THE X-BAR THEORY OF PHRASE STRUCTURE

ANDRÁS KORNAI

Stanford University
and
Hungarian Academy of Sciences

GEOFFREY K. PULLUM

University of California, Santa Cruz

X-bar theory is widely regarded as a substantive theory of phrase structure properties in natural languages. In this paper we will demonstrate that a formalization of its content reveals very little substance in its claims. We state and discuss six conditions that encapsulate the claims of X-bar theory: LEXICALITY — each nonterminal is a projection of a preterminal; SUCCESSION — each X^{n+1} dominates an X^n for all $n \geq 0$; UNIFORMITY — all maximal projections have the same bar-level; MAXIMALITY — all non-heads are maximal projections; CENTRALITY — the start symbol is a maximal projection; and OPTIONALITY — all and only non-heads are optional. We then consider recent proposals to 'eliminate' base components from transformational grammars and to reinterpret X-bar theory as a set of universal constraints holding for all languages at D-structure, arguing that this strategy fails. We show that, as constraints on phrase structure rule systems, the X-bar conditions have hardly any effect on the descriptive power of grammars, and that the principles with the most chance of making some descriptive difference are the least adhered to in practice. Finally, we reconstruct X-bar theory in a way that makes no reference to the notion of bar-level but instead makes the notion 'head of' the central one.*

1. INTRODUCTION. The aim of this paper is to undertake a systematic examination of the content of X-bar theory and its consequences for language description. The topic is truly central to modern grammatical theory. X-bar theory is discussed in almost all modern textbooks of syntax, and it is routinely assumed as a theory of phrase structure in a variety of otherwise widely differing schools of grammatical thought such as government-binding theory (GB), lexical-functional grammar (LFG), and generalized phrase structure grammar (GPSG).[1]

One of the primary tasks of syntactic theory is to explain how sentences are built from words. This explanation is generally conceived of in terms of assigning syntactic structures to sentences. The exact form and content of the structures that should be assigned has been strongly debated, but there is overwhelming agreement that constituency information is a crucial element in any adequate analysis of sentential structure. However, there is very little agreement among syntacticians concerning the phrase structure of even the simplest sentences. Questions concerning the exact labeling and bracketing of the constituents in such sentences as *Mary will have been swimming* (what is the constituency of the auxiliary and main verbs?) or *Is John a good boy?* (how many branches does the root node have, and how many does the predicate NP *a good boy* have?) or *Mary gave Sue the book* (what are the constituency relations between the nonsubject NPs and the verb?) receive different answers in virtually each new study that addresses these issues.

There is a clear need to develop new ways of using linguistic evidence to rule out hypotheses about phrase structure and force choices of structure in such cases. By embodying substantive principles of phrase structure, X-bar theory should narrow down the range of choices to a small, preferably universal set of possible analyses. A maximally strong version of X-bar theory would in fact narrow down the set of possible choices to one, and would thereby eliminate the need for language-specific phrase structure rules in the manner suggested by Stowell (1981).

We will show that the specific proposals in the literature on X-bar theory fall short of providing such an ideally strong theory of phrase structure. We undertake a systematic examination of the content of intuitively-based claims in the literature and re-express them in theory-neutral mathematical terms. We lay out a set of restrictions on phrase structure rule systems that can be seen as constitutive of X-bar theory and discuss each with regard to the extent to which it is adhered to in linguistic practice. We then investigate the technical details of X-bar theory in light of the problem of delimiting the space of possible phrase structure analyses. We argue that the interest of the standard X-bar restrictions resides mainly in the notion of 'headedness', bar-levels as such being epiphenomenal and even eliminable. Finally, we restate the standard restrictions in terms of a 'bar-free' version of X-bar theory.

2. CONSITUTIVE PRINCIPLES OF X-BAR THEORY. In this section we will lay out a set of restrictions that are jointly constitutive of (a rather strong variety of) X-bar theory. We will discuss each with regard to the extent to which it is adhered to in practice in the syntactic literature. Although it turns out that none of the restrictions we draw out of theoretical discussions (especially Jackendoff 1977) are fully respected in descriptive works over the past two decades, it is nonetheless useful to study the properties of an X-bar theory that imposes all of them. The idealized theory provides a neutral standard for comparison of X-bar theories.

2.1. LEXICALITY. The primary defining property of X-bar systems is what we shall call LEXICALITY, which requires all phrasal categories to be projections of lexical categories. A primitive set of preterminals

---

[1]For example, X-bar theory is discussed in all three chapters of Sells (1985): in connection with government-binding theory (p. 27), with generalized phrase structure grammar (p. 81), and with lexical-functional grammar (p. 139).

(categories that can immediately dominate terminals; in linguistic terms, lexical categories) is assumed. Lexicality means that the complete category inventory (the NONTERMINAL VOCABULARY) is exhausted by two sets: (a) the set of preterminals (categories like noun or verb), and (b) a set of PROJECTIONS of these (such as various kinds of noun phrase or verb phrase).

Bar-level originates as a notation for phrasal category labels that makes it clear how they are based on lexical category labels. Thus in a complex noun phrase like [a [[student] [of linguistics]]], the head noun *student* might be labeled N (with bar-level zero), the noun-plus-complement group *student of linguistics* might be labeled N′ (with bar-level one; primes are used instead of overbars for typographical convenience), and the full phrase *a student of linguistics* might be labeled N″ (with bar-level two).

There are many ways in which Lexicality could be built into the theory of categories. It is customary in current work to regard categories as sets of feature specifications and bar-level as a feature taking integers as values (see Gazdar and Pullum 1982; Abney 1987:236; Gazdar et al. 1988). But since issues of lexical feature structures and phrase structure are largely orthogonal to our concerns, we will adopt the proposal of Bresnan (1976) and treat categories as ordered pairs, the first member in each pair being a preterminal and the second being an integer denoting bar-level — the number of bars (or primes). The pair $\langle X, n \rangle$ will be written X^n, in keeping with a frequently used notation for bar-level.

Since the process of rewriting preterminals as terminal symbols (i.e. the process of lexical insertion) is now generally assumed to be context-free, in the mathematical parts of the following discussion we will generally treat the zero bar-level categories (i. e. preterminals, i. e. lexical categories) as the terminal vocabulary of the grammar. That is, we will concentrate on describing 'languages' that are in fact strings of lexical categories rather than strings of words. This is because we are much more concerned with assigning analyses to sentences (what is sometimes called STRONG GENERATIVE CAPACITY) than with what sets of strings of words can be obtained (WEAK GENERATIVE CAPACITY).

We will use the standard notations for context-free grammars (CFGs): V_N denotes the set of nonterminals (the entire category set); V_T is the set of terminals (here, lexical categories); S, a specific member of V_N, is the start symbol (which labels the root node of constituent structure trees); and P is the set of phrase structure rules. The rules are written in the form $\alpha \rightarrow W$, where $\alpha \in V_N$ and $W \in (V_N \cup V_T)^*$. (The expression '$(V_N \cup V_T)^*$' denotes the set of all strings, of length zero or greater, that are formed from the symbols in the union of V_N and V_T. More generally, if A is a set of symbols, A^* is the set of all strings formed using symbols from A.) It is often necessary to make specific mention of a string of zero length, the EMPTY STRING; for this we will use the notation e.

We now give an exact statement of the Lexicality condition. What it says is basically that a CFG observes Lexicality if and only if all its nonterminals are formed by addition of superscript integers to terminals (to lexical categories, that is).

(1) Definition: a CFG observes Lexicality iff every nonterminal is X^i, where $X \in V_T$ and $i > 0$.

It should be noted that a CFG containing some X^i (for $i > 0$) need not contain every X^j for $j < i$. That is, there is no provision in the definition of Lexicality for an unbroken projection line between lexical head and maximal projection. However, other conditions do guarantee this. As we shall see, the rule system of an X-bar grammar is such that the bar-level of a node may be interpreted as the distance of the node from the preterminal in a connected chain of gradually more inclusive phrases founded on the preterminal.

3

Lexicality is by no means generally observed by linguists who claim to be assuming X-bar theory. In works such as Chomsky (1970) and Emonds (1976), for example, the category S does not participate in the bar system at all. More generally, any use of such frequently encountered minor categories as Adv[erb], Af[fix], Agr[eement], Aux[iliary], Art[icle], Cl[itic], C[omplementizer], Conj[unction], Cop[ula], Deg[ree], Det[erminer], I[nflection], M[odal], Neg[ative], P[a]rt[icle], Spec[ifier], T[ense], or Top[ic] will constitute a departure from Lexicality, unless they are assigned to some bar-level. Furthermore, they are a serious problem for the feature analysis of category systems. Even Jackendoff (1977), who clearly sets out to incorporate all categories in a single syntactic feature system, employs additional categories with no place in the system of features and bar-levels ('T' on p. 50 and 'Comp' on p. 100 are examples).

In recent GB work there is a growing tendency to take such minor categories to be zero-bar lexical categories and to create a full set of nonterminals projected from them. For instance, in Chomsky 1986 the category S′ is replaced by C′ and C″, which are projections of C (COMP).[2] Moreover, S is also abandoned as the label for sentences, being reanalyzed in terms of an entirely distinct projection founded on I (INFL). And in some works (see Abney 1987 for an extended discussion), another projection is based on D (Determiner), the traditional 'noun phrase' being reanalyzed as a DP (Determiner Phrase), with a subconstituent labeled N″ that contains the noun. Under such analyses, the structure of a sentence as simple as *Birds eat worms*, assuming 2 as the maximal bar-level and assuming also the DP hypothesis, contains a minimum of twelve nonterminals, two of them (I and D) having null heads. The more recent analysis of Pollock (1989) adds full phrasal projections for the categories T (tense), Agr (agreement), and Neg (negative particle), so that the structure of *Birds eat worms* has at least fifteen nonterminals, and *Birds do not eat worms* has over twenty.

