12 Phrase Structure

NAOKI FUKUI

0 Introduction

That sentences and phrases in human language have abstract hierarchical structure, not merely sequences of words and formatives, is one of the fundamental discoveries of modern linguistics. Accordingly, any theory of human language must have a component/device that deals with its "phrase structure," regardless of the analyses it offers for other properties of language (such as transformations). In this sense, the theory of phrase structure is a kind of backbone for contemporary linguistic theory.

In earlier generative traditions, the properties of phrase structure were coded in terms of the formal mechanism called "phrase structure rules" of the following form, where α is a single symbol and φ, ψ, and χ are strings of symbols (χ non-null; φ and ψ possibly null):

(1) $\varphi\alpha\psi \rightarrow \varphi\chi\psi$

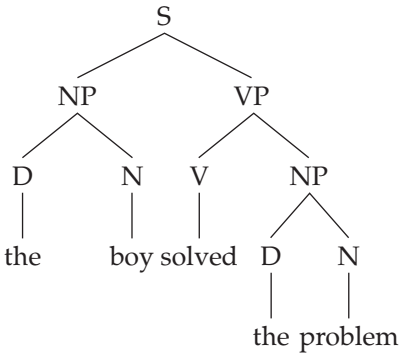
Phrase structure rules express the basic structural facts of the language in the form of "phrase markers" they generate,[1] with terminal strings drawn from the lexicon. In particular, phrase markers generated by phrase structure rules express three kinds of information about syntactic representations:

(2) i. the hierarchical grouping of the "constituents" of the structure (Dominance);
ii. the "type" of each constituent (Labeling);
iii. the left-to-right order (linear order) of the constituents (Precedence).

For example, the phrase marker (3), generated by the phrase structure rules in (4), indicates that the largest constituent, whose label is S (the designated initial symbol), is made up of a constituent NP (Noun Phrase) preceding the other constituent VP (Verb Phrase); that the NP consists of two constituents,

D(eterminer) and a N(oun), in this order; and that the VP is composed of V(erb) and NP (in this order), and so on:

(3)



(4) i. $S \rightarrow NP\ VP$
ii. $VP \rightarrow V\ NP$
iii. $NP \rightarrow D\ N$
iv. $D \rightarrow$ the
v. $N \rightarrow$ boy
vi. $N \rightarrow$ problem
vii. $V \rightarrow$ solved

Phrase structure rules of the kind represented by (4iv)–(4vii), which directly insert lexical items into appropriate places in the structure, were later abolished in favor of the lexicon with subcategorization features (Chomsky 1965). This separation of lexicon from the "computational system" (phrase structure rules) makes it possible to simplify the form of phrase structure rules for human language from the "context-sensitive" (1) to the "context-free" (5) (with φ, ψ necessarily null; other qualifications are the same):

(5) $\alpha \rightarrow \chi$

In (5), α is a single "non-terminal" symbol, and χ is either a non-null string of non-terminal symbols or the designated symbol "Δ," into which a lexical item is to be inserted in accordance with its subcategorization features (see Chomsky 1965 for details).

Thus, context-free phrase structure rules, coupled with the lexicon containing the information about idiosyncratic properties of each lexical item, were assumed in the "Standard Theory" of generative grammar (Chomsky 1965) to be responsible for expressing the properties of phrase structure. However, toward the end of the 1960s, it became apparent that certain important generalizations about the phrase structure of human language cannot be stated in terms of phrase structure rules alone. Recognition of the inadequacies of phrase structure rules, as we will see in the following section, led to the emergence

and development of the general theory of phrase structure, "X'-theory," which is a main topic of this chapter.

The organization of this chapter is as follows. Section 1 discusses the basic insights of X'-theory. The section provides a brief explanation as to how this has emerged as an attempt to overcome the deficiencies of phrase structure rules in capturing the basic properties of phrase structure of human language, and summarizes the development of X'-theory from its inception to the Barriers version (Chomsky 1986). Section 2 is concerned with the "post-Barriers" development of the theory of phrase structure, which can be characterized as minimizing the role of X'-theory as an independent principle of Universal Grammar (UG), while maintaining its basic insights, which led to the eventual elimination of X'-theory in the "Minimalist program" (Chomsky 1994). It should be mentioned that the historical overview of these sections is by no means meant to be comprehensive, and the remarks to be made in the presentation are rather selective and schematic. It also goes without saying that the overview benefits from hindsight. Section 3 deals with one of the current issues in the theory of phrase structure, namely, the role of "linear order," in general, and that of the "head parameter," in particular. This section takes up some of the most recent works on the issue of linear order, and examines their basic claims. Section 4 is a summary and conclusion.

