, the observations in Peters (1973) suggest that transformational grammars for natural languages have further properties that yield recursiveness.
mappings known to me is more than a single deletion elementary necessary (though there may be more than one elementary).\(^6\)

2.4. Target Predicates

A transformation may have the effect of "moving" or deleting a single substructure of the structure to which it applies, as in example (3), where the substructure \([\text{NP}[s[\text{NP}[\text{N} \text{children}('s)]]\text{NP}[\text{VP}[\text{vsmoking}]_{v}]\text{VP}]_{s}]_{\text{NP}}\) is displaced by the transformation (2). By the form of (2) itself, we can tell that it is an NP that will be removed when (2) applies to any structure, for the predicate "NP" must hold of term 6, and the mapping specifies that term 6 is deleted. We will call this NP-predicate in (2) the target predicate.

In general, given any \(n\)-term transformation \(T = (\mathcal{C}, \mathcal{M})\) such that \(\mathcal{C} = \mathcal{P}_1 \& \cdots \& \mathcal{P}_m\), where each \(\mathcal{P}_r\), \(1 \leq r \leq m\), is a basic predicate, then if there are \(s, h, i, j, k\), with \(1 \leq s \leq m\) such that \(\mathcal{P}_s = \mathcal{P}_{h, k}\) or \(\mathcal{P}_{h, k}\), and \([T_D, (h, i), (j, k)] \in \mathcal{M}\), we call \(\mathcal{P}_s\) a target predicate in \(\mathcal{C}\).

A target predicate in \(\mathcal{C}\) is one that describes either something that \(\mathcal{M}\) deletes or something that what \(\mathcal{M}\) deletes is identical to. By our conditions on \(\mathcal{C}\) and \(\mathcal{M}\), each transformation will have at most two target predicates (and possibly none, if it involves no deletion). A context predicate in \(\mathcal{C}\) is any predicate \(\mathcal{P}_{t, u}\) other than a target predicate, whose term indices \((t, u)\) are not included within those of a target predicate in \(\mathcal{C}\). Context predicates describe things "outside of" the target predicates.

3. The Functioning of Transformations

The preceding section has provided us with the concepts necessary to formulate the Relativized A-over-A Principle. For convenience, I will continue to use the familiar informal representations of rules and structures as illustrations in this section.

3.1. The A-over-A "Paradox"

Chomsky (1973, 235) remarks of the A-over-A Principle that it

does not establish an absolute prohibition against transformations that extract a phrase of type A from a more inclusive phrase of type A. Rather, it states that if a transformational rule is nonspecific with respect to the configuration defined, it will be interpreted in such a way as to satisfy the condition.

Thus, applications of Question Movement in sentences like (12) and (13) are not prevented, even though the interrogative phrases are of type NP and are contained within larger phrases of type NP:

(12) Who would you approve of my seeing?

\(^6\) Note that whether or not the Passive transformation performs simultaneous movement of the subject and the object, only one deletion elementary is necessary. Similar restrictions on combinations of operations are proposed by McCawley (1970).
Which books are you uncertain about giving to John?

To see why this is so, compare the structure (14) to the rule (15). (Question Movement is formulated more accurately in section 4.2.)

\[(14)\]
\[
\begin{array}{c}
S \\
\downarrow \\
NP \quad \text{Aux} \\
\downarrow \\
you \quad \text{would} \\
\downarrow \\
V \\
\downarrow \\
approve \quad \text{PP} \\
\downarrow \\
S \\
\end{array}
\]

\[(15)\] \[Q - X - [NP \text{wh} + Y] - Z \]

Rule (15) can extract the NP who from the more inclusive NP my seeing who because the latter does not satisfy the structural condition of (15), which specifies that the NP that undergoes movement must begin with the interrogative morpheme wh.

From this it is clear that the A-over-A Principle is a condition on proper analyses: it requires that if a transformation \(T\) applies under proper analysis \(\pi\) to a substructure \(|A|_\pi\) of \(\phi\), then for all proper analyses \(\pi_1, \ldots, \pi_n\) of \(\phi\) with respect to \(T\), \(|A|_\pi\) must be maximal in \(|A|_{\pi_i}\) if \(i \leq n\). \(|A|_\pi\) is maximal in \(|A|_{\pi_i}\) if for each \(i = 1, \ldots, n\), \(|A|_{\pi_i} > |A|_\pi\) only if \(|A|_{\pi_i} = |A|_\pi\), where "\(\alpha > \beta\)" means \(\alpha\) dominates \(\beta\); a well-formed labelled bracketing dominates all of its well-formed substrings.

It would seem, then, that a maximal proper analysis of \(\phi\) for \(T = (\mathcal{C}, \mathcal{M})\) would be one that assigned maximal values to all the predicates in \(\mathcal{C}\). However, a maximal proper analysis in this sense need not exist. Consider the structural condition in (16) and the structure in (17).

\[(16)\] \[A - B - C \]

\[(17)\]
In (17), \( |A|_{\pi_1} \) is maximal if and only if \( |C|_{\pi_1} \) is not maximal.

3.2. Maximal Proper Analyses

The A-over-A “paradox” shows that we cannot define a maximal proper analysis of \( \phi \) for \( T = (\mathcal{C}, \mathcal{M}) \) as one that assigns maximal values to all the predicates in \( \mathcal{C} \). But we can define it as one that assigns maximal values to all the target predicates in \( \mathcal{C} \).

If there is only one target predicate \( F \) in \( \mathcal{C} \), then \( F \) has a value in \( \phi \) only if \( F \) has a maximal value in \( \phi \). Otherwise, for every proper analysis \( \pi \) of \( \phi \) for \( T \), there would exist a \( \pi' \) of \( \phi \) for \( T \) such that \( |F|_{\pi'} > |F|_{\pi} \) and \( |F|_{\pi'} \neq |F|_{\pi} \), and \( \phi \) would have an infinite branch.