The postulation of growing numbers of invisible heads and their abstract projections makes it harder to ascribe any content to claims about the relation between the terminal and nonterminal vocabularies under X-bar theory. We shall see below (in section 3.2) that arbitrary CFGs can be emulated by X-bar grammars largely by virtue of this sort of expansion of the category set and use of categories realized as the empty string. Yet in GB work such abstract projections and empty categories have multiple uses (see Chomsky 1986 and Pollock 1989 for many examples), and they are not eliminable.

2.2. Succession. It would be possible to comply with Lexicality simply by renaming all the nonterminals of the grammar as projections of some arbitrarily chosen preterminal. If we change the rules accordingly, the resulting Lexicality-observing grammar will be not only weakly equivalent to the original one but also strongly equivalent (at least if isomorphism between structural descriptions is taken as the criterion for strong equivalence). This makes it clear that the Lexicality constraint, as defined above, has no substantive content in itself. In order to capture the notion of bar-levels, it is also necessary to have unbroken projection lines, so that the bar-level number corresponds to the number of steps up from the head preterminal (as opposed to being arbitrarily chosen). The condition we call Succession guarantees precisely this, by constraining the rule system, rather than the category system.

(2) Definition: a Lexicality-observing CFG observes Succession iff every rule rewriting some nonterminal X^n has a daughter labeled X^{n-1}.

Hellan (1980:67), who assumes that this daughter is unique, calls X^{i-1} the KERNEL of the right hand side, and calls the other elements in the right hand side the *dependents*. Although it is more usual to use

---

[2]The idea of making complementizers heads of complement clause constituents goes back at least to Langendoen (1975:540), but has gained general favor only in recent years.

the term HEAD for kernel (see e. g. Gazdar et al. 1985), we adopt Hellan's usage when talking about the requirement imposed by the Succession condition because we wish to separate out several distinct notions that are commonly conflated. First, when we use the term 'head', it will denote the head daughter in a local tree rather than any element in a rule. Second, the notion 'category (in a rule) guaranteeing that Succession is met' (i.e. kernel) must be distinguished from the notion 'obligatory substring in the right hand side of a rule'. For the latter notion we will use the term CORE (see (6) below). Later we shall see that when the Optionality condition is imposed, kernels and cores coincide.

Under the parsimonious interpretation of bars found in the original work of Harris (1951, ch. 16), new bar levels are introduced only for non-repeatable substitutions; but this is incompatible with Succession. The Succession-observing interpretation of bars necessitates a new level wherever additional dependents are introduced; this is basically the approach taken by Jackendoff (1977). Present-day linguistic practice seems to be an uneasy mixture of the two. Although Succession is considered definitive by many linguists like Jackendoff (1977), it is weakened in some way in nearly all works that assume X-bar theory. It is not obeyed in any theory that allows exceptions to Lexicality, for example, in theories like that of Chomsky (1970) and Emonds (1976) where S has no head (and the rule expanding S has no kernel).

Emonds also violates Succession in allowing a rule to introduce a head of bar-level 0 under a mother category of bar-level 2 (1976: 15, Base Restriction II), as do Gazdar et al. (1985:61f). Succession is also violated strikingly by the phrase structure rule $N^1 \rightarrow N^2$ proposed by Selkirk (1977:312), and by any variant of the familiar rule $NP \rightarrow S$ used for sentential subjects, as stressed by Richardson (1984).

Succession is incompatible with the recursive introduction of modifiers such as the adjective modifier *very* (cf. the rule $Adj \rightarrow very + Adj$ in Chomsky (1957:73) or prenominal adjectives in the noun phrase (cf. the rule $N^1 \rightarrow A^2N^1$ tacitly assumed in Gazdar et al. 1985:126) or postnominal PP modifiers (Gazdar et al. 1985:129 give '$N^1 \rightarrow H, PP$', which is equivalent to '$N^1 \rightarrow N^1 PP$'). Interestingly, such analyses are crucial to the argument of Hornstein and Lightfoot (1981:17–24), who defend the explanatory power of a system incorporating the X-bar theory on the basis of anaphora facts (such as that in *She told me three* [$_{N'}$ *funny* [$_{N'}$ *stories*]] *but I didn't like the one about Max*, the anaphoric phrase *the one* can mean either 'the stories' or 'the funny stories'). Succession-violating structures like [$_{N^1}$ $N^1 P^2$] are an essential part of the analysis (due to Baker 1978) that Hornstein and Lightfoot advocate and elaborate. Thus Hornstein and Lightfoot's extended argument for the explanatory force of X-bar theory is inconsistent with the single most commonly cited restriction imposed on rules by X-bar theory (Jackendoff 1977:34, Stowell 1981:70, etc.; Radford 1981:96ff gives a pedagogical review of the argument, and notes the inconsistency with Succession on p. 104).

Jackendoff relaxes Succession in a number of cases, which he refers to as 'a principled class of exceptions' (p. 235). These exceptions come in two classes. One covers coordination rules, where a category X^k can immediately dominate a string of other X^k categories without any daughter being labeled X^{k-1} as Succession would require,[3] and another covers rules in the form '$X^k \rightarrow t Y^k$' where t is a syncategorematic terminal (a grammatical formative belonging to no category).

We will use the term Weak Succession for the condition that the head of a phrase X^k is that daughter which (i) is a projection of X, (ii) has bar-level j equal to or less than k, and (iii) has no other daughter that is a projection of X with fewer than j bars. This condition is adopted by Emonds (1976:15), Gazdar

---

[3]It is not necessary that coordination should constitute an exception to the notion that every constituent has a head; cf. the treatment in Gazdar et al. 1985, more fully expounded in Sag et al. 1985. The usual formulation of Succession does seem to exclude it, however. Sag et al. assume a different principle, under which a head has the same bar-level as its mother unless some statement in a grammar forces things to be otherwise.

(1982), and Gazdar, Pullum and Sag (1982). We will discuss Weak Succession in greater detail in 3.2. Here it is sufficient to note that it has the same language-theoretic consequences as Succession as long as the requirement of Maximality (see below) is imposed.

2.3. UNIFORMITY. Many linguists assume that the maximum possible bar-level is the same for every preterminal. We will refer to this condition as UNIFORMITY. The property of having the maximum permitted value for bar-level constant across all the preterminals makes it possible to fix a single number m as defining the notion MAXIMAL PROJECTION, and we will define Uniformity by stipulating that such a number exists:

(3) Definition: a Lexicality-observing CFG observes Uniformity iff

$$\exists m \in N[V_N = \{X^i \mid 1 \leq i \leq m,\ X \in V_T\}]$$

Notice that this makes V_N identical with the set of ALL X^i with i between 1 and m, and thus ensures that there are no gaps (if there is an X^i and an X^{i-2} there must be an X^{i-1}). Uniformity does not require that all these categories be used in rules (but Succession will force that result once any given X^m is used). In the remainder of this paper, if we are assuming Uniformity we use the symbol m as a constant.

Not every proponent of X-bar theory has accepted Uniformity. Jackendoff maintains it strictly; in Jackendoff (1977), $m = 3$. But Dougherty (1968), Williams (1975), Bresnan (1976), and others have defended systems that do not satisfy it. And there have also been many different proposals about the optimal value for m if Uniformity is assumed, from the logical low of one (entertained as a possibility by Emonds (1976:16), and assumed by Starosta 1984 and Stuurman 1985) to numbers as high as six or seven.

The Uniformity hypothesis, i.e. that maximal projections are on the same bar-level, played a prominent role in the early development of X-bar theory and it is still retained in most frameworks. However, since Uniformity can always be achieved by introducing new nonterminals,[4] the general acceptance of this constraint does not signify a real consensus. Moreover, the idea that a great number of significant generalizations could be captured in terms of syntactic features relating uniform projections of different categories turned out not to be very fruitful (Kean 1978 offered an early complaint along these lines). As the additional motivation offered by Jackendoff (1977, ch. 4), namely that each level has a separate semantic characterization, has been persuasively criticized (cf. Verkuyl 1981), we must conclude that the Uniformity hypothesis has received very little support.

If Uniformity does not hold, the number of necessary bars has to be fixed individually for every category. In this context, the question whether rules in the form '$X^n \to \dots X^n \dots$' should be permitted or not becomes significant. If we allow such rules, we have a larger variety of rules to choose from, and more complex constructions can be analyzed with the same number of bar-levels.

2.4. MAXIMALITY. Let us now turn to another important condition, familiar from most variants of X-bar theory: the requirement that EVERY NON-HEAD DAUGHTER IN A RULE IS A MAXIMAL PROJECTION. There is an intuitively interesting claim here: that syntax is NEVER a matter of putting words together

---

[4]In Succession-observing grammars it is also necessary to add non-branching rules to the rule set in order to maintain the unbroken projection lines.

with other words to make phrases (though in some theories of grammar, such as the one proposed in Hudson (1984), this is ALL that syntax does). Under Maximality, no syntactic rule can introduce or combine two lexical categories; some rules combine a lexical (head) category with a (maximal projection complement) phrase, while others combine (head) phrases with other (maximal projection) phrases.