As the discussion proceeds, I will occasionally touch on some of the issues of movement (transformations) as well. This is because the theory of phrase structure and the theory of movement have been progressing side by side in the history of generative grammar. Transformations are formal operations applying to linguistic representations constructed in accordance with the general principles of phrase structure. Thus, a substantive change in the theory of phrase structure necessarily has important implications for the theory of transformations.

Throughout the chapter, I will basically confine myself to the discussion of the development of X'-theory, with only scattered references to other approaches to phrase structure, such as categorical grammars (Lambek 1958; see also Wood 1993 and references there), generalized phrase structure grammar (Gazdar et al. 1985) and its various ramifications (head driven phrase structure grammar (Pollard and Sag 1994, for example)), lexical-functional grammar (Bresnan 1982b), etc. This is of course not to dismiss the other approaches, but mainly to keep the discussion coherent and to manageable proportions. In addition, there are also more substantive reasons. First of all, the empirical insights offered by X'-theory are to be captured by any theory of phrase structure, regardless of the difference in formalism. Second, at least given the current version of "X'-theory" (this name may no longer be appropriate, as we will see later), there do not seem to be, as far as the treatment of phrase structure is concerned, so many fundamental differences between "X'-theory" and the other approaches mentioned above. The differences, if any, seem to be only concerned with the way other properties of language (the property of "displacement," for instance) are handled in a given framework.

1 From "Remarks" to *Barriers*: Formulating and Enriching X′-Theory

The basic motivations for X′-theory come from the following two considerations:

(6) i. the notion of "possible phrase structure rules";
ii. cross-categorical generalizations.

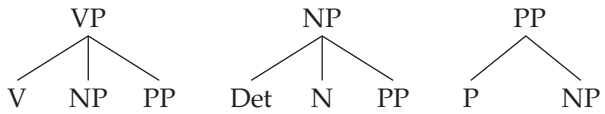
The first consideration has to do with what counts as "a possible phrase structure rule" in natural languages. It is observed that while phrase structure rules of the kind in (7) (cf. also the phrase structure rules in (4) above) are widely attested in natural languages, those represented in (8) are systematically excluded in any grammar of human language:

(7) VP → V (NP) (PP)
NP → (Det) N (PP)
PP → P (NP)

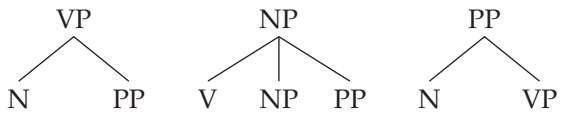
(8) VP → N (PP)
NP → V (NP) (PP)
PP → N (VP)

In other words, structures such as those in (9), which are generated by the phrase structure rules in (7), are permitted in human language, whereas structures like those in (10), generated by the phrase structure rules in (8), are systematically excluded in human language:

(9)



(10)



The reason for the impossibility of the phrase structure rules in (8) (and the corresponding structures in (10)) is intuitively clear. VP, for example, is a "Verb Phrase," rather than, say, a "Noun Phrase," and since it is a phrase of a verb, it must have a verb in it. However, the right-hand side of the phrase structure rule VP → N (PP) does not contain any verb. Hence the structure generated by such a phrase structure rule (i.e., the first structure in (10)) is ill formed. The same is true for the other phrase structure rules in (8) (and the corresponding structures in (10)).

In general, an "XP" cannot be a "phrase of X" if there is no X. Put another way, phrase structure in human language is "endocentric," in the sense that it is constructed based on a certain central element (called the "head" of a phrase), which determines the essential properties of the phrase, accompanied by other non-central elements, thus forming a larger structure. This is the right intuition, but, as pointed out by Lyons (1968), the theory of phrase structure grammar simply cannot capture this. Recall that in the general scheme of context-free phrase structure rules in (5), reproduced here as (11), the only formal requirements are that α is a single non-terminal symbol and χ is a non-null string of non-terminal symbols (or the designated symbol Δ):