If there are two target predicates \( F, G \) in \( \mathcal{C} \) that have values in \( \phi \) under some \( \pi \), then \( |F|_{\pi} \) and \( |G|_{\pi} \) are structurally identical (see sections 2.3, 2.4). From this fact we can show that, if \( F \) and \( G \) have values in \( \phi \) for some \( \pi \), then there is some \( \pi' \) under which \( F \) and \( G \) have maximal values in \( \phi \). By the preceding argument, there must be a \( \pi_1 \) such that \( |F|_{\pi_1} \) is maximal. Now consider \( |G|_{\pi_1} \): if it is maximal, we have nothing left to show, so assume that \( |G|_{\pi_1} \) is not maximal. Then there is a \( \pi_2 \) such that \( |G|_{\pi_2} \neq |G|_{\pi_1} \) and \( |G|_{\pi_2} \) is maximal. Now since \( |F|_{\pi_1} \) is maximal, either \( |F|_{\pi_1} > |F|_{\pi_2} \) or one of \( |F|_{\pi_2} \), \( |F|_{\pi_2} \) precedes the other. The former case leads to a contradiction, because \( |F|_{\pi_2} \) dominates something (\(|F|_{\pi_2} \)) that is structurally identical to \( |G|_{\pi_2} \), which properly dominates something (\(|G|_{\pi_2} \)) that is structurally identical to \( |F|_{\pi_1} \): in other words, \( |F|_{\pi_2} \) would be structurally identical to a proper substructure of itself, so \( \phi \) would have an infinite branch. We have to conclude, therefore, that one of \( |F|_{\pi_2} \), \( |F|_{\pi_2} \) precedes the other. Now if \( |F|_{\pi_2} \) is not already maximal, we can pick a \( \pi_3 \) such that \( |F|_{\pi_3} \neq |F|_{\pi_2} \) and \( |F|_{\pi_3} \) is maximal. Notice that at this point we can repeat the entire argument that we used for \( |F|_{\pi_1} \) on \( |F|_{\pi_2} \), arriving at the conclusion that there is a \( \pi_4 \) such that \( |G|_{\pi_4} \) is maximal and \( |F|_{\pi_4} \) cannot dominate or be dominated by either \( |F|_{\pi_3} \) or \( |F|_{\pi_1} \). (The situation is depicted in the diagram below.) If we repeat this argument \( k \) times, we have to conclude that \( \phi \) has at least \( k \) distinct branches, each going from the root of a maximal \( |F|_{\pi_{2k+1}} \) to the root of \( \phi \). Thus if
we maintained the assumption that there is no \( \pi' \) under which both \( F \) and \( G \) have maximal values in \( \phi \), we would be forced to conclude that \( \phi \) has infinitely many branches.

\[
\begin{array}{c}
F_4 \\
F_3 \\
F_2 \\
F_1 \\
G_1 \\
G_2 \\
G_3 \\
G_4
\end{array}
\]

Thus we have established the following theorem:

**Theorem**

If there is a proper analysis \( \pi \) of \( \phi \) for \( T = (C, \mathcal{M}) \), then there is a proper analysis \( \pi' \) of \( \phi \) for \( T \) under which all target predicates in \( C \) have maximal values.

This justifies the following definition:

**Definition**

\( \pi \) is a maximal proper analysis of \( \phi \) for \( T = (C, \mathcal{M}) \) if and only if \( \pi \) assigns maximal values to all target predicates in \( C \).

We can now reformulate Chomsky's A-over-A Principle as follows:

**A-over-A Principle**

No transformation \( T \) can apply to a structure \( \phi \) under a proper analysis \( \pi \) unless \( \pi \) is a maximal proper analysis of \( \phi \) for \( T \).

By narrowing down what it means for a transformation to "apply" to a phrase of type A, we have obtained a well-defined A-over-A Principle. The reader can easily verify that the paradoxical situation in (17) can no longer arise: if both \( A \) and \( C \) are target predicates in (16), then \( |A|_{\pi_2} \) must be structurally identical to \( |C|_{\pi_2} \), in which case it would be impossible for \( |A|_{\pi_1} \) also to be structurally identical to \( |C|_{\pi_1} \). Therefore, if \( \pi_2 \) is a proper analysis of (17), \( \pi_1 \) cannot also be a proper analysis of (17), so \( \pi_2 \) assigns maximal values to the target predicates \( A \) and \( C \) in (16), thereby satisfying the A-over-A Principle. (The symmetrical argument holds for \( \pi_1 \), of course.) Note that if \( T \) has no deletion elementary, every proper analysis for \( T \) will be maximal, because the condition that all target predicates have maximal values will be vacuously satisfied when there are no target predicates.7

---

7 Therefore, we do not expect copying transformations to apply only to "maximal" phrases. The reader may wish to examine the rules of Right Node Raising and English Topicalization from this point of view: if,
This way of resolving the A-over-A “paradox” was in fact chosen for empirical reasons, to which I now turn. There is in English a class of complex verbs that permit “double passives”:

(18) Someone took advantage of my absence.
(19) My absence was taken advantage of.
(20) Advantage was taken of my absence.

As proposed by Bresnan (1972) and Berman (1974), (19) and (20) are derived from a structure something like (21), which is shown properly analyzed with respect to the Passive transformation (rule (2) in section 1).

\[(21)\]
\[
\begin{array}{c}
S \\
NP \quad \text{Aux} \\
\Delta \\
\qquad \text{was} \\
\qquad \text{V} \\
\qquad \text{P} \\
\qquad \text{NP} \\
\end{array}
\]

\[
\begin{array}{c}
\pi_1: \\
1 \\
2 \\
3 \\
4 \\
5 \\
6 \\
7 \\
\end{array}
\quad \begin{array}{c}
\pi_2: \\
1 \\
2 \\
3 \\
4 \\
5 \\
6 \\
7 \\
\end{array}
\]

Transformation under both \(\pi_1\) and \(\pi_2\) is possible, despite the fact that \(|V|_{\pi_1}\) is not maximal. But this is permitted by the A-over-A Principle, because \(V\) is not a target predicate in the Passive transformation.

Evidence that \textit{take advantage} is in fact a verb (albeit a complex verb) comes from the rule of Verb Gapping, which deletes a verb between two constituents that are S-initial and S-final: \textit{John hates Mary and Mary hates John} \rightarrow \textit{John hates Mary, and Mary, John}. We find that \textit{take advantage} undergoes Verb Gapping:

(22) John took advantage of Mary, and Mary took advantage of John.

(23) John took advantage of Mary, and Mary, of John.

as is commonly supposed, both rules involve adjunction of a phrase to a structure that properly contains that phrase, then by the nonoverlap condition on transformational mappings, they must be broken down into separate subrules of copying and deletion. See Bresnan (1974) and Postal (1974, 125ff.) for some discussion of Right Node Raising.
The fact that *take is not maximal in *take advantage is consistent with its failure to undergo Verb Gapping by itself, although there are probably other reasons for the ungrammaticality of (24) as well (see Stillings (1975) and the references cited there):

(24) *John took advantage of Mary, and Mary, advantage of John.

Finally, we note that a sequence V NP that is not dominated by V cannot undergo Verb Gapping between two S-initial and S-final constituents:

(25) They regard John as a fool, and we regard John as a pushover.

(26) *They regard John as a fool, and we, as a pushover.

In (25), regard John as a fool has the structure V NP Pred, where all three constituents are immediately dominated by VP. (On the ungrammaticality of *They regard John as a fool, and we, Mary as a pushover, see Stillings (1975) and the references cited there.)