(4) Definition: a CFG observing Lexicality and Succession observes Maximality iff for every rule $X^n \to Y X^{n-1} Z$, the strings Y and Z are in $V_M{}^*$, where $V_M = \{X^m \mid X \in V_T\}$.

Maximality is observed by most varieties of X-bar theory, but explicit departures from it can be found in Gazdar (1982) and Gazdar, Pullum and Sag (1982), where S is taken to be a projection of the category 'verb' (specifically V^2), VP is distinct from it only in bar-level (VP = V^1), and V^1 complements are permitted, allowing some lexical heads to have complements of non-maximal bar-level.[5]

Jackendoff in fact subscribes to only a weakened version of Maximality. According to our definition, non-head daughters (more precisely, non-kernel daughters) must be maximal projections. But Jackendoff (1977:36) also permits 'specified grammatical formatives' such as perfect *have*, number morphemes, case markers, or tense particles in non-head position: his requirement will be called WEAK MAXIMALITY in order to distinguish it from the stronger version of Maximality as defined above, which we can call Strong Maximality to avoid ambiguity.

It is obvious that a grammar satisfying Strong Maximality also satisfies Weak Maximality, for it simply represents the special case where the number of grammatically specified formatives introduced in nonlexical rules is zero. Furthermore, from any CFG satisfying Weak Maximality an equivalent grammar satisfying Strong Maximality can be constructed by carrying out three operations: (i) assign each grammatical formative t to a previously unused lexical category α; (ii) add the rules $\alpha^k \to \alpha^{k-1}$ for each k such that $1 \le k \le m$; (iii) replace t by α^m in the right-hand side of every rule where t appears.

In other words, Maximality (both strong and weak) can be complied with in ways that deprive it of consequences. When Jackendoff encounters categories like Art (article) or Prt (verb particle) or M (modal) which show no evidence at all of having a three-level structure of specifiers and complements, he postulates the X-bar skeleton that would allow such a structure nonetheless; thus he posits the rules 'Art^3 $\to$ Art^2', 'Art^2 $\to$ Art^1', 'Art^1 $\to$ Art^0', where 'Art^0' dominates *the* or *a*. The categories 'Art^3', 'Art^2', and 'Art^1' are postulated only to comply with Maximality, and the claim is made that it is a lexical accident that no articles allow subcategorized phrasal complements in 'Art^1', phrasal specifiers in 'Art^3', etc.

Again, when Jackendoff considers the possibility that it is correct to postulate subjectless ('orphan') VP complements for some verbs, he proposes that V^3 complements containing only a V^2 could be employed. Clearly he is assuming some way of ensuring that in root sentences and in complements to verbs like *think* the subject of an S (= V^3) will always appear, but in the complement of a verb like *try* it will never appear. But this gives exactly the effect of a violation of Maximality.

Given the possibility of such analyses, it seems intuitively clear that nonmaximal complements could always be replaced by maximal complements with missing daughters, so that Maximality has no consequences at all. And this is indeed the case under a wide range of assumptions. By introducing dummy terminals and renaming every nonterminal as the first (and only) projections of these, every CFG can be turned into a Maximality-observing grammar. However, we shall show in section 4 that if

---

[5]The more recent work of Gazdar et al. (1985) observes Maximality in such cases.

Succession is observed and lexical items (terminals) are required to retain their category memberships, it is possible for the requirement of Maximality to decrease the power of a grammar.

Maximality might be interpreted as an explanatory principle in terms of acquisition, if one takes seriously the idea of drawing direct links between the formalism of grammatical theory and the infant's acquisition task (a point on which we remain neutral here). If we assume that children have knowledge of the distinction between heads and dependents, a child finding some lexical element X in a dependent position will automatically assume that its maximal projection X^{max} can also appear there. Thus, a productive pattern like 'S → NP VP' could be acquired solely on the basis of data in the form 'N V'. For some speculations along these lines, see Grimshaw (1981) and Pinker (1984).

2.5. CENTRALITY. The usual definition of a grammar demands that a single designated symbol be admissible as the start symbol in a derivation. Maximality covers constituents introduced in right hand sides of rules, and thus says nothing about the start symbol. Centrality requires that the start symbol must be the maximal projection of some preterminal.

(5) Definition: a Lexicality-observing CFG observes Centrality iff the start symbol is the maximal projection of a distinguished preterminal.

Succession guarantees that each right hand side of a rule will contain a category that is the kernel. But it does not require that kernels in rules be unique. It guarantees only that rewriting any X^i will give at least one X^{i-1}. For $i = 1$, this will be a preterminal symbol, and for $i > 1$, X^{i-1} can be rewritten to yield a X^{i-2} and so on until we arrive at some preterminal X^0, which will be called the LEXICAL HEAD of the X^i constituent.[6]

Since the lexical head of the start symbol must appear in every preterminal string if both Lexicality and Succession are to be observed, Centrality requires that there must be one lexical category such that every string in the language contains at least one instance of that category. Under certain assumptions, this can decrease the descriptive power of CFGs. For example, suppose e-rules (rules of the form $A \rightarrow e$, where the right hand side is empty) are disallowed; then under Lexicality, Succession, and Centrality, some quite simple languages are not describable. For example, a language in which bare verbs and bare nouns can both constitute sentences would be excluded.

At least some natural languages appear not to be compatible with the claim that some lexical category is overtly realized in every sentence. For example, in Hungarian, Russian, Tagalog, Jamaican Creole, and many other languages, sentences with an adjectival predicate do not necessarily exhibit a copular verb. In these cases, either the initial symbol is not V^n (as Jackendoff proposes), which would mean that the category label for root nodes is not a universal, or a zero copular verb must be postulated. A prohibition against null kernels would thus seem to be too stringent to be compatible with reasonable natural language grammars.

Jackendoff's system obeys Centrality, for his start symbol (the label associated with the category of ordinary sentences) is analyzed as V^m (i.e. V^3). But not all X-bar systems observe Centrality. It is not assumed by Harris (1951) or by Chomsky (1970), with exocentric S as initial symbol, and it is also

---

[6]We say 'lexical head' rather than 'lexical kernel' because it is defined in the tree, not on the rule. We assume uniqueness of heads here, as most developers of X-bar theory have done; notice that it is not logically necessary that heads should be unique, and some works, e.g. Gazdar et al. 1985 and Sag et al. 1985, have developed analyses that crucially assume otherwise.

denied by Emonds (1976), who assumes an exocentric non-embeddable initial symbol E (p. 52 et seq.). Centrality does seem to be observed in recent GB analyses that analyze root clauses as either I^m or C^m.

2.6. OPTIONALITY. In intuitive terms, Optionality is the condition that NON-HEADS ARE ONLY OPTIONALLY PRESENT. More formally:

(6) Definition: a CFG $G = \langle V_N, V_T, P, S \rangle$ observes Optionality iff for every rule in P of the form $\alpha \to W$ there exist β, W_1, W_2 such that

    i. $\beta \in (V_N \cup V_T)$;

    ii. $W_1, W_2 \in (V_N \cup V_T)^*$;

    iii. $W = W_1 \beta W_2$; and

    iv. the rule $\alpha \to W_1' \beta W_2'$ is also in P for all strings W_1' and W_2' constructible by deleting some elements from W_1 and W_2, respectively.

In such rules, β will be called the CORE.

If a grammar observes Optionality, that fact together with the identity of the core of each rule can be inferred from the set of rules. First we collect the right hand sides of rules rewriting the same nonterminal α. In the resulting set $R(\alpha)$, all strings of length one (and only these) are cores with respect to α. Then we check every position where cores with respect to α appear in the strings in $R(\alpha)$: CORE POSITIONS will have the property that the deletion of arbitrary elements from their left and (or) right hand sides will always yield strings that are in $R(\alpha)$. A CFG is Optionality-observing if and only if we can identify cores and core positions for every rewrite rule.

Moreover, if cores are unique (we will discuss this assumption later on), bar-levels can be assigned to the nonterminals of an equivalent grammar in accordance with the principle of SIZE-DEPENDENCY[7] in the following way.

First we collect those nonterminals that can be directly rewritten with a preterminal core into the set B_1: these will be the one-bar elements. Then we collect those nonterminals that are not members of B_1, but can be directly rewritten with a one-bar core: the resulting set B_2 will contain the two-bar elements. Because the number of nonterminals is finite, repeating the process will give us finitely many disjoint nonempty sets B_1, B_2, ..., B_n: for the sake of completeness, we can collect the preterminals in B_0. Certain dependents (those that appear in core position in other rules) will also receive bar-levels in the process. The remaining nonterminals can be simply eliminated from the grammar without loss because these will never appear in derivations resulting in strings of terminals (thus the algorithm we sketch here gives an X-bar grammar without useless symbols).

Now, if for every $i \geq 0$ and for every k-tuple of nonterminals in B_i, we can find at least k elements in B_{i-1} appearing in core position in the right sides of the rules rewriting the elements of the k-tuple in question, it will be possible to rename the nonterminals (without changing the previously assigned bar-levels) in such a manner that the resulting grammar will be Lexicality-observing and Succession-observing. The actual construction is somewhat tedious: see Ore (1963, ch. 4.1) for an elementary exposition of the

---

[7]A key principle for assignment of level-number to an expression is the following: if the expression X belongs to the category C^n, and the expression XY (or YX) has roughly the same distribution possibilities as X, then XY belongs to the category C^m, where $m > n$ (Hellan 1980:65).

graph-theoretic lemma that has to be applied at every bar-level. It should be mentioned here that the above condition on k-tuples is only a SUFFICIENT one; but as grammars describing natural languages seem to meet it without exception, there seems to be no need to develop weaker conditions entailing Succession for Optionality-observing grammars.