(11) $\alpha \rightarrow \chi$

The phrase structure rules in (7) (which are attested in human language) and those in (8) (which are excluded in human language) are no different as far as the formal "definitions" of phrase structure rules are concerned. Thus, in each of the phrase structure rules in (8), the left-hand side is a single non-terminal symbol ("VP," "NP," and "PP"), and the right-hand side of the rule is a non-null string of nonterminal symbols ("N (PP)," "V (NP) (PP)," and "N (VP)"). These are all legitimate phrase structure rules, satisfying the formal definitions of (context-free) phrase structure rules, just like the phrase structure rules in (7), despite the fact that only the latter type of phrase structure rule is permitted and the former type is never allowed (at least, has never been attested) in human language. Phrase structure rules are too "permissive" as a theory of phrase structure in human language in that they generate phrase structures that are indeed never permitted in human language. We thus need some other mechanism which correctly captures the endocentricity of phrase structure that appears to be a fundamental property of human language.

The second major motivation for X'-theory is concerned with some observed parallelisms that exist across different categories. Historically, the discussion started out with the treatment of two types of nominal in English, as represented by the following examples:

(12) a. John's refusing the offer
b. the enemy's destroying the city

(13) a. John's refusal of the offer
b. the enemy's destruction of the city

Nominals of the type represented in (12) are called "gerundive nominals," whereas those shown in (13) are called "derived nominals." These two types of nominal were treated uniformly in terms of a "nominalization transformation," which derives nominals like, say, (12b) and (13b) from the same source,

namely, (the underlying form of) the sentence "the enemy destroyed the city" (see Lees 1960 for details).

Chomsky (1970), however, refutes this "Transformationalist Hypothesis," and argues that the theory of grammar should not allow a nominalization transformation (or any other transformation with similar expressive power) because it performs various operations that are never observed in any other well-argued cases of transformations. Thus, the alleged nominalization transformation (i) changes category types (it changes S to NP and V to N), (ii) introduces the preposition *of*, (iii) changes the morphological shape of the element (*destroy* is changed to *destruction; refuse* is changed to *refusal*, etc.), (iv) deletes all auxiliaries, and so on. These are the operations that other well-attested transformations never perform, and hence should not be allowed, Chomsky argues, if we are to aim at restricting the class of possible grammars.

In particular, Chomsky points out (i) that derived nominals are really "noun-like," not sharing various essential properties with sentences, and (ii) that the relationship between derived nominals and their sentential counterparts is rather unsystematic and sometimes unpredictable (see Chomsky 1970 for more arguments establishing these points). He then concludes that derived nominals should be handled in the lexicon, rather than in terms of transformations which deal with formal and systematic relationships between phrase structure trees. This proposal defines the "Lexicalist Hypothesis," which has become standard for the analysis of derived nominals in particular, and for the characterization of transformations in general.

Once we adopt the Lexicalist Hypothesis, however, an important problem immediately arises as to how to capture certain similarities and parallelisms holding between verb/noun and sentence/nominal pairs. More specifically, the strict subcategorization properties of a verb generally carry over to the corresponding noun, and the identical grammatical relations are observed in both sentences and the corresponding nominals (see Lees 1960 and Chomsky 1970 for detailed illustrations of these points; see also van Riemsdijk and Williams 1986 for a lucid summary). Under the Transformationalist Hypothesis, these parallelisms are captured by the nominalization transformation. With the elimination of such a transformation under the Lexicalist Hypothesis, we now have to seek an alternative way to express the parallelisms in the grammar.

Chomsky (1970) proposes that these parallelisms can be successfully captured if the internal structure of noun phrases is made to be sufficiently similar to that of sentences so that the strict subcategorization properties and grammatical relations can be stated in such a general form as to apply to both verbs/sentences and nouns/nominals. As a concrete means to express these cross-categorical generalizations, Chomsky introduces a preliminary version of X′-theory of the following kind (adapted from Chomsky 1970):

(14) a. $X' \rightarrow X \ldots$
b. $X'' \rightarrow$ [Spec, X'] X'

The "X" in (14) is a variable ranging over the class of lexical categories N(ouns), V(erbs), A(djectives), and (perhaps) P(repositions). The symbol X′ (called "X bar," although, for typographical reasons, it is common to use primes rather than bars) stands for a constituent (phrase) containing X as its "head" (the central and essential element of the phrase), as well as those elements appearing in the place indicated by ". . ." in (14a), the elements called the "complement" of X. The schema (14b) introduces a still larger phrase X″ (called "X double bar") containing X′ and pre-head elements associated with X′, called the "specifier" (Spec) of X′ (notated as [Spec, X′]).[2] Examples of specifiers include, according to Chomsky, determiners as [Spec, N′], auxiliary elements as [Spec, V′], comparative structures and elements like *very* as [Spec, A′], etc. X′ and X″, which share the basic properties of the head X, are called "projections" of X, with the latter (X″) referred to as the "maximal projection" of X (since it does not project any further).