We see, then, that context predicates appear not to be maximized by the A-over-A Principle, a fact that empirically supports the reformulation of the A-over-A Principle given here.

3.3. The Relativized A-over-A Principle

Although the A-over-A Principle formulated in 3.2 is already "relativized" to structural conditions, it differs slightly from the formulation that I will adopt in the remainder of this study under the name The Relativized A-over-A Principle. The latter will require that a transformation \( T = (C, M) \) apply under a proper analysis \( \pi \) that is maximal relative to all proper analyses that agree with \( \pi \) on all context predicates in \( C \):

Definition

\( \pi \) is an \( r \)-maximal proper analysis of \( \phi \) for \( T = (C, M) \) iff: for every proper analysis \( \pi' \) of \( \phi \) for \( T \) such that for every context predicate \( N \) in \( C \), \( |N|_{\pi} = |N|_{\pi'} \), \( |F|_{\pi'} > |F|_{\pi} \Rightarrow |F|_{\pi'} = |F|_{\pi} \) for each target predicate \( F \) in \( C \).

If there are no target predicates in \( C \), any proper analysis of \( \phi \) for \( T \) will be \( r \)-maximal. Therefore, we state the Relativized A-over-A Principle as follows:

The Relativized A-over-A Principle

No transformation \( T \) can apply to a structure \( \phi \) under a proper analysis \( \pi \) unless \( \pi \) is an \( r \)-maximal proper analysis of \( \phi \) for \( T \).

The Relativized A-over-A Principle has the same effects as the A-over-A Principle in the cases considered so far. The reason for fixing the values of the context predicates before "maximizing" the target predicates appears when we consider the behavior of other transformations, such as VP Deletion.

VP Deletion can be stated approximately as in (27).


\[
\begin{array}{cccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
1 & 2 & 3 & 4 & 5 & 6 & 7 & \phi & 9
\end{array}
\]
Unlike Verb Gapping, VP Deletion leaves behind a finite or nonfinite auxiliary element: compare *John didn’t go, but Bill didn’t φ either with John didn’t go, but Bill φ either, and John didn’t go, because he didn’t want to φ with *John didn’t go, because he didn’t want φ. Furthermore, VPs not preceded by an auxiliary element cannot undergo VP Deletion (though there are other ellipsis rules possible in some cases):

(28) First people began pouring out of the building, and then smoke began pouring out of the building. ⇒ *First people began pouring out of the building, and then smoke began φ.

(29) First people began to pour out of the building, and then smoke began to pour out of the building. ⇒ First people began to pour out of the building, and then smoke began to φ.

(Thus, however auxiliaries are analyzed, some appropriate context predicate to the left of the target VP is needed in (27).)

Recast in the notation of section 2, (27) would be (30):

\[(S^9_{1-4} & \text{Aux}^9_{2-2} & \text{VP}^9_{3-3} & \text{Conj}^9_{5-5} & S^9_{5-9} & \text{Aux}^9_{7-7} & \text{VP}^9_{8-8},
\{[T_D, (8, 8), (3, 3)]\})\]

I give this formulation to make clear the fact that variables are not predicates in a rule and hence do not have maximizable values under proper analyses. The target predicates in (30) are VP^9_{3-3} and VP^9_{8-8}; all the other predicates are context predicates.

Now consider the possible proper analyses of (31) with respect to (27):

\[(S^{31})\]
By the A-over-A Principle (3.2), only the highest \( \text{VP} \) can be deleted, but by the Relativized A-over-A Principle (3.3), each \( \text{VP} \) can be deleted, because it is only for each assignment to the context predicate "Aux" that the target predicate "VP" must have maximal value. Thus, the Relativized A-over-A Principle correctly permits derivation of (32)–(34), while the A-over-A Principle permits only (32):

\[
\begin{align*}
(32) & \quad \text{Frankie will seem to want to leave St. Louis, but Johnny won't.} \\
(33) & \quad \text{Frankie will seem to want to leave St. Louis, but Johnny won't seem to.} \\
(34) & \quad \text{Frankie will seem to want to leave St. Louis, but Johnny won't seem to want to.}
\end{align*}
\]

I am aware that VP Deletion is considered by some linguists not to be a transformation at all, but rather a rule of anaphora. To them, this evidence in favor of the Relativized A-over-A Principle as against the A-over-A Principle may not seem compelling. Further evidence comes from the case of Pied Piping discussed in 4.2, for which an understanding of this possibly hypothetical example of VP Deletion will be useful.\(^8\)

Other cases in the literature given as counterexamples to the A-over-A Principle are not counterexamples to the Relativized A-over-A Principle (e.g. the one given in Akmajian (1975, 124)). Rather than survey all relevant transformations to demonstrate this, I will turn now to a few of what I consider to be the most interesting applications of the principle, the cross-categorial transformations.

4. Applications

In this section the Relativized A-over-A Principle is applied to cross-categorial transformations. Section 4.1 is a brief introduction to the "\( \text{X} \)" notation used in 4.2.

4.1. Natural Classes of Syntactic Categories

An inadequacy of phrase-structure grammars was pointed out by John Lyons (1968, 330–332). Consider the phrase structure rules given in (35).

\[
\begin{align*}
(35) & \quad S \rightarrow \text{NP Aux VP} \\
& \quad \text{VP} \rightarrow \text{V NP} \\
& \quad \text{VP} \rightarrow \text{V} \\
& \quad \text{NP} \rightarrow \text{Det N}
\end{align*}
\]

\(^8\) I do not really consider this example hypothetical, because an adequate interpretive account of VP Deletion requires "ramified-delta" structures, which must themselves undergo deletion. (See Wasow (to appear).) It is easy to recapture the Relativized A-over-A Principle in a system that permits only delta deletion, by using the concept of structural nondistinctness instead of structural identity, taking \( a \) to be structurally nondistinct from \( \beta \) if \( a \) differs from \( \beta \) only in having some structures dominating \( \Delta s \) where \( \beta \) has structures dominating lexical material, and using the presence of \( \Delta s \) in surface structure as a filtering device. But, lacking crucial evidence for such a reformulation, I leave it unformulated.
Lyons observes that

the rules given above fail to formalize the fact that there is an essential, language-independent, relationship between N and NP and between V and VP. As far as the formalization of phrase-structure grammars is concerned, it is a matter of "accidental" coincidence that linguists will include in their grammars of different languages rules which always expand NP into a string of symbols containing N and rules which always expand VP into a string of symbols containing V. In other words, phrase-structure grammars fail to formalize the fact that NP and VP are not merely mnemonically-convenient symbols, but stand for sentence-constituents which are necessarily nominal and verbal, respectively, because they have N and V as an obligatory major constituent. . . .