This formalization brings into sharp relief a split between linguistic theory and practice. While most linguists espouse the principle of Optionality[8] for phrasal non-heads (cf. Emonds' Base Restriction III (1976:16) and Jackendoff 1977:45), in practice they routinely use analyses that violate Optionality. For example, Emonds (1976:16) clearly recognizes that there are lexical items whose lexical entries show that they 'require obligatory complements of various sorts' but he also asserts that 'there are no object, adverb, predicate attribute, relative clause, or comparative complements that must appear within ALL [maximal projection] structures given by a particular base rule.'

The trouble with this position is that it seems to be based on an equivocation. Consider the claim that NP is optional under VP. In a VP where the verb is *have*, the accompanying NP is absolutely obligatory (cf. *Lee doesn't have a car* but **Lee doesn't have*). If subcategorization is treated separately from major category membership, it can be said that NP is optional in VP, since there happen to be verbs like *elapse* which do not require an NP complement. But it does not follow that removal of an NP daughter from a VP will preserve grammaticality. And moreover, it is surely not claimed that a language MUST have lexical items of each subcategorial type necessary to allow for the full array of expansions for each category. English happens to have prepositions like *over* which can be used without their NP objects, but in all likelihood this is an accident of the lexicon; we would not be too surprised if some other language happened to have only transitive prepositions like *at*.

A separate problem, distinct from the issue of subcategorization of obligatory complements, is that 'nonphrasal' nodes can be obligatory in a base rule. Emonds (1976:17n) cites examples such as Tense in English and the Article constituent in French. (There are many other such examples; consider, for instance, the obligatory *ang/ng* specifier in Tagalog NPs.) Such cases are highly problematic for Jackendoff, who makes Art a preterminal with a full projection to the phrasal category Art‴, the latter being an optional non-head immediate constituent of noun phrases. Under Optionality, there is no way to guarantee the presence of any such non-head, non-subcategorized, maximal projection specifiers.

The worst problems for Optionality occur within S. Consider two current assumptions about S in English, namely the assumption of Jackendoff that S is a projection of V and the assumption of Chomsky (1981) and others that the former 'S' is a projection of I (the former 'INFL'), a category subsuming information about tense and verb-subject agreement inflection.

Under the Jackendoff view, Optionality makes tense optional, even in root clauses, predicting sentences like **She be nice*. Tense is in fact doubly optional for Jackendoff, being a non-head daughter of a non-head maximal projection M‴ that is itself permitted under Optionality to be absent from S. Jackendoff refers to this double anomaly as 'a minor exception' (p. 50).

Under the Chomsky view, tense can be made syntactically obligatory in S by including it as part of the feature structure of I; but now not only subjects but also verb phrases are optional within S: *-ed* is predicted to be the only obligatory element in the sentence *Sandy walked home*, and **Did* is predicted to be a grammatical and non-elliptical sentence.

Because of these and similar problems, it is generally assumed that Optionality can be overridden

---

[8]Chomsky (1970) does not really address the issue, although he uses the phrase 'optional complements' once (p. 210). Jackendoff (1977:36) claims that 'probably' all non-heads are optional, and refers back to the section later for the claim that 'only heads are obligatory constituents' (1977:43).

by considerations 'extrinsic to the phrase structure rules' (see Jackendoff 1977: 44, 50) and thus cannot force the choice between competing phrase structure analyses.

3. The content of X-bar theory. The picture emerging from the previous discussion permits little confidence in X-bar theory as a strong and substantive theory of phrase structure. Some constraints (Lexicality and Uniformity) are effectively without any consequences in other than esthetic terms, and others (Succession, Optionality) have content but cannot be maintained under assumptions that linguists generally wish to make about the phrase structure of natural languages. Yet the fundamental problem of using linguistic evidence to rule out hypotheses about phrase structure is perhaps even more acute today than it was in 1970, given the proliferation of competing syntactic theories in the decades since the introduction of the X-bar convention.

In this section we present a version of X-bar theory that contributes directly to the solution of this problem. In section 3.1 we analyze the content of standard X-bar theory and conclude that its apparent failure to delimit the range of possible analyses is due to an excessive emphasis on the universal over the parochial. We argue that X-bar theory, which serves primarily as a heuristic tool in present-day syntactic research, should have a more overt role in the description of individual languages. In section 3.2 we present some mathematical results concerning the effects produced when phrase structure rules are constrained by X-bar theory.

3.1. X-bar grammar without phrase structure rules. In works like Emonds (1976) and Jackendoff (1977), X-bar theory is construed as a set of constraints on phrase structure rules of the base. More recent works, starting with Stowell (1981), have preferred to interpret X-bar theory as a set of conditions directly applied to structural representations. Stowell (1981:70) gives a list of 'plausible and potentially very powerful restrictions on possible phrase structure configurations at D-structure':

(7)a. Every phrase is endocentric.

   b. Specifiers appear at the X^2 level; subcategorized complements appear within X^1.

   c. The head always appears adjacent to one of the boundaries of X^1.

   d. The head term is one bar-level lower than the immediately dominating phrasal node.

   e. Only maximal projections may appear as non-head terms with a phrase.

We take (a) together with (d) to mean that Succession (hence also Lexicality) is observed; (b) is just a definition of the notions 'specifier' and 'complement'; and (e) corresponds to Maximality. Stowell's (c) is new, and not mentioned elsewhere in the literature as a part of X-bar theory. We will call it Peripherality, because it requires that lexical heads must be phrase-peripheral:

(8) Definition: a Lexicality-observing CFG observes Peripherality iff in any rule rewriting X^1 as YX^0Z, either $Y = e$ or $Z = e$.

Note that in any grammar limited to binary branching, as in the proposals of Kayne (1981) or Pollard (1984), Peripherality is trivially satisfied.

Stowell proposes to keep these purportedly restrictive generalizations but to 'eliminate' phrase structure rules from grammars. A more precise way of putting it is that he proposes to eliminate

PAROCHIALITY in PS rules: he wants to remove the possibility of one language having different PS rules from another. This goal, of course, is familiar to students of the history of generative grammar: it used to be known as the Universal Base Hypothesis (UBH), a phrase which was coined some time in the late 1960s (we have been unable to date it precisely) to denote a strengthened version of suggestions made by Chomsky concerning the universality of 'much of the structure of the base' (1965:117). Under the UBH, PS rules can be said to be 'eliminated' inasmuch as the base component is part of Universal Grammar and does not vary from one language to another, so that no individual grammar has to say what its PS rules are.

One novel element in this new version of the UBH is that the implicit universal base, far from having no PS rules, has infinitely many. If we enforce Lexicality, Succession, Maximality, Uniformity, Centrality, Optionality, and Peripherality on PS rules, and do no more than that, then for any given choice of finite nonterminal and terminal symbol vocabularies, there is a unique infinite set $\mathcal{R}$ of rules that meet these conditions. Assuming that the available D-structure trees for all languages are all and only those that satisfy the X-bar principles (as rephrased to apply to structures rather than rules) is exactly equivalent to assuming that the D-structure trees for all languages are the trees generated by the rules in $\mathcal{R}$.[9]

Infinite grammars are not technically CFGs, and infinite sets of CFG rules are not necessarily equivalent to any CFG. But in this case we can show that the infinite set of rules is equivalent to a CFG. The set of all right hand sides of rules meeting the X-bar conditions is a regular set. Specifically, for each X and each k such that $0 < k \leq m$, the set of W such that $X^k \to W$ is given by the following regular expression:

(9) $(X^{k-1} V_M{}^*) + (V_M{}^* X^{k-1})$

It follows that the particular infinite grammar involved here generates a context-free language (see Langendoen 1976), i. e. generates a set of strings that is also generated by some CFG.

Stowell's version of the UBH simply claims that all natural languages share the infinite set of pre-lexical-insertion D-structures generated by $\mathcal{R}$, and that the set of surface structures, logical forms, and phonetic representations for any natural language can be derived from this universal base language by means of transformations and other components of grammar.[10] What is the exact membership of the universal D-structure language generated by $\mathcal{R}$? The conditions listed above permit every rule of the form $X^1 \to X(Y^m)$ or $X^1 \to (Y^m)X$. From this it is easily seen that the universal base language generated is the regular language $V_T{}^*$ (i. e. the set of all strings of elements from the terminal symbol vocabulary V_T). Only if we maintained Centrality without permitting empty categories would $V_T{}^*$ not be generated.[11] But Universal Grammar (as conceived by Stowell, Chomsky, and others) does not, of course, exclude empty categories, so we conclude that the D-structure language generated by the universal X-bar system is $V_T{}^*$.

---

[9] This is true because CF PS rules and local subtrees are in one-to-one correspondence: a set of n rules generates a set of trees in which there are exactly n distinct local trees.

[10] Under this assumption, Chomsky's claim that 'X-bar theory permits only a finite class of possible base systems' (1981:11), reiterated in subsequent works (e.g. Chomsky 1986, fn. 3), becomes trivially true: there is only one X-bar system. As a claim about CFGs obeying the X-bar principles, it is false, unless a fixed length limit is placed on daughter sequences (e.g. by adoption of Kayne's suggestion about binary branching), in which case it is again true but trivial, since regardless of X-bar principles, there are only finitely many distinct rewriting systems with vocabulary V and maximum rule length k; see Pullum 1983 for discussion.