The X′-schemata in (14) are proposed as a principle of UG on phrase structure, and express the manner in which phrases are constructed in human language. Note that given the X′-schemata, the problem concerning the "possible phrase structure rules" in human language discussed above is immediately resolved. That is, the "endocentricity" of phrases in human language is directly encoded in X′-theory as the generalization that phrases are all projections of their heads. Thus, the non-existing phrase structure rules in (8) are excluded on principled grounds as rules generating the illegitimate structures in (10), which contain phrases lacking the proper heads, in violation of X′-theory.

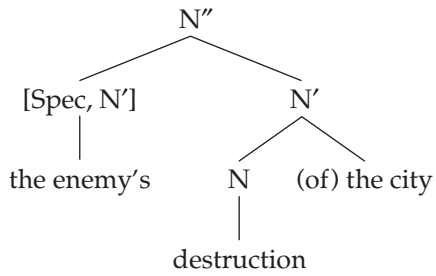
With respect to the problem of expressing cross-categorical parallelisms, X′-theory provides a generalized structure by which we can uniformly express basic grammatical relations. Thus, the notion of "object-of" X can be stated as an NP that is immediately dominated by X′, and the notion of "subject-of" X can be expressed as an NP that is immediately dominated by X″, where X in both cases ranges over V, N, etc. Likewise, the strict subcategorization properties of, say, verbs and nouns are stated uniformly in terms of the general X′-scheme. For example, if an X (a verb or a noun) has a subcategorization frame +[_PP], then the PP is realized as the complement of X (the verb or noun).

However, the X′-theoretic generalizations were not complete at this stage of the development of the theory. This is because sentences did not quite fit into the general X′-scheme and were introduced by the following phrase structure rule, which does not really conform to X′-theory (see Chomsky 1970):

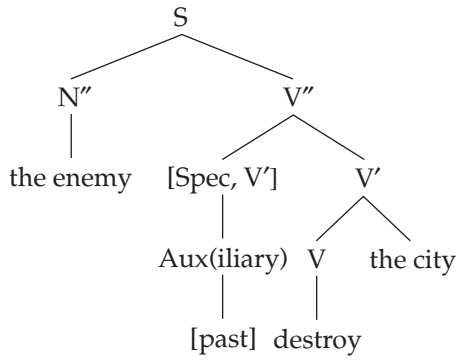
(15) $S \rightarrow N'' \; V''$

Given the X′-schemata in (14) and the "S-introducing" phrase structure rule (15), the internal structures of noun phrases such as *the enemy's destruction of the city* and sentences like *the enemy destroyed the city* should be as follows (omitting much detail):

(16) a. **Noun phrases**



b. **Sentences**



While these internal structures of noun phrases and sentences are sufficiently similar to permit a generalized cross-categorial formulation of grammatical relations and strict subcategorization properties for noun/verb pairs, it is also apparent that further (and rather complete) parallelism could be obtained if sentences are to be analyzed as V″. This issue, however, turns out to be complex and controversial, and in fact motivates much of the subsequent development of X′-theory after Chomsky (1970), as we will see below.

A final tenet of Chomsky's X′-theory concerns the feature analysis of syntactic categories, according to which categories are defined in terms of the two primitive features [±N] (substantive) and [±V] (predicative). The major "lexical categories" are thus defined as follows, using these two primitive features:

(17) N = [+N, –V]
A = [+N, +V]
P = [–N, –V]
V = [–N, +V]

This feature analysis claims that categories in syntax are not really "atoms," but rather, they are decomposable feature complexes characterized by the primitive features, pretty much as "phonemes" are decomposed in terms of distinctive features in phonology. And, as in phonology, this approach makes it

possible to define certain "natural classes" of syntactic category with respect to various syntactic operations and principles. Thus, we can capture the generalization that NPs and PPs behave in the same way (as opposed to VPs and APs) with respect to certain transformations, by attributing it to the feature specification [–V]; we can define the class of possible (structural) Case assigners, V and P (as opposed to N and A), by referring to the [–N] feature; we (correctly) predict that N and V never form a natural class because of their completely conflicting feature specifications, and so on (see, among many others, Bresnan 1977, Chomsky 1981).