What is required, and what was assumed in traditional grammar, is some way of relating sentence-constituents of the form XP to X (where X is any major category: N, V, etc.). It would not only be perverse, but it should be theoretically impossible for any linguist to propose, for example, rules of the following form. . . . [emphasis added/JWB]

(36)  S → VP Aux NP
      NP → V VP
      NP → V
      VP → Det N

Chomsky (1970) provides a means for solving this problem. Categories are decomposed into features and types. The "major class features" of the major categories of English are shown in (37).9

(37)  V  [+] verbs
       N  [-V] nouns
       A  [+V] adjectives (and adverbs)
       P  [-V] prepositions (and adverbial particles)

The categories in (37) are of type o. Categories of type i occur on the left hand side of the phrase structure rules in (38). (The illustrative phrase structure rules to follow are based on a modification of Bresnan (1973); I hope to give a fuller treatment in a more complete study.)

(38)  V → V (NP)  We are nearing the meadow.
      P → P (NP)  Near the meadow, we built a house.
      A → A (PP)  The house was much nearer to the meadow after the tornado than it was before.
      N → N (PP)  Nearness to the meadow is the great virtue of our house.

9 This system was given in Chomsky's public lectures at the 1974 Summer Institute of the Linguistic
The symbols "NP" and "PP" in (38) are equivalent to categories $\overline{N}$ and $\overline{P}$ of type 2, whose expansion rules are given in (39).

\[(39) \quad \overline{V} \rightarrow (\text{Perf}) (\text{Prog}) (\text{Pass}) \overline{V} \\
\overline{P} \rightarrow (\overline{Q}) \overline{P} \\
\overline{A} \rightarrow (\overline{Q}) \overline{A} \\
\overline{N} \rightarrow (\overline{Q}) \overline{N} \]

"$\overline{Q}$" is the category of measure-phrases, numerals, and quantifiers studied in Bresnan (1973). (The syntactic features of $Q$-categories appear to depend upon the features of the categories they modify; I leave them unspecified here.) The expansion of $\overline{Q}$ is given by $\overline{Q} \rightarrow (\overline{Q}) \overline{Q}, \overline{Q} \rightarrow (D) Q$, where $D$ is the category of the, this, each, some; too, so, as, -er; who, how, what, etc.; and $Q$ is the category of many, much, few, 6, one, etc. I assume that there are also categories of type 3 ($\overline{V} \rightarrow \text{Aux} \overline{V}$), type 4 ($S \rightarrow N V$), and type 5 ($\overline{S} \rightarrow (\text{COMP}) S$), but these details will not be of concern here.

Every category of type $i + 1$ has as an obligatory major constituent a category of type $i$, which is its head. If $A$ is the head of $B$, $A$ and $B$ must have the same syntactic features. (This represents Lyons's generalization.) The syntactic features and types can be used to designate natural classes of categories in the following way.

Every category can be represented as an ordered pair $\langle i, M \rangle$, where $i$ is the type of the category and $M$ is the feature-matrix of the category. We will call all categories (other than $\overline{S}$) that are the heads of no categories, type $\phi$ categories; examples are $D$ and COMP, which also can be represented as ordered pairs $\langle \phi, M \rangle$ for suitable matrices $M$. (In addition to major class features, there are other syntactic features including "count" features.)

Given any type $i$ (including $i = \phi$) and any submatrix of syntactic features $M$ (including the null matrix), we form a class predicate $^iX$, which characterizes the set of categories $\langle i, N \rangle$ for which $M$ is a submatrix of $N$. We will say that a category $B$ satisfies a class predicate $^iX$ if $B = \langle i, N \rangle$ and $M$ is a submatrix of $N$. Examples of sets of categories that satisfy several class predicates are in (40).

\[(40) \quad ^0X = \{V\} \\
[+V] \quad ^0X = \{V, A\} \\
[-N] \quad ^0X = \{V, A, N, P\} \]

---

Society of America at the University of Massachusetts, Amherst. Note that not all combinations of features need occur in every language and that there may be additional features.
FORM AND FUNCTIONING OF TRANSFORMATIONS

\[ 1X = \{ \nabla, \bar{P} \} \]
\[ [-N] \]

\[ 2X = \{ \bar{N}, \bar{P} \} \]
\[ [-V] \]

For those class predicates that characterize singleton sets of categories, we will abbreviate \( X \) with the name of the category, thus: the first class predicate in (40) will be written "\( OV \)". We will also use the obvious abbreviations "\( S \)" and "\( \bar{S} \)" for class predicates.

It is from these class predicates that we now form the structural conditions of transformations. For any class predicate \( \chi \) and natural numbers \( n, j, k \), we form an \( n \)-ary predicate \( \chi_{j-k}^n \). \( \chi_{j-k}^n \) holds of a factorization \( (\phi_1, \ldots, \phi_n) \) iff \( 1 \leq j \leq k \leq n \) and \( \phi_j \cdot \phi_k \) is an \( \chi \). \( \phi_j \cdot \phi_k \) is a \( \chi \) if there is a category \( B \) such that \( \phi_j \cdots \phi_k \) is a \( B \) and \( B \) satisfies \( \chi \). As before, we add \( n \)-ary predicates \( (\chi)^j_{j-k} \) that hold of a factorization \( (\phi_1, \ldots, \phi_n) \) if either \( \chi_{j-k}^n \) holds of it or \( \phi_j \cdot \phi_k = e \). Structural conditions are now defined exactly as before (2.1), as conjunctions of \( n \)-ary predicates.

It is important to distinguish constituent structure trees (terminal labelled bracketings) from structural conditions: the categories in the former are fully specified for major syntactic features; the predicates in the latter may be only partially specified. This is analogous to phonological strings and rules. Just as phonological rules may apply to natural classes of phonological segments, so transformations may apply to natural classes of syntactic categories, as we will see in the next subsection.