[11] The reason for this exception is that Centrality would allow only one single-preterminal structure; i.e., if C^m were the initial symbol, the only admissible structure containing a single projection would be one with C^0, and one-word sentences containing other categories could not be described.

Hence Stowell's claim reduces to the claim that a natural language over vocabulary V_T is a subset of V_T^*. But this is simply the definition of 'language' in the formal language theory sense. Therefore, the content of Stowell's theory is to be found not in the universal base component but in other components of the grammar. From Stowell's discussion, it appears that these components include at least the following: parochial conditions on theta-grids (p. 81), stylistic movement rules (p. 107), conditions on Case assignment (p. 113), constraints on 'instantiation of the adjacency condition' (p. 113), filters (p. 116), rules of 'restructuring' (p. 115), rules of 'absorption' (p. 119), structure-preserving movement (Move Alpha; p. 126), rules of insertion (p. 127), rules of adjunction (p. 136), and word formation rules (pp. 296ff). The design of some such components of the theory may, of course, incorporate universal and thus potentially restrictive elements, but as far as we can discern, the possibility of parochial variation in all of the above elements of the theory is left open.

One way of enriching the content of X-bar theory would be to add further constraints until a properly restricted set of D-structure strings (smaller than V^*) is arrived at. However, it seems highly unlikely that any of the rules Stowell allows can be ruled out by any universal principle. After all, these rules represent the extremely simple case of an endocentric construction A^1 formed by a head A and an optional phrasal complement B^m, e.g. a verb-object construction. Given Optionality and Succession, B^m can be rewritten as B^{m-1}, which in turn can be rewritten as B^{m-2}, and so on until we arrive at B^1. At this point we can rewrite B^1 as B and introduce the next preterminal as C^m, and so on until any desired form analyzable by the preterminal sequence $ABC\ldots$ is derived.

Notice that this argument does not really hinge on the principles of Optionality and Succession: it will go through as long as there are no principles that exclude rules of the form $X^n \to X^{n-1}$. In fact, such rules are a special case of the general principle of what Hellan has called LEVEL PIED-PIPING: 'If C^m and C^n are both defined, and $m > n$, then any occurrence of an expression belonging to C^n is also an occurrence of an expression belonging to C^m' (Hellan 1980:65). We regard it as quite clear that Universal Grammar cannot be formulated to exclude such rules if anything recognizable as X-bar theory is to be maintained.

One final observation is that there can be no doubt that the base determined by Stowell's revival of the UBH is sufficient to allow arbitrary recursively enumerable languages to be derived from it by appropriate transformations. According to Peters (1970:37–38), 'any phrase structure grammar is a universal base if it permits recursive embedding of one Sentence in another (not necessarily in a self-embedded manner) in such a fashion that every formative of the language can be introduced into trees of arbitrarily deep embedding.' Stowell's universal base clearly satisfies this description. Whether a variant of the Peters and Ritchie theorem (that all recursively enumerable languages have transformational grammars) actually goes through for specific current versions of transformational grammar depends on what is achievable through combinations of the various movements, adjunctions, insertions, and deletions that transformations are permitted to carry out. Deletions are the most critical, and we are not aware of any recent (post-1981) studies that fully clarify the status of deletion transformations in current theories; we assume that deletion rules (or something tantamount to them) are still permitted.

In sum, we believe that the effort to resuscitate the UBH by making X-bar theory a universal set of constraints on D-structure representations is a mistaken one. The only way X-bar theory can contribute to the problem of selecting among competing analyses is by allowing it to restrict the rules of grammar themselves. It does not matter whether we think of the rules in the traditional way as rules of a CFG or as constraints licensing well-formed local trees; both of these interpretations are compatible with the view that phrase structure admits of cross-linguistic variation. This view has never been subject

to a substantive (as opposed to programmatic) challenge. The idea that universal X-bar principles and constraints on lexical entries could conspire to produce (for example) the complexity seen in the structure of the English noun phrase as analyzed by Jackendoff (1977) or Hellan (1980) is one that no one has ever attempted to make plausible; yet this is what Stowell's program would entail.

3.2. X-BAR THEORY AS A CONSTRAINT ON PHRASE STRUCTURE RULES. If X-bar theory is viewed as constraining the set of parochial rules or licensing conditions, a number of questions arise. Does X-bar theory permit the description of every language describable by CFGs? Does it permit the assignment of any structure that could be assigned by some CFG?

Informally, the results to be presented below can be summarized as follows. As long as we permit 'empty categories', the constraints discussed in section 2 do not affect the generative power of CFGs at all. If 'empty categories' are disallowed, the constraints do decrease the descriptive power of CFGs, but the resulting family of languages is not formally coherent: appealing mathematical properties of the family of context-free languages (such as closure under various operations) are lost, while less desirable properties (such as undecidability of certain predicates) are generally retained.[12]

The interpretation of these results is somewhat obscured by the nature of the relation that holds between lexical categories (preterminals) and lexical entries (terminals). The traditional category system of linguistics does not provide a partitioning of the set of words: on the one hand, the same word can belong to more than one category, and on the other hand 'particles' and other 'function-words' are sometimes treated as syncategorematic (belonging to no syntactic category). In order to increase the transparency of the relationship between words and lexical categories it is convenient to assume that each particle has its own (one-member) lexical category; and in order to avoid the issue of overlap between categories, at some points below we will do what is often done in formal language theory and discuss results stated (in effect) in terms of a very broad lexical category that encompasses every lexical entry.

Recall the definition of a CFG: if G is a CFG, then $G = (V_N, V_T, P, S)$, where V_N and V_T are disjoint finite sets, S is a distinguished member of V_N, and P is a subset of $V_N \times (V_N \cup V_T)^*$. Lexicality can be interpreted as the claim that members of V_N have the form X^k, where X is a preterminal (lexical category) and $k \in N$, and we can define the set of maximal projections as follows:

(10) $V_M = \{X^i \mid (X^j \in V_N \ \wedge \ j \geq i) \quad \Rightarrow \quad j = i\}$

We now define the notion STANDARD X-BAR GRAMMAR (SXBG) as a grammar observing Lexicality, Maximality, and Succession, and define an OPTIONALITY-OBSERVING STANDARD X-BAR GRAMMAR (OSXBG) as an SXBG that additionally observes Optionality.

(11) Definition: a Standard X-bar Grammar (SXBG) is a Lexicality-observing CFG in which the rules have the form $X^n \rightarrow Y X^{n-1} Z$, where $Y, Z \in V_M^*$.

(12) An SXBG is an Optionality-observing Standard X-bar Grammar (OSXBG) iff for every rule $X^n \rightarrow Y X^{n-1} Z$, the grammar also contains all rules of the form $X^n \rightarrow Y' X^{n-1} Z'$ such that Y' and Z' are derivable by deleting zero or more symbols from Y and Z respectively.

---

[12]For a proof that the family of languages describable by (Optionality-observing) X-bar grammars is not closed under union, intersection, complementation, product, Kleene closure, substitution, homomorphism, gsm-mapping, or operations with regular sets, see Kornai (1985, Theorem 2.2). See Kornai (1982, Theorem 2.15) for a proof that it is undecidable whether the intersection of two languages generated by X-bar grammars is trivial (i.e. whether the two languages share any strings other than the lexical head of the initial symbol).

The definitions do not allow 'empty categories', because there can be no rules in which the right hand side is null. When we want to allow the empty string e to appear as the right hand side of a rule, we will introduce the possibility separately.

We now state a theorem entailing that Uniformity cannot have any effect on languages generated by SXBGs.

(13) Theorem: For every SXBG (resp. OSXBG) there exists an equivalent Uniformity-observing SXBG (resp. OSXBG) generating the same language and satisfying $V_M = \{X^1 \mid X \in V_T\}$.

This theorem, which is proved in Kornai (1985:526ff), entails that every SXBG can be converted into one that uses only zero-bar and one-bar categories (a 'one-bar normal form'), and thus Uniformity (and the number of permitted bar-levels) has no effect on the class of languages describable by SXBGs. In order to gain a better understanding of this result, we will give a full characterization of the languages that can be generated by some SXBG using only one lexical category.

(14) Theorem: If an SXBG G generates the language L over a one-symbol vocabulary of preterminals, j is the length in symbols of a sentence in L iff

$$j = 1 + (k_1 \cdot n_1) + (k_2 \cdot n_2) \ldots + (k_s \cdot n_s)$$

where each n_i is a natural number and $k_1 \ldots k_s$ are constants determined by the rules of G.

Keep in mind that our terminal symbols here are like the linguist's lexical categories. It follows from the theorem that over a one-symbol preterminal vocabulary $\{\sigma\}$, no languages containing a finite set of lexical category sequences can be generated by an SXBG, except for the trivial language $\{\sigma\}$, containing the single string in which σ appears just once. Thus if there is only one lexical category in an SXBG, the language generated either is infinite or contains a single one-category construction type.

From this it follows that not every regular preterminal language can be described by an SXBG (since every finite language is a regular language) and therefore not every CF preterminal language can be described by an SXBG (since every regular language is CF). For example, the preterminal language $\{\sigma\sigma\}$ is not generated by any SXBG. And from this, it is easy to see that the SXBG languages are not closed under renaming of preterminals, since the language $\{\sigma\tau\}$ can be generated, but the result of renaming τ as σ in this language cannot be generated.