Summing up the discussion so far, the basic claims of X′-theory of Chomsky (1970) can be stated as follows:

(18) **The basic claims of X′-theory**
   a. Every phrase is "headed," i.e., has an endocentric structure, with the head X projecting to larger phrases.[3]
   b. Heads (categories) are not atomic elements; rather, they are feature complexes, consisting of the primitive features [±N] and [±V].
   c. UG provides the general X′-schemata of the following sort (cf. (14)), which govern the mode of projection of a head:
      X′ → X . . .
      X″ → [Spec, X′] X′

The version of X′-theory presented in Chomsky (1970) was in a preliminary form, and there certainly remained details to be worked out more fully. However, it is also true that all the crucial and fundamental insights of X′-theory were already presented in this study and have been subject to little substantive change in the following years. More specifically, the claims (18a) and (18b) above have survived almost in their original forms throughout the following development of grammatical theory and are still assumed in the current framework, while the claim (18c), the existence of the universal X′-schemata, has been subjected to critical scrutiny in recent years, as we will see in the next section.

The proposal of X′-theory was followed by a flux of research on phrase structure in the 1970s, trying to fix some technical problems associated with the initial version of the theory and to expand the scope of X′-theory to extensive descriptive material. The relevant literature in this era is too copious to mention in detail, but to name just a few: Siegel (1974), Bowers (1975), Bresnan (1976, 1977), Emonds (1976), Hornstein (1977), Selkirk (1977), and perhaps most importantly, Jackendoff (1977). From our current perspectives, two important and interrelated problems emerged during this period. They are (i) the analysis of sentences (or clauses) *vis-à-vis* X′-theory, and (ii) the proper characterization of "Spec." Let us look at these issues in some detail.

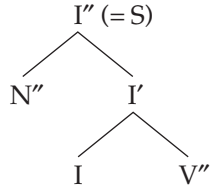
As we saw above, the sentential structure was handled in Chomsky (1970) by the phrase structure rule (15), which does not conform to the general X′-schemata in (14), thereby making the structure of a sentence a kind of an

exception to X′-theory. And this is the main reason for the rather incomplete parallelism between sentences and noun phrases as depicted by (16). Naturally, a proposal has been made, most notably by Jackendoff (1977) (cf. also Kayne 1981c), that a sentence be analyzed as the (maximal) projection of V, with its subject being treated as [Spec, V′] (or [Spec, V″] in Jackendoff's system, since he assumes that X‴ is the maximal level for every category). While this proposal has the obvious advantage of making the internal structures of sentences and noun phrases (almost) completely parallel, there exists some evidence against this claim (see Hornstein 1977, among others). The most crucial evidence that counters the $S = V^{max}$ (the maximal projection of V) analysis comes from the close relationship holding between the subject of a sentence and I(nflectional elements, including the traditional notion of AUX) of that sentence. For example, it is the I of a sentence that assigns nominative Case to the subject, and it is also I that the subject agrees with (in terms of number, person, etc.). And this kind of formal relation cannot be straightforwardly stated if the subject is generated inside the projection of V, with I outside of that projection. Thus, even in Chomsky (1981), S is still generated by the following phrase structure rule (adapted from Chomsky 1981), where the subject N″ is placed outside the maximal projection of V:

(19) $S \rightarrow N'' \; I \; V''$

Huang (1982) proposes (cf. also Stowell 1981, Pesetsky 1982) that S should in fact be analyzed as the maximal projection of I, a natural extension of the spirit of X′-theory. His arguments for this claim mainly come from considerations of the behavior of the subject and I with respect to general principles such as the Empty Category Principle (ECP). In particular, Huang argues that I really behaves like a head in that it governs (but does not properly govern, at least in English) the subject (see Huang 1982 for much detailed discussion). The internal structure of a sentence now looks like the following, which conforms to X′-theory:

(20)