4.2. Cross-Categorial Transformations

4.2.1. Question Movement (QM). Question Movement applies to Noun-, Adjective-, Adverb-, and Q-Phrases that begin with an interrogative morpheme:

\[
(41) \quad \text{What book did you read?} \\
\text{How long is it?} \\
\text{How quickly did you read it?} \\
\text{How much did it cost?}
\]

The phrases moved in (41) are all type 2 categories, so QM can be formulated as in (42).

\[
(42) \quad [sQ - W_1 - [\bar{\chi}wh - W_2] - W_3] \\
\begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 \\
\phi \end{array} \rightarrow \\
\begin{array}{cccccc}
3 & 4 & 2 \\
\phi \end{array} 5
\]
The term "Q" in (42) represents the interrogative complementizer of Bresnan (1970, 1974). In the notation of sections 4.1 and 2, (42) is written as (43).

\[(43) \quad (\text{COMP}_{1-1}^5 \& 2X_{3-4}^5 \& D_{3-3}^5 \& S_{1-5}^5, + \text{wh})
\]

\[
\{[T_S, (1, 1), (3, 4), (1, 1)], [T_D, (3, 4), (3, 4)]\}
\]

In accordance with the conditions of 2.3, the sequence 3–4 is deleted under identity to itself; this is possible because the sequence is preserved in the output of any mapping defined by the set of operations in (43), since 3–4 is substituted for 1–1. Also, 1–1 is substituted for under identity to itself; this is possible only if the COMP node dominates a designated element (which I assume to be \(\Delta\)). The target predicate in (43) is 2X_{3-4}^5; the context predicates are the first and last predicates in the conjunction. (D_{3-3}^5 is + \text{wh} neither a context predicate nor a target predicate, because its term indices 3–3 are properly included within those of a target predicate; see 2.4.)

Now the Relativized A-over-A Principle predicts that QM can apply only when a proper analysis assigns a maximal value to \(\bar{X}\) for fixed values of \(Q\) and \(S\). Thus, it correctly predicts the following data.

(44) a. How difficult a problem do you want to solve?
   b. *How difficult do you want to solve _____ a problem?
   c. How difficult is that problem?

(45) a. How many feet tall does the girl stand?
   b. *How many feet does the girl stand _____ tall?
   c. *How many does the girl stand _____ feet tall?

(46) a. How many feet is a mile?
   b. *How many is a mile _____ feet?

(47) a. How much more money is it?
   b. *How much more is it _____ money?
   c. *How much is it _____ more money?

(48) a. How much more skillfully can you phrase this?
   b. *How much more can you phrase this _____ skillfully?
   c. *How much can you phrase this _____ more skillfully?

To show how (44)–(48) are explained by the Relativized A-over-A Principle, I will display the sources for these sentences. The structures are simplified except in essential points; the analysis derives from Bresnan (1973, 1975). The only important feature of these structures is that the modifiers are left-nested in categories of the same type, which was justified in detail in Bresnan (1973) on completely independent grounds.

The source for (44a) is (49):
All of the circled phrases in (49) satisfy rule (42), but only the circled $\bar{N}$ is maximal; hence (44a) can be derived, but not (44b). (44c) shows that when an $\bar{A}$ is maximal, it can undergo QM.

The source for (45a) is (50):

(50)
Here only the $\bar{A}$ is maximal, so only (45a) in (45) is derived. The source for (46a) is (51):

\[
(51)
\]

Here $\bar{N}$ but not $\bar{Q}$ is maximal, so (46a) but not (46b) is derived. The source for (47a) is (52):

\[
(52)
\]

Only $\bar{N}$ is maximal, so only (47a) in (47) is derived.
The source for (48a) is (53):

\[
\begin{array}{c}
\text{S} \\
\text{COMP} \\
+ \text{wh} \\
\Delta \\
\text{S} \\
\text{Aux} \\
\text{you} \\
\text{can} \\
\text{V} \\
\text{NP} \\
\text{phrase} \\
\text{this} \\
\text{VP} \\
\text{A} \\
\text{\bar{Q}} \\
\text{D} \\
\text{Q} \\
\text{\bar{Q}} \\
\text{\hat{Q}} \\
\text{\hat{A}} \\
\text{skillfully}
\end{array}
\]

Only \( \bar{A} \) is maximal, so only (48a) in (48) is derived.

4.2.2. **Comparative Deletion (CD).** Comparative Deletion applies to type 2 categories that begin with a measure phrase:

- **“Q” Phrases**
  
  \[
  \text{(54) It costs more than it weighs.} \\
  \text{[It costs [\(\tilde{\text{A}}\)-er much] than it weighs [\(\tilde{\text{A}}\Delta\) much]]}
  \]

- **Adjective Phrases**
  
  \[
  \text{(55) It looks more costly than it is.} \\
  \text{[It looks [\(\tilde{\text{A}}\)-er much] costly] than it is [\(\tilde{\text{A}}\Delta\) much] costly]]}
  \]

- **Adverb Phrases**
  
  \[
  \text{(56) She drives more dynamically than he drives.} \\
  \text{[She drives [\(\tilde{\text{A}}\)-er much] dynamically] than he drives [\(\tilde{\text{A}}\Delta\) much] dynamically]]}
  \]

- **Noun Phrases**
  
  \[
  \text{(57) She has more friends than he had because of his rotten character.} \\
  \text{[She has [\(\tilde{\text{A}}\)-er many] friends] than he had [\(\tilde{\text{A}}\Delta\) many] friends] because of his rotten character]}
  \]

The same rule applies in equative constructions (e.g. *She has as much money as he has*).
I formulate CD as in (58). Note that each occurrence of $\bar{X}$ in (58) is a predicate that can be satisfied by any of several type 2 categories: $\bar{X}$ is not a variable.

\[
(58) \quad [\bar{x}[\bar{Q} - W_1] - W_2][\bar{s}W_3 - [\bar{Q} - W_4] - W_5]
\]

\[
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
& & & & & \phi & \\
\end{array}
\]

In the notation of 4.1 and 2 we have (59):

\[
(59) \quad (\hat{2}X_1 - 2 & \hat{3}Q_{1-1} & \hat{2}X_1 - 2 & \hat{3}S_{1-4} & \hat{2}Q_{1-5} & \hat{2}X_{5-6}, \\
\{[T_D, (5, 6), (1, 2)]\})
\]

The target predicates in (59) are $\hat{2}X_{1-2}$ and $\hat{2}X_{5-6}$; the context predicates are the first and fourth predicates in (59).

To see how this rule works, let us apply it to derive the comparative construction in (57):

\[
(60)
\]

10 At this point it must be mentioned that the sequence of factors 1–2 is never going to be strictly identical to 5–6. As we see in examples (54)–(57), the determiner of the compared constituent to be deleted is a designated element $\Delta$, while that of the higher compared constituent is -er (or as). Therefore, we must revise the notion of structural identity to structural nondistinctness, to permit deletion of structures dominating designated elements under “identity” to lexically saturated structures. The appropriate definition is not hard to formulate, but I prefer to reserve it for a more complete study where its detailed justification and general applications can be given. I will continue to speak of (structural) identity in this section.