These results depend crucially on the prohibition against e-rules. As soon as we permit zero preterminals (rules of the form $X^1 \to e$), EVERY CF preterminal language can be described by SXBGs. We now state and prove this result.

(15) Definition: a Standard X-bar Grammar with e-rules (SXBG^e) is a CFG in which rules are either of the form $X^n \to Y X^{n-1} Z$, where $Y, Z \in V_M{}^*$, or of the form $X^0 \to e$.

(16) Theorem: Given a CFG $G = (V_N, V_T, P, S)$, we can construct an SXBG^e (in fact, a Peripherality-respecting SXBG^e) generating the same language.

Proof: We will show that there is an SXBG $G' = \langle V_N{}', V_T{}', P_1 \cup P_0 \cup P_e, S^1 \rangle$ that generates the same language as G. Let $V_T{}' = V_T \cup V_N$ (i. e. G' generates a language over a vocabulary containing all the terminal and nonterminal symbols of G). Let $V_N{}' = \{X^1 \mid X \in V_T{}'\}$ (i. e. the nonterminals for G' are made from the terminals by adding one bar). Let $P_1 = \{X^1 \rightarrow X\,W^1 \mid X \rightarrow W \in P\}$, where the string W^1 results from W by adding a bar to each symbol in W. (Recall that W will be entirely composed of nonterminals or preterminals from V_N. The rules of G' thus generate a string XY^1Z^1 wherever G has a rule $X \rightarrow YZ$.) Let $P_0 = \{X^1 \rightarrow X \mid X \in V_T\}$ (i. e. for every X in the terminal vocabulary of G, the rules of G' include $X^1 \rightarrow X$). Let P_e contain the rule $A' \rightarrow e$ for each new preterminal A' in G'. (Note that such new preterminals correspond to nonterminals in G, so this allows for erasure of nonterminals from G wherever they turn up.)

By definition, G' is a Uniform (1-bar) CFG satisfying Lexicality, Succession, Maximality, Centrality, and Peripherality. Let us denote the language generated by G and G' by L and L' respectively. We know that $L \subset L'$, because if we derive some string w in L using the rules in P, the parallel derivation using rules from P_1 will result in a string w' to which we can apply the rules in P_0. Since the extraneous preterminals in V_N will be realized as zero by the rules in P_e, the string resulting from such a derivation is w.

To prove $L' \subset L$, take a string of symbols W in L'. All that has to be shown is that omitting the symbols in V_N from W gives us a string in L. Let us define W' as the string containing A^1 at each position where W contained some $A \in V_T$, and α at each position where W contained $\alpha \in V_N$. Since elements of V_T can only result from applications of rules in P_0, W' is a sentential form over G'. Moreover, W' can be derived from S^1 according to P_1. Therefore a parallel derivation can be constructed using G. The fact that this derivation results in a string in L can be seen from the observation that elements of V_N are terminals in G' and thus cannot be rewritten by rules of G'. $\square$

As is customary in mathematical linguistics, we give in the above proof a simple and direct way of constructing an equivalent grammar from another given one, in order to show as briefly as possible that an equivalent grammar always exists. The fact that the construction may yield an ugly or linguistically inappropriate grammar is beside the point. The grammar that our procedure constructs will not be the only one that exists; in fact, there will always be infinitely many others. This means that the possibility of some elegant and insightful X-bar grammar corresponding to some arbitrary CFG can never be dismissed out of hand; our proof shows that every CFG has a non-empty class of equivalent SXBGs that must be considered. And this shows that maintaining the SXBG conditions commits a linguist to nothing at all as regards limits on what is describable by CFGs.

The theorem in (16) depends crucially on the fact that Optionality is not enforced — i. e. that obligatory non-heads can be introduced by the rules of G'. There are context-free languages that cannot be described by any Optionality-observing SXBG. One example is the familiar language $\{a^n b^n \mid n \geq 1\}$. In an Optionality-observing grammar, it is never possible to guarantee in some language or construction type that (say) a certain number of nouns will be followed by exactly the same number of verbs.

Provided that all other SXB conditions are met, Optionality will decrease the descriptive power of the system even if e-rules are not permitted. For instance, $\{a^{2n+1} \mid n \geq 0\}$ can be generated by an SXBG, but not by an SXBG observing Optionality. Of all the X-bar conditions, Optionality is the one with the most effect on descriptive power of grammars.

4. THE THEORY OF HEADS. Historically, Chomsky's (1970) introduction of the X-bar theory seems to be a response to Lyons' (1968:235) observation that CFGs are inadequate for expressing the relation

that holds between endocentric constructions and their heads. We argue here that subsequent work on X-bar theory has concentrated too much on bars and bar-levels, losing sight of the central idea (due originally to Harris 1951) of overtly expressing the relationship between constructions and their heads.

In this section we will restate the basic content of X-bar theory without mention of bar-level. The primitive element in our account is a partial function defined on the nonterminal vocabulary; intuitively, it corresponds to the notion 'labels the head daughter of', a form of words which we will sometimes abbreviate as 'head-label of'. This function is bijective; that is, we ensure that if α labels the head daughter of β then α labels only the head of β and nothing else labels the head of β; maximal projections are not in the domain of the function (they do not label the head daughter of anything, because they are never heads), and preterminals are not in its range (nothing labels the head daughter of a preterminal, because preterminals do not have heads).

4.1. HEADEDNESS WITH STRICT SUCCESSION. A set X on which a function $f : X \mapsto X$ or a relation $R \in X \times X$ is defined will be said to be ENDOWED WITH f or with R. We will call a partial function $f : X \mapsto X'$ INVERTIBLE iff there is a function $f^{-1} : X' \mapsto X$ such that for all $x \in X$, $f(f^{-1}(x)) = x$ holds if f^{-1} is defined and $f^{-1}(f(x)) = x$ holds if f is defined; and we will call a partial function ACYCLIC iff there are no $x \in X$ and $n > 1$ such that $f^n(x) = x$, where f^n means n iterated applications of f, thus e.g. $f^2(x) = f(f(x))$.

Let V be a finite set and $h : V \mapsto V$ an invertible acyclic partial function. We define V_P, the set of preterminals, as those elements of V that have no image under h:

(17) $V_P = \{\alpha \in V \mid h(\alpha)$ is undefined$\}$

We define V_M, the set of maximal projections, as those elements of V that are not the image of anything under h:

(18) $V_M = \{\alpha \in V \mid h^{-1}(\alpha)$ is undefined$\}$

Using such a set V and function h, we can express the content of X-bar theory directly on trees as follows.

We define trees in a standard way, as a set of nodes endowed with the binary relation M 'mother of', and denote the label of a node a by $L(a)$. A tree Δ with labeling L will be called X-BAR COMPATIBLE iff the set of node labels $L(\Delta) = V$ is endowed with an invertible acyclic partial function h as described above, and every local subtree δ (i.e. a mother node m and its daughters) of Δ satisfies the conditions in (19).

(19)a. δ has a daughter d such that $L(d) = h(L(m))$;

    b. for every daughter d' other than d, $h(L(d'))$ is undefined;

    c. d is either the leftmost or rightmost daughter of m.

Conditions a, b, and c in (19) correspond to Succession, Maximality, and Peripherality, while Lexicality is built into the structure of h. Accordingly, we will say that a set of trees $\mathcal{T}$ SATISFIES X-BAR THEORY iff there is a set of labels V endowed with an invertible acyclic partial function h such that each tree $\Delta \in \mathcal{T}$ is labelled from V and is X-bar compatible, and further, $\mathcal{T}$ is closed under the operation of erasing non-head daughters and the subtrees they dominate (Optionality).[13]

---

[13]To say that $\mathcal{T}$ is closed under erasing of non-head daughters is to say that every tree formed by erasing non-head daughters from a tree in $\mathcal{T}$ is itself in $\mathcal{T}$.

This formulation describes exactly the same set of labelled trees as the SXB definitions given in the previous section, but without any reference to bars or bar-level.[14] Therefore, bar-levels are not an intrinsic part of the internal structure of category labels, and their use is only an implementation decision. It may be mnemonically helpful that in standard notations N (= N^0) labels the head of N^1 and N^1 labels the head of N^2, and similarly I (= I^0) labels the head of I^1 and I^1 labels the head of I^2. But the same description could be given with categories named in ways familiar from older work: we could stipulate that Noun labels the head of Nom and Nom labels the head of NP, or that Aux labels the head of Predicate-phrase and Predicate-phrase labels the head of Sentence. The content of the X-bar claims lies in the function 'labels the head of', not in any notation for category labels designed to make this perspicuous. The above reformulation brings into sharp relief the fact that (as Halitsky 1975 argued), X-bar theory should be regarded as a notion rather than a notation.

4.2. ADJUNCTION STRUCTURES AND WEAK SUCCESSION. In many works that make use of some X-bar principles, it is assumed that there is a class of exceptions to Succession in which a node labeled X^n may lack a daughter labeled X^{n-1}, having a daughter labeled X^n instead. The analysis of attributive adjective phrase modification that has $A^2 N^1$ immediately dominated by N^1 (mentioned in section 1.2) is a typical example. In such cases, the immediately dominated X^n is generally regarded as the head of its mother X^n.

A simple modification of our reconstruction of X-bar theory allows for this. We simply assume that maximal projections can label the head of a mother with the same label, but cannot label any other heads. Formally, instead of requiring that h be an acyclic function, we require that it be a particular type of partial ordering. A function is simply a relation R that respects the uniqueness condition $\forall x \forall y \forall z[(xRy \wedge xRz) \Rightarrow y = z]$. We replace the function h described in the previous section by a binary relation H meeting the weakened condition we can call FIDELITY:

(20) A binary relation H meets the Fidelity condition iff $\forall x \forall y \forall z[(xHy \wedge xHz) \Rightarrow (x = y \vee x = z)]$.