In (20), the subject is the [Spec, I′] and the sentential structure now looks quite "normal" in the sense that there is nothing special with it in light of X′-theory, now extended to a "non-lexical" category I. Note, however, that the incompleteness of parallelism between sentences and noun phrases still remains even under this modified analysis: the subject of a noun phrase is inside its own projection, whereas the subject of a sentence is generated outside of the projection of a verb. This problem was resolved when the new analysis of

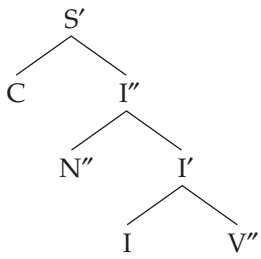
subjects (called the "Predicate-Internal Subject Hypothesis") was introduced, as we will discuss in the following section.

Returning to the historical discussion of the analysis of sentences, Bresnan (1972), based on extensive study of *wh*-movement phenomena, introduced a larger clausal unit that includes the core part of the sentence (S) and the "sentence-introducer," called C(omplementizer) (e.g., *that, for, whether*, etc.) Thus, the structure of a full clause (notated as S′ (S-bar)) should be introduced by the following phrase structure rule (see Bresnan 1972 for details):

(21) $S' \rightarrow C\ S$

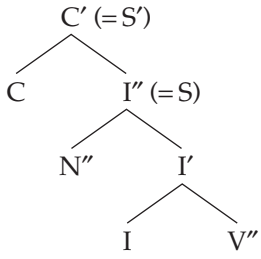
Given the structure of S in (20), the structure of a full clause is:

(22)



In (22), the top portion of the structure is still an exception to X′-theory. S′ is not headed by anything, but rather, branches to two coordinated elements, C and I″ (= S). Evidence has been accumulated, however, to show that C functions as a head, in terms of, particularly, the ECP (Fassi Fehri 1980, Stowell 1981, Lasnik and Saito 1984, among others). This led to the proposal of analyzing C as the head of S′, thus reanalyzing the latter as C′:

(23)



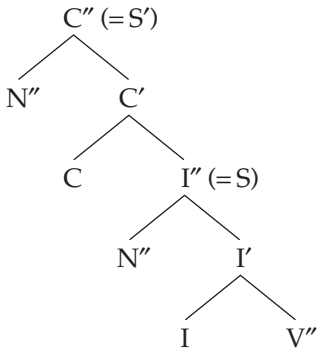
Now the clausal structure is made to fall under X′-theory almost completely, the only problem being the "defectiveness" of the complementizer phrase, i.e., it projects only to C′, not to C″. To see how this final gap was filled, we should turn to the other major problem that motivated the development of X′-theory, namely, the characterization of Spec.

In Chomsky's (1970) version of X′-theory, "Spec" constituted a rather heterogeneous set, including a variety of "pre-head" elements. Thus, Chomsky suggested that [Spec, V′] includes auxiliary elements of various sorts (with

time adverbials associated), [Spec, N′] is instantiated as determiners, [Spec, A′] contains the system of qualifying elements such as comparative structures, *very*, etc. As the research progressed, however, it became increasingly apparent that those pre-head elements can be classified into different types, and that the notion of Spec should be more narrowly defined to capture the true generalization. Accordingly, some elements that were initially identified as Spec were later reanalyzed as heads (e.g. auxiliary elements, now analyzed as instances of the head I), or "adjuncts" (modifiers) that are optionally generated to modify heads (e.g. *very*), although many descriptive questions remain (even now) with the analysis of the latter.

The notion of Spec that resulted from these efforts has the following properties: (i) it is typically an NP, and (ii) it bears a certain relationship with the head. Of the pre-head elements in English, the fronted *wh*-phrase, the subject of a sentence, and the subject of a noun phrase exhibit these properties. Thus, the subject of a sentence is identified as [Spec, I′], and the subject of a noun phrase (as in *the enemy's destruction*) is characterized as [Spec, N′].[4] The fronted *wh*-phrase apparently shows the two properties just discussed: it is typically an NP (or at least a maximal projection), and it bears a certain relationship with the head C (it is a [+wh] C that triggers *wh*-movement; see Bresnan 1972). Thus, it is well qualified to be [Spec, C′], patterning with the other Specs. However, to characterize a fronted *wh*-phrase as [Spec, C′] requires a reanalysis of *wh*-movement. Namely, *wh*-movement should now be analyzed as "movement to [Spec, C′]," rather than "movement to C," as has been long assumed ever since Bresnan's pioneering work (Bresnan 1972). This is in fact what Chomsky (1986b) proposes, with some additional arguments to support this conclusion (see Chomsky 1986b for details). If a fronted *wh*-phrase occupies [Spec, C′], then the structure of a full clause now looks like the following, with the projection of C completely on a par with other projections (i.e., no "defectiveness" of C^{max}):