Let me also forestall another possible objection of the reader, who might reasonably ask, How does CD “pick out” the correct phrase of the form $[\bar{x}\bar{Q} \ldots]$ for deletion? The answer is that a designated determiner $\Delta$ will cause a sentence to be filtered out if it is not deleted. The correct distribution of $\Delta$s is a function of seman-
The reader can check that \( \pi_1 \) is a proper analysis of (60) for (59). Note that factor 1 is a \( 2Q \); the sequence of factors 1–2 is a \( 2X \) (because \( N \) satisfies \( 2X \) ); and the sequence of factors 1–3 is a \( 2X \) (again, because \( N \) satisfies \( 2X \) ). (Recall the definition of \( is \) a in 4.1 and 2.) \( \pi_1 \) is also a maximal proper analysis of (60), for there is no other proper analysis that assigns greater values to the target predicates of (59). So CD deletes the contents of the sequence of factors 5–6, and the comparative in (57) is derived.

But there is another proper analysis of (60) that is nonmaximal: namely, \( \pi_2 \), which is shown in (61):

\[
\begin{array}{c}
1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \\
\pi_2:
\end{array}
\]

\( \pi_2 \) is nonmaximal because \( |X^*_1| \neq |X^*_2| \) and \( |X^*_5| \neq |X^*_6| \). If we applied (59) to (61) under \( \pi_2 \), we would derive (62), which is ungrammatical.

(62) *She has more friends than he had friends because of his rotten character.

But now consider (62), which I believe is grammatical:

(63) She has more friends than he had enemies because of his rotten character.

The source for (63) is (64):

---

*tic interpretation: “\( \Delta \)” corresponds to a semantic variable over “degrees”, “extents”, “amounts”, etc. Further, for semantic recoverability, the \( Q \) to be deleted must be bound by the \( Q \) that it is “identical” to. So the reader could visualize rule (58) heuristically as “[\( x \underline{Q}_1 W_1 W_2 \underline{Q}_4 W_3 \underline{Q}_4 W_5 \underline{Q}_4 W_6 \) W_7]”. For a formal semantics of comparatives based on a syntax similar to this, see Davis and Hellan (1975).
\( \pi_3 \) factors (64) in the same way that \( \pi_2 \) factors (61): the difference is that \( |^{2X_{1-2}}|_{x_9} \) and \( |^{2X_{6-8}}|_{x_9} \) are maximal, for the larger \( \tilde{\mathbf{N}} \)-phrases dominating these phrases fail the identity condition. Thus, the Relativized A-over-A Principle has the result that CD deletes the maximal constituent identical to a corresponding constituent in the "head" of the construction; it is the maximally recoverable constituent that undergoes Comparative Deletion.

For some speakers, the ungrammaticality of (62) and similar examples may be less than "plangent". There are several reasons for this, I believe. One is that it is very easy to restore simple deletions: there is no reordering over a variable in CD, as there is in QM, and the structures are more easily perceived. (On the issue of movement in CD, see Bresnan (1975).) Easy perceptibility no doubt favorably influences grammatical judgments. However, there are other factors involved as well.

For in many cases, partial repetitions in the compared constituent of the subordinate clause are perfectly grammatical. Consider first examples (65)–(67).

(65) He’s as sick of you as you are sick of him.

(66) Are there more degrees of grammaticality than there are degrees of ungrammaticality?

(67) There aren’t as many nuggets of gold in the jar as there appear to be nuggets of pyrite.
Taking (67) as representative, let us analyze the structure of the compared constituent as many nuggets of gold. I assume that of gold is a constituent. It appears that nuggets of gold is also a constituent:

(68) a. I have 20, and she has 50, nuggets of gold.
b. I have many, and she has too many, nuggets of gold.

In (68) we have applied Right Node Raising to nuggets of gold; Right Node Raising is a sufficient (but not a necessary) test for constituency (see Bresnan (1974)). Therefore, we assign the structure (69) to this phrase.

(69)

Now consider how CD applies to derive the comparative in (67):

(70)
There is no way for CD to delete $\Delta$ many nuggets in (70) because that string is not a constituent in $\Delta$ many nuggets of gold and hence cannot satisfy the target predicate $^2X^L_{\Delta}$. Thus (67) is derived by CD. How then do we account for (71)?

(71) There aren't as many nuggets of gold in the jar as there appear to be of pyrite.

In (71), nuggets has been elided—but not by CD. Rather we hypothesize a separate, optional rule of deletion at work in (71), the same rule that derives (72b):

(72) a. There are many nuggets of gold in the jar, but there don't appear to be many nuggets of pyrite. →

b. There are many nuggets of gold in the jar, but there don't appear to be many of pyrite.

(This rule could well be a “$\Delta$-deletion” rule; see Andrews (1975).)

Thus, when evaluating examples of CD that bear on the maximality principle, it is very important to consider with care the internal structure of the compared constituents. One will find that many subtle factors are at play. Consider, for example, (73).

(73) I have more pictures than she has pictures on her studio wall.

In general, the prepositional phrase on her studio wall could be an immediate constituent either of a noun phrase, pictures on her studio wall, or of a verb phrase, has pictures on her studio wall. In the latter situation, CD must delete the full NP $\Delta$ many pictures; but in the former case, it cannot delete the phrase pictures on her studio wall, which is not structurally identical to that in the head. Thus (73) can be derived, but only on one reading. By contrast, the closely related example I have more pictures than she has on her studio wall has two distinct sources, one involving the optional elision mentioned above.

The interaction of CD and the Relativized A-over-A Principle accounts for many hitherto unexplained facts. Consider the following sentence. (This type of example was brought to my attention by James Gee.)

(74) I have as many too many marbles as there are fingers on my right hand.

This sentence is ambiguous. The most natural and pragmatically preferred reading is the one that implies that I have five (or whatever the number of fingers on my right hand) too many marbles. But there is another possible interpretation that entails that I have “x many too many fingers on my right hand”. If this interpretation seems strained in (74), compare (75):

(75) Linda swam as many too many laps as Joan swam—which is why they both died of exhaustion.
FORM AND FUNCTIONING OF TRANSFORMATIONS

(75) has as a possible entailment, "Joan swam \(x\) many too many laps". (74) has a similar possibility, but it is less favored, because I would not normally wish to suggest that "there are \(x\) many too many fingers on my right hand."

We have the following two sources for the comparative construction in (74):

(76)

(77)
(Observe that $Q_{l-1}^2$, not being a target predicate, does not have to be maximal in (76) or (77).) Now observe that (78) is unambiguous:

(78) Linda swam as many too many laps as Joan swam laps (in all).

The sentence means only that Joan swam a certain number of laps, say $x$ many laps, and Linda swam $x$ many too many laps. This is predicted by our A-over-A Principle: for since laps has not undergone CD (which is obligatory), it must not have been in the maximal $\tilde{X}$-phrase that was structurally identical to one in the head. That is, the underlying structure of the compared constituent in the clause could not have been $\Delta$ many too many laps. Rather, it was $\Delta$ many laps, and CD applied exactly as in (76). Finally, notice that the optional elision rule mentioned in connection with (71) may operate on the structure underlying (78), to give (79):

(79) Linda swam as many too many laps as Joan swam (in all).