This says that a label x cannot bear the relation H to two different labels (distinct from itself). We also replace the definition of acyclic function given in the previous section by a definition applicable to binary relations: we call a relation H acyclic iff its positive transitive closure H^+ is irreflexive. That is:

(21) A binary relation H is acyclic iff $\forall x \forall y[(xH^*y \wedge yH^*x) \Rightarrow x = y]$.

The revised X-bar theory is as follows. Let V be a finite set and H a binary acyclic relation on V satisfying Fidelity. We define V_T, the set of preterminals, as those elements of V to which no distinct element bears the relation H:

(22) $V_T = \{\alpha \in V \mid \exists \beta[\beta H \alpha] \Rightarrow \beta = \alpha\}$

We define V_M, the set of maximal projections, as those elements of V which bear the relation H only to themselves:

---

[14] If it were considered desirable to add a condition corresponding to Uniformity, this could be done by requiring the existence of a number m such that in each tree in $\mathcal{T}$, each node labeled by a maximal projection α has a descendant with label $h^m(\alpha)$ and no node α has a descendant with label $h^{m+1}(\alpha)$. In other words, the length of all paths from maximal projections via head daughters to preterminals is fixed at m links. This makes it very clear, however, that Uniformity is a rather isolated and stipulative condition.

(23) $V_M = \{\alpha \in V \mid \exists\beta[\alpha H\beta] \Rightarrow \beta = \alpha\}$

A tree Δ with labeling L will be called X-BAR(WS)-COMPATIBLE iff the set of node labels $L(\Delta) = V$ is endowed with an acyclic binary relation H observing Fidelity as described above, and every local subtree δ (i.e. a mother node m and its daughters) of Δ satisfies the conditions in (24).

(24)a. δ has a daughter d such that $L(d)HL(m)$;

b. for every daughter d' other than d, $\alpha HL(d') \Rightarrow \alpha = L(d')$;

c. d is either the leftmost or rightmost daughter of m.

4.3. MAXIMAL PROJECTION COMMAND DOMAINS. The theory of heads developed here can be used to clarify a rather subtle point concerning the command relations that obtain in adjunction structures. In the recent literature the original notion of c-command (the 'superior to' relation of Reinhart 1974:94) is replaced by a notion called 'm-command' by Chomsky (1986) and called MAX-command by Barker and Pullum (1990). Published definitions have varied, but basically a node x is said to MAX-command a node y iff y is a subconstituent of the smallest constituent properly containing x and labeled by a maximal projection (see Barker and Pullum 1990 for a more careful formulation). The intent is to define command in such a way that when a phrase α is adjoined to a phrase X^2 to yield a structure like (25), the lexical head X^0 of X^2 MAX-commands the α (as it does any number of phrases similarly adjoined at the X^2 level).

(25)

```
              X²
           /      \
        α            X¹
                     |
                     X⁰
```
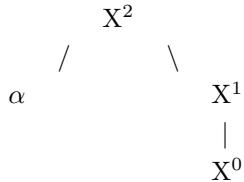
Under the earlier, stricter conception of Succession, such structures cannot be generated in the base at all: they must be the result of later (adjunction) rules. Under Weak Succession, these structures can also be base-generated. Either way, the proper interpretation of MAX-command will be contingent on the variant of Succession we adopt.

Under strict Succession, the smallest constituent properly containing x and labeled by a member of V_M is determined by the following procedure (where L is the labeling function on nodes and M is the partial function 'mother-of'). If $h(L(M(x)))$ is undefined (which means the mother of x is labeled by a maximal projection), then $M(x)$ is the root of the desired constituent, and x MAX-commands y iff $M(x)$ dominates y. Otherwise, consider whether $h(L(M^2(x)))$, and so on. Since trees are finite, either there is some k such that $h(L(M^k(x)))$ is undefined, in which case x MAX-commands y iff $M^k(x)$ dominates y, or the tree contains no node labeled by a maximal projection, in which case the question is trivial.[15]

If we adopt Weak Succession, matters are somewhat different. To determine whether a node x MAX-commands a node y we have to ask whether $L(M(x))$ bears the H relation to $L(M^2(x))$. If not, then x MAX-commands y iff $M(x)$ dominates y. Otherwise, consider whether $L(M^2(x))$ bears the H relation to $L(M^3(x))$, and so on. Since trees are finite, either there is some k such that the pair

---

[15]Under the assumptions of Barker and Pullum (1990), x trivially MAX-commands y in such a case because of the lack of any maximal projection-labeled node dominating x but not y.

$\langle L(M^k(x)), L(M^{k+1}(x)) \rangle$ is not in H, in which case x MAX-commands y iff $M^k(x)$ dominates y, or the tree contains no node labeled by a maximal projection, in which case the question is again trivial.

Now consider the adjunction configuration (25). Notice that it is not true in this structure that the smallest constituent properly containing x and labeled by a maximal projection also contains y, so the normal intuitive concept of MAX-command does not work here. The correct result regarding what x dominates is obtained by tracing the path of heads from $M(x)$ upward until it terminates, rather than looking for the lowest maximal projection dominating x.

The procedures decribed above for the cases of strict and Weak Succession are based exactly on this idea of path tracing, and it is easy to see that in the structure (26) the node labeled X^0 MAX-commands not only α but β as well.[16]

(26)

$$
\begin{array}{c}
X^2 \\
\diagup \qquad \diagdown \\
\beta \qquad\qquad X^2 \\
\qquad \diagup \qquad \diagdown \\
\qquad \alpha \qquad\qquad X^1 \\
\qquad\qquad\qquad | \\
\qquad\qquad\qquad X^0
\end{array}
$$
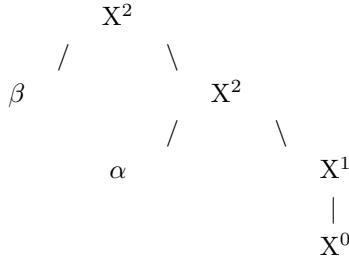
5. CONCLUSIONS. We will first summarize what we have done in this paper, and then proceed to some general conclusions. We have expressed the content of a fairly stringent form of X-bar theory in terms of six conditions: Lexicality, Succession, Maximality, Uniformity, Centrality, and Optionality on phrase structure rules. We have shown that these conditions, while intuitively restrictive, actually have little or no effect on what languages can be described. Constraints like Maximality and Uniformity have no implications at all for description of languages; hence linguists can (and regularly do) assume that all lexical categories project to the same level (usually X^2) and all non-heads are maximal projections, but this does not impose any real constraints on what syntactic phenomena can be described. Optionality does impose constraints on description of languages ($\{a^n b^n \mid n \geq 0\}$ being an example of a language it excludes), but is hardly respected at all in actual descriptive work; devices such as strict subcategorization, predication requirements, etc. are used to evade it by rendering non-head subconstituents obligatory.

We have argued that there is nonetheless some content in X-bar theory that deserves to be recognized and elucidated. The key concept is headedness and its connection with a mapping (or relation) on the nonterminal vocabulary of a grammar that constrains the distribution of categories in root-to-frontier paths in local trees (or equivalently, in phrase structure rule sets). The representation of bar-levels in the structure of categories, as implied by Lexicality, is only one way of capturing this. While the X-bar conditions can be made precise in terms of internal structure assigned to syntactic category labels (as in Gazdar et al. 1988), we have argued here that an elegant reformulation of the essential content of X-bar theory can be developed on the basis of headedness without reference to the concept of bar-level. Whether a category is labeled C′ or S′ or S[COMP THAT] is entirely unimportant; what is important is whether the category in question must have as its head the complementizer node (as in current transformational work) or the clausal node (as in Gazdar et al. 1985). It is possible for the form of category labels themselves to bear information from which headedness can be deduced, but it is not necessary.

---

[16]The linguistic predictions obtained from this result depend, of course, on the precise role that command plays in mediating between the theoretically derived phrase structure and more directly observable facts (such as pronoun-antecedent relationships).

The set of maximal projections and the set of preterminal categories can be defined in terms of the headedness function just as easily as in terms of bar-level: the maximal projections are those categories that never label a head, and the lexical categories are those that never label a node that has a head. The notion 'MAX-command' can be captured appropriately whether or not adjunction structures are assumed.

We conclude with some speculative remarks about methodology in syntax. Although there is a useful set of ideas embedded in the X-bar syntax literature, X-bar theory clearly falls short of being a genuinely restrictive theory of natural language phrase structure. We are inclined to think that a mistaken research strategy may be at fault. The problem may lie with the strategy of seeking theories that, in Chomsky's words (1981:3), 'are fairly intricate in their internal structure, so that when a small change is introduced there are often consequences throughout this range of phenomena.' Theories of the sort Chomsky characterizes are peculiarly unstable, in the sense that minor modifications in their definition lead to major (and almost always unpredictable) changes in the class of objects described. X-bar grammars are a good example of this. We will cite three examples of their instability.

First, the introduction of e-rules to SXBGs changes their descriptive abilities radically (whereas the addition of e-rules to e-free CFGs makes only the trivial addition that languages containing the empty sentence become generable); adding e-rules to SXBGs makes them equivalent to ordinary CFGs, in spite of the apparently limiting constraints on rule format.