(24)

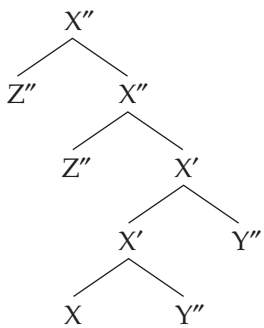


X′-theory is now in full force, regulating the clausal structure, which has always been an exception to the theory in one way or another, as well as the structure of other phrases. The basic ideas of the version of X′-theory presented in Chomsky (1986b) can be stated as follows:[5]

(25) **X-bar schemata** (cf. Chomsky 1986b)
a. X′ = X/X′ Y″
b. X″ = Z″ X′/X″

In (25), X means X^0, a zero-level category (i.e., a head), the "/" sign between symbols indicates that there is a choice between them (e.g. either X or X′ can be chosen in (25a)), and X, Y, Z are variables ranging over possible categories (now including non-lexical categories). Notice that by allowing the same symbol (viz., X′ in (25a) and X″ in (25b)) to occur on both sides of the same equation, we permit "recursion" of the same bar-level structures in a phrase. For example, (25) licenses the following structure, where X′ and X″ each appear twice:

(26)



We call the lower Y″ in (26) the "complement" of X, the lower Z″ the "Spec" of X′ (or the Spec of X″ ([Spec, X″]); see n. 2), and the upper Z″ an "adjunct" of X″. The status of the upper Y″ in (26) is ambiguous (it could be a "quasi-complement" or an adjunct, for instance), depending on further articulation of the theory of phrase structure (see Chomsky and Lasnik 1993). Note incidentally that these notions (complement, Spec, and adjunct) are "relational" notions defined in terms of their structural positions, not inherent and categorical ones (unlike notions such as "Noun Phrase," which are categorical). This is an assumption that has been pretty much constant throughout the history of X′-theory.

A few more general remarks are in order with respect to the X′-scheme in (25). First, one might notice the use of equations in (25), rather than X′-"rules" that have been exploited in previous works on X′-theory. In most earlier works, X′-theory was taken to be a principle of UG that provides the general "rule schemata" that regulate the general form of phrase structure rules of human language. This traditional conception of X′-theory collapsed when the very notion of phrase structure rules was subjected to critical scrutiny, and was eventually eliminated around 1980, when the "principles-and-parameters" approach was first set forth in a systematic way (see Chomsky 1981). Specifically, it was pointed out that phrase structure rules are redundant and dubious devices, recapitulating the information that must be presented in the lexicon. For example, the fact that the verb *persuade* takes an NP and S′ (= C″) complement has to be stated as the verb's lexical property, quite independently

from the phrase structure rule that generates the sequence V-NP-S′/C″. And since descriptions of lexical properties in the lexicon are ineliminable, it is the phrase structure rules that ought to be eliminated. Subsequent work such as Stowell (1981) showed that the other information expressed by phrase structure rules (most of which have to do with linear ordering) can in large part be determined by other general principles of UG (such as Case theory; see Stowell 1981). Thus, it was generally believed in the principles-and-parameters approach that phrase structure rules could be entirely eliminated, apart from certain parameters of X′-theory. With the notion of phrase structure rules eliminated from the grammar, X′-theory has become a principle of UG that directly regulates phrase structure of human language.

Second, the X′-scheme in (25) is formulated only in terms of the structural relation "dominance," and does not encode the information regarding linear order. Thus, of the three types of information listed in (2) before, i.e., (2i) Dominance, (2ii) Labeling, and (2iii) linear order (Precedence), only the first two ((2i) and (2ii)) are regulated by X′-theory itself. The linear order of elements (2iii) is to be specified by the "parameter" (called the "head parameter") associated with X′-theory. This is in accordance with the general guidelines of the principles-and-parameters approach, under which UG is conceived of as a finite set of invariant principles each of which is associated with a parameter whose value is to be fixed by experience. There are two values of the head parameter, "head initial" and "head last." If the parameter is set for the value "head initial," the English-type languages follow, in which complements generally follow their heads, whereas if the value is set as "head last," the Japanese-type languages obtain, where complements typically precede their heads.[6] With this move to parametrized X′-theory, the phrase structure system for a particular language is largely restricted to the specification of the parameter(s) that determine(s) the linear ordering of elements.