The Relativized A-over-A Principle is therefore confirmed by the behavior of CD; behind many intricacies, the important generalization holds that CD deletes the maximal phrase recoverable from the head.

Let me conclude by emphasizing the main features of this application of the Relativized A-over-A Principle, so as not to leave them embedded in the intricacies.

The basic idea is that the A-over-A Principle is relativized to the structural conditions of rules. The structural condition of CD contains two $\tilde{X}$ target predicates, which must have maximal values for CD to apply. Thus CD must derive (80) and not (81):

(80) John isn’t as bad a poker player as he was.

(81) *John isn’t as bad a poker player as he was a poker player.

In (81) I have removed the $\tilde{A}$-phrase $\Delta$ (much) bad, which is properly included in the $\tilde{N}$-phrase $\Delta$ (much) bad a poker player. The example is ungrammatical because the $\tilde{A}$-phrase is not the maximal value for one $\tilde{X}$ target predicate in CD. However, in (82) the same $\tilde{A}$-phrase is maximal:

(82) John isn’t as bad a poker player as he was a bridge player.

The reason is that though the $\tilde{A}$-phrase is properly included in the $\tilde{N}$-phrase $\Delta$ (much) bad a bridge player, the latter is not a possible value for one target predicate, since it is not structurally identical to any corresponding phrase in the head that could be a value of the other target predicate.11

11 Note that examples like *John is as heavy a man as he is tall a man can be derived by CD, since the deleted $Q$ is maximal. The ungrammaticality of such examples is not accounted for by CD itself, but should be
4.2.3. Complex NP Shift. We turn now to the rule of Complex NP Shift (also known as Heavy NP Shift) studied by Ross (1967, 3.1.1.3 and 4.3.2.3). As observed by Ross, this rule applies to both NPs and PPs, so we will formulate it informally as in (83).

\[(83) \quad W_1 - [v_P W_2 - \overline{X} - W_3] - W_4\]

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 & 5 \\
1 & 2 & \phi & 4 & 3 & 5
\end{array}
\]

This is only a rough formulation of the rule, but it will serve our purposes.

The Relativized A-over-A Principle correctly predicts the following data:

\[(84)\]

a. He considers many of my best friends stupid.
b. He considers stupid many of my best friends.
c. *He considers many of stupid my best friends.

The NP that is shifted in (84c) was not maximal for (83) because it was dominated by another NP that satisfied (83); cf. (84b).

Now consider (85).

\[(85)\]

a. He talked to many of my best friends about their stupidity.
b. He talked about their stupidity to many of my best friends.
c. *He talked to about their stupidity many of my best friends.

In (85c), the NP that was shifted was not maximal for (83), because it was dominated by another type 2 category that satisfied the target predicate in (83), namely the PP; cf. (85b).

Now consider (86).

\[(86)\]

a. I consider arguing with women who ride motorcycles (to be) silly.
b. I consider (to be) silly arguing with women who ride motorcycles.
c. *I consider arguing (to be) silly with women who ride motorcycles.

The PP that has been shifted in (86c) is not maximal, because it was dominated by another type 2 category that satisfied the target predicate in (83), namely, the gerundive NP; cf. (86b). Notice that such a PP can be extracted by other rules with impunity:

\[(87)\]

a. With whom would you consider arguing (to be) silly?
b. Who would you consider arguing with (to be) silly?

Nor does Ross's "rightward bounding" principle seem to be the explanation for (86c); recall that this principle prohibits rightward movement over more than one

---

attributed, I think, to the same factors that condition "AP Shift", discussed in Bresnan (1973). Cf.: *John is taller a man than I thought and John is a taller man than I thought. There are in fact speakers who accept examples like John is as heavy a man as he is a tall man, though I find them marginal.
S-boundary. But unlike NPs dominating S, the gerundive NP in (86) can "pied pipe"; that woman, arguing with whom I consider silly.... (See Ross (1967, 4.3) and Nanni (1975).)

In summary, we find that only a maximal category of the $X$ type—whether an NP or PP—can undergo Complex NP Shift.

4.2.4. Pied Piping. I now give a formulation of relativization that, subject to the Relativized A-over-A Principle, correctly derives exactly the set of facts recorded by Ross (1967, 4.3). There is much variation among speakers in judgments of such cases, and slightly different formulations of rules give different results. I choose this formulation primarily to show how a difficult problem can be solved using the techniques developed in this study; the facts given by Ross are more difficult to account for than some others. For an analysis of slightly different grammaticality judgments, see Nanni (1975).

Ross records the following facts (his (2.3), (4.163), and (4.167)):

(88) The government prescribes the height of the lettering on the covers of the reports.

(89) a. Reports which the government prescribes the height of the lettering on the covers of are invariably boring.
   b. Reports the covers of which the government prescribes the height of the lettering on almost always put me to sleep.
   c. Reports the lettering on the covers of which the government prescribes the height of are a shocking waste of public funds.
   d. Reports the height of the lettering on the covers of which the government prescribes should be abolished.

(90) a. *Reports of which the government prescribes the height of the lettering on the covers are invariably boring.
   b. *Reports on the covers of which the government prescribes the height of the lettering almost always put me to sleep.
   c. *Reports of the lettering on the covers of which the government prescribes the height are a shocking waste of public funds.

Consider now the following rule. (rel is just an expository convenience: it designates the class of relative pronouns, which, who, etc.)

(91) \[ NP - [gCOMP - W_1 - (P) - [NPW_2 - rel - W_3] - W_4] \]

\[
\begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\end{array}
\]

The NP-predicate covering terms 5–7 in (91) is the target predicate. Therefore, by
the Relativized A-over-A Principle, it must be maximal for each assignment to the context predicates. The predicate "(P)" is a context-predicate, which is satisfied either by a Preposition or by e. Note that for each of the prepositions in the following tree, adapted from Ross (1967, example (4.162)), the sister NP to its right is maximal.

Thus, there are three distinct maximal proper analyses \( \pi \) of (92) in which \( |(P)\|_\pi \neq e \). Therefore, (89a–c) are derived by rule (91).

There are four proper analyses \( \pi \) of (92) in which \( |(P)\|_\pi = e \) (one for each of NP_1, ..., NP_4), but only one is maximal (the one that assigns NP_1 to the target predicate in (91)). Therefore, of examples (89d) and (90a–c), only (89d) can be derived by (91).