Second, the SXBG languages are not closed under finite state transductions, so that morphological substitutions like *du* for *de le* in French could in principle have unpredictable effects on the type of language generated. Simply altering the shape of a word, or replacing one word by another in some syntactic construction in a language, could change the language from SXB to non-SXB.

Third, the class of languages generated by SXBG grammars is not closed under union, so that even the relatively minor change of allowing more than one start symbol (i.e. root label), as suggested by Brame (1981), increases the class of describable languages considerably.

We believe that the pursuit of theories in which minor changes can have radical effects is likely to prove unprofitable. Conceivably such theories will be shown to have some of the usefulness that is widely claimed for them in connection with the problem of language acquisition, but at present the question seems entirely open.

If such theories are to be seriously developed, it seems to us essential that they should be built step by step. Instead of proposing highly specific (and in all probability language-particular) constraints whose effects are formally unexplored, linguists should adopt a more conservative stance, and avoid theorizing in ways that render the descriptive power of their theory a mystery. The full ramifications of the effect of introducing a particular condition should be explored with care before it is assumed as a part of linguistic theory.

The explorations that are called for will include investigations into the little-studied issue of strong generative capacity: the study of what structures can be assigned to sentences by different types of grammars, what interpretations such structures will support, and what claims particular types of grammar can be regarded as making about the structures of sentences. Clarification of some such questions has been the aim of the present examination of X-bar theory.

REFERENCES

Abney, Stephen. 1987. The English noun phrase in its sentential aspect. Cambridge MA: MIT dissertation.

Baker, C. L. 1978. Introduction to generative-transformational syntax. Englewood Cliffs, NJ: Prentice-Hall.

Barker, Chris, and Geoffrey K. Pullum. 1990. A theory of command relations. Linguistics and Philosophy (in press).

Brame, Michael. 1981. Trace theory with filters vs. lexically based syntax without. Linguistic Inquiry 12.275–293.

Bresnan, Joan W. 1976. Transformations and categories in syntax. Basic problems in methodology and linguistics, ed. by Ronald Butts and Jaakko Hintikka, 261–282. Dordrecht: D. Reidel.

Chomsky, Noam. 1957. Syntactic structures. The Hague: Mouton.

Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge, Massachusetts: MIT Press.

Chomsky, Noam. 1970. Remarks on nominalization. Readings in English transformational grammar, ed. by Roderick A. Jacobs and Peter S. Rosenbaum, 184–221. Waltham, Massachusetts: Ginn.

Chomsky, Noam. 1981. Lectures on government and binding. Dordrecht: Foris Publications.

Chomsky, Noam. 1986. Barriers. Cambridge, Massachusetts: MIT Press.

Dougherty, Ray. 1968. A transformational grammar of coordinate conjoined structures. Cambridge MA: MIT dissertation.

Emonds, Joseph E. 1976. A transformational approach to English syntax. New York: Academic Press.

Gazdar, Gerald. 1982. Phrase structure grammar. The nature of syntactic representation, ed. by Pauline Jacobson and Geoffrey K. Pullum, 131–186. Dordrecht: D. Reidel.

Gazdar, Gerald; Ewan Klein; Geoffrey K. Pullum; and Ivan A. Sag. 1985. Generalized phrase structure grammar. Oxford: Basil Blackwell; Cambridge, Massachusetts: Harvard University Press.

Gazdar, Gerald, and Geoffrey K. Pullum. 1982. Generalized phrase structure grammar: a theoretical synopsis. Bloomington, Indiana: Indiana University Linguistics Club.

Gazdar, Gerald; Geoffrey K. Pullum; Robert Carpenter; Ewan Klein; Thomas E. Hukari; and Robert D. Levine. 1988. Category structures. Computational Linguistics 14.1–19. Also technical report no. CSLI–87–102, Center for the Study of Language and Information, Stanford University, 1987.

Gazdar, Gerald; Geoffrey K. Pullum; and Ivan A. Sag. 1982. Auxiliaries and related phenomena in a restrictive theory of grammar. Language 58, 591–638.

Grimshaw, Jane. 1981. Form, function, and the language acquisition device. The logical problem of language acquisition, ed. by C. L. Baker and John J. McCarthy, 165–182. Cambridge, Massachusetts: MIT Press.

Halitsky, David. 1975. Left branch S's and NP's in English: A bar notation analysis. Linguistic Analysis 1.279–296.

Harris, Zellig. 1951. Methods in structural linguistics. Chicago: Chicago University Press.

Hellan, Lars. 1980. Toward an integrated theory of noun phrases. Trondheim: University of Trondheim dissertation.

Hornstein, Norbert, and David Lightfoot. 1981. Introduction. Explanation in linguistics: The logical problem of language acquisition, ed. by Norbert Hornstein and David Lightfoot, 9–31.

22

London: Longman.

HUDSON, RICHARD A. 1984. Word grammar. Chicago: University of Chicago Press.

JACKENDOFF, RAY S. 1977. $\overline{\mathrm{X}}$ syntax: A study of phrase structure. Cambridge, Massachusetts: MIT Press.

KAYNE, RICHARD S. 1981. Unambiguous paths. Levels of syntactic representation, ed. by Jan Koster and Robert May, 143–183. Dordrecht: Foris.

KEAN, MARIE-LOUISE. 1978. Some features of features. Social Science Research Report R41, University of California, Irvine.

KORNAI, ANDRÁS. 1982. X-vonás nyelvtanok. (X-bar Grammars.) Budapest: Eötvös Loránd University dissertation.

KORNAI, ANDRÁS. 1983. "... some version of the X-bar theory" Unpublished manuscript, Institute of Computer Science, Hungarian Academy of Sciences

KORNAI, ANDRÁS. 1985. X-bar grammars. Algebra, combinatorics, and logic in computer science, ed. by J. Demetrovics, G. O. H. Katona, and A. Salomaa, 523–536. Amsterdam: North-Holland.

LANGENDOEN, D. TERENCE. 1975. Finite-state parsing of phrase-structure languages and the status of readjustment rules in grammar. Linguistic Inquiry 6.533–554.

LANGENDOEN, D. TERENCE. 1976. On the weak generative capacity of infinite grammars. CUNYForum 1.13–24. New York: Graduate Center, City University of New York.

LYONS, JOHN. 1968. Introduction to theoretical linguistics. Cambridge: Cambridge University Press.

ORE, OYSTEIN. 1963. Graphs and their uses. New Haven: Yale University Press.

PETERS, P. STANLEY. 1970. Why there are many 'universal' bases. Papers in Linguistics 2.27–43.

PINKER, STEVEN. 1984. Language learnability and language development. Cambridge, Massachusetts: Harvard University Press.

POLLARD, CARL J. 1984. Generalized phrase structure grammars, head grammars, and natural language. Stanford CA: Stanford University dissertation

POLLOCK, JEAN-YVES. 1989. Verb movement, universal grammar, and the structure of IP. Linguistic Inquiry 20.365–424.

PULLUM, GEOFFREY K. 1983. How many possible human languages are there? Linguistic Inquiry 14.447–467.

PULLUM, GEOFFREY K. 1985. Assuming some version of X-bar theory. CLS 21 part I: Papers from the general session at the twenty-first regional meeting, ed. by William H. Eilfort, Paul D. Kroeber, and Karen L. Peterson, 323–353. Chicago, Illinois: Chicago Linguistic Society.

RADFORD, ANDREW. 1981. Transformational syntax: A student's guide to Chomsky's extended standard theory. Cambridge: Cambridge University Press.

REINHART, TANYA. 1974. Syntax and coreference. NELS 5: Papers from the fifth annual meeting, Northeastern Linguistic Society, 92–105. Amherst, MA: Graduate Linguistic Student Association, University of Massachusetts.

RICHARDSON, JOHN F. 1984. Let $\overline{\mathrm{X}} = \overline{\mathrm{X}}$. CLS 20: Papers from the twentieth regional meeting, ed. by Jospeh Drogo, Veena Mishra, and David Testen, 321–333. Chicago, Illinois: Chicago Linguistic Society.

SAG, IVAN; GERALD GAZDAR; THOMAS WASOW; and STEPHEN WEISLER. 1985. Coordination and how to distinguish categories. Natural Language and Linguistic Theory 3, 117–171.

SELKIRK, ELISABETH O. 1977. Some remarks on noun phrase structure. Formal syntax, ed. by Peter W. Culicover, Thomas Wasow, and Adrian Akmajian, 285–316. New York: Academic Press.

Sells, Peter. 1985. Lectures on contemporary syntactic theories. CSLI Lecture Notes 3. Stanford: Center for the Study of Language and Information. Chicago: University of Chicago Press.

Starosta, Stanley. 1984. Lexicase and Japanese language processing. Tokyo: Musashino Electrical Communication Laboratories, Nippon Telegraph and Telephone Public Corporation.

Stowell, Timothy A. 1981. Origins of phrase structure. Cambridge MA: MIT dissertation.

Stuurman, Frits. 1985. Phrase structure theory in generative grammar. Dordrecht: Foris.

Verkuyl, H. 1981. Numerals and quantifiers in X-bar syntax and their semantic interpretation. Formal methods in the study of language (Amsterdam Mathematical Centre tracts), ed. by Jeroen Groenendijk, Theo Janssen, and Martin Stokhof, 567–599. Amsterdam: University of Amsterdam.

Williams, Edwin. 1975. Small clauses in English. Syntax and semantics 4, ed. by John Kimball, 249–273. New York: Academic Press.

ACKNOWLEDGEMENTS