Rule (91) needs one modification, however: as it is formulated, only NPs can be moved by relativization. But as is well known, PPs can also be moved:

\[
\text{(93) a. The reports about which I was talking. . . .}
\]
\[
\text{b. The reports which I was talking about. . . .}
\]

This is probably one reason why Ross (1967) analyzes PPs as NPs. We will accept Ross's insight that PPs and NPs have features in common, and reformulate (91) as (94):

\[
\text{(94) NP - } [\text{S COMP } W_1 - (P) - [\text{V } W_2 - \text{rel } W_3] - W_4]
\]

\[
\begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\hline
1 & 5 & 6 & 7 & 3 & 4 & \phi & 8
\end{array}
\]
Recall that $\bar{X}$ designates the set of categories \{NP, PP\}. Now (94) will apply as follows: for each assignment to (P), the maximal NP or PP next to it will be moved. When $|\langle P \rangle|_w \neq e$, all of (89a–c) and (93b) are derived. When $|\langle P \rangle|_w = e$, the highest NP or PP must be moved. Thus only (89d) and (93a) are derived in this case. This formulation also accounts for the sentences like (95) that Ross (1967, examples (4.168), (4.169)) included in his description of pied piping:

(95) a. The Greek authors which he has books by...
   b. ?*The Greek authors by which he has books...

Notice that if we had substituted the disjunction $\{\text{NP}, \text{PP}\}$ for $\bar{X}$ in (94), we would have wrongly derived (90c) and (95b), since the moved PPs would have been maximal in the NPs. The same remark can be made for QM, CD, and Complex NP Shift: in every case, a disjunction of particular categories would have yielded ungrammatical sentences. What we have discovered, then, is that with respect to the Relativized A-over-A Principle, certain sets of different categories \{B, C, D, E\} behave as though they were a single category A under transformation. These sets, which are denoted by our $\bar{X}$ predicates, thus appear indeed to be natural classes of categories.

Further study of the form and functioning of transformations will permit further development of the theory of syntactic categories.

4.3. Alternatives

Some of the facts accounted for in 4.2 have been attributed to separate constraints.

For the Question Movement facts, the Left Branch Condition has been proposed (Ross (1967, example (4.181))):

(96) No NP which is the leftmost constituent of a larger NP can be reordered out of this NP by a transformational rule.

To sustain this formulation, Ross proposed (1967, 4.3.2.1) that "how is analyzed as deriving from an underlying NP, and [adjectives] and [adverbs] are dominated by NP at the stage of derivations at which questions are formed." If Ross is correct that all these categories are to be derived from NP, the Relativized A-over-A Principle still accounts for the facts, making (96) unnecessary.

The Comparative Deletion facts have not hitherto been accounted for (but see Bresnan (1975)). The Relativized A-over-A Principle automatically explains them, given a formulation of CD like the one I have proposed. The Left Branch Condition cannot be consistently extended to these cases. For further discussion of (96), see Bresnan (1975) and Grosu (1975).

For the Complex NP Shift facts Ross proposes a special constraint (1967, (4.231)).
(97) No NP may be moved to the right out of the environment \([\text{NP} \text{P } \text{---}]\).

However, (97) does not carry over to cases like (86).

The Pied Piping facts are accounted for by the Pied Piping Convention, a version of which is stated by Ross (1967, (4.166)) as follows:

(98) Any transformation which is stated in such a way as to effect the re-ordering of some specified node NP, where this node is preceded and followed by variables, can reorder this NP or any NP which dominates it.

(98) could be made obligatory for Complex NP Shift and other rightward movement rules, but it remains unexplained why this is so. On our account, the asymmetry follows from the form of the rules: relativization can optionally "strand" a preposition in English, a fact that is expressed in the rule itself by the context predicate "(P)". From this optionality, the optionality of "pied piping" is predicted by the Relativized A-over-A Principle. If we deleted this context predicate from the rule, pied piping would be obligatory in English (as it is in many other languages).

Evidence that "pied piping" is the by-product of the form of rules under the Relativized A-over-A Principle, and not a special "convention", is that QM also pied pipes—but in a different way from relativization. In general, QM cannot pied pipe NPs, but only PPs. This can be easily captured by slightly modifying the form of QM itself. Ross does not describe pied piping for QM.

The Pied Piping Convention, as Ross observes, wrongly permits the ungrammatical examples (goa-c) to be derived, so he adds a special further constraint (see Ross (1967, 4.3.1)):

(99) ... there seems to be a constraint, in my dialect at least, which prohibits noun phrases which start with prepositions [i.e. PPs/JWB] from being relativized and questioned when these directly follow the NP they modify.

In summary, the Relativized A-over-A Principle provides a unified explanation for phenomena described by these special constraints (96)–(99), as well as hitherto unexplained facts. Moreover, it provides a principled explanation for the evident nonuniversality of "left branch" and "pied piping" effects; such effects of functioning vary with the form of individual transformations and structures and therefore cannot be expected to be universal, although the general conditions on the form and functioning of transformations proposed in this study are hypothesized as universals.

The Relativized A-over-A Principle does not account for many of the other phenomena studied by Ross (1967). Nor is it intended to: Ross has convincingly shown that further conditions govern the applicability of transformations. (For other conditions relevant to some of the facts discussed in 4.2, see Bresnan (1975) and Nanni (1975), and the references cited there.)
5. Conclusion

In conclusion, let us return to the issues raised at the beginning of this study. Two views of the relation between the logical form and the grammatical form of natural language sentences led to divergent lines of research on the problem of constructing a theory of syntactic categories. The generative semantics program was to reduce the categories of syntax to the categories of a variant of the predicate calculus using transformational exception features. The alternative proposed by Chomsky (1970) was to define the concept of syntactic category by means of a set of universal syntactic features and types (the $X$ theory). Cross-categorial transformations can then be formulated without disjunction or exception features by using the "$X$" predicates defined in section 4.

By relativizing the A-over-A Principle to the structural conditions of transformations, we have been able to study the relation between the form of transformations and their functioning. We found that the sets of categories designated by the $X$ predicates behave under different transformations like a single category "A" with respect to the Relativized A-over-A Principle, providing new evidence that these are indeed natural classes of categories. Further, the $X$ theory and the maximality principle together provide a more general and unified explanation for left-branch, pied-piping, and other phenomena than previous accounts based on category-reduction and applicational constraints. These results suggest that those extensions of the classical theory of transformational grammar which preserve the structure-dependence of transformations and the autonomy of syntax will be promising areas for further research.\(^\text{12}\)

References


\(^\text{12}\) I would like to thank Emmon Bach, Debbie Nanni, Barbara Partee, and Lisa Selkirk for helping me to clarify some of the ideas in this study. I am also grateful to Emmon Bach and Stanley Peters for suggesting various improvements in the manuscript. None of them should be associated with any errors of analysis or interpretation, for which I take full responsibility.
FORM AND FUNCTIONING OF TRANSFORMATIONS


20C–132
MIT
Cambridge, Massachusetts 02139