Finally, given the narrower characterization of Spec as a place for a maximal projection (typically a noun phrase), we now have a much simplified theory of movement. Chomsky (1986b) proposes that there are two types of movement: (i) X^0-movement (movement of a head), and (ii) X″ (or X^{max})-movement (movement of a maximal projection). We put aside the discussion of X^0-movement (see Chomsky 1986b, 1995a; see also Roberts in this volume for much detailed discussion of this type of movement). Movement of a maximal projection is divided into two subtypes: (i) substitution, and (ii) adjunction. Chomsky then argues that, apart from X^0-movement to a head position (which we put aside), various principles of UG ensure that substitution (NP-movement and *wh*-movement) always moves a maximal projection to a specifier position (see Chomsky 1986b for details).[7] Thus, the notion of "Spec" now receives a uniform characterization as a landing site for X^{max}-movement: [Spec, C′/C″] is the landing site for *wh*-movement, [Spec, I′/I″] is the landing site for NP-movement (passive and raising), and [Spec, N′/N″] is the landing site for "passive" in a noun phrase. We will return to adjunction later on.

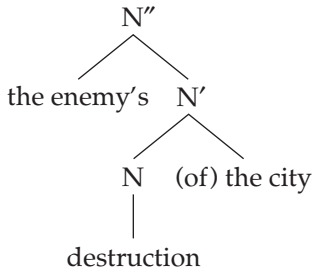
To sum up, Chomsky's (1986b) version of X′-theory has the following characteristics. First, it includes two "non-lexical" categories, I and C, as members of

"X" relevant for X'-theory, so that a full clausal structure is now in full conformity with the principles of X'-theory and "sentences" are no longer exceptions to the theory, a great improvement over earlier versions of X'-theory for which "sentences" have always been treated as exceptions. Second, X'-theory is now parametrized in accordance with the general guidelines of the principles-and-parameters approach, and the theory no longer specifies the linear ordering of elements in the scheme. The ordering restrictions are determined by the value for the parameter (the head parameter) associated with X'-theory, not by X'-theory itself. And finally, the notion of Spec is further sharpened as a landing site for movement of a maximal projection (substitution), with a remarkable simplification of the theory of movement. Some important problems, however, remained open in this version of X'-theory, which motivated further development of the theory in the decade that followed.

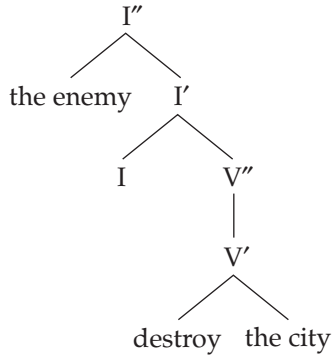
2 Minimizing and Deriving X'-Theory

An obvious point in Chomsky's (1986b) version of X'-theory that calls for further improvement is the incomplete parallelism it expresses between noun phrases and clauses/sentences. Compare the following structures which are assigned to noun phrases and clauses in this theory:

(27) a. **Noun phrases**



b. **Clauses**

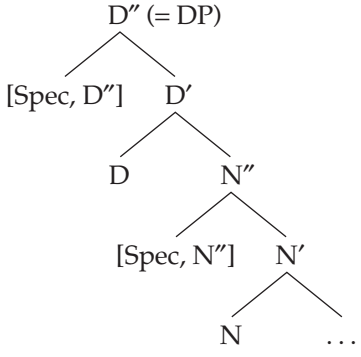


There are various problems with the structures in (27). The source of the problems is the fact that in (27a), all the "arguments" (subject and object) are located within the maximal projection of a single head (N = *destruction*), while in (27b), subject and object are split in two different projections. In other words, in a sentential structure (27b), there is an "additional" structure, due to the existence of the head I; in (27a), on the other hand, there is no such additional structure and all the arguments are located within the projection of N. From this discrepancy, a variety of problems arise. Why is the subject of a sentence located in [Spec, I″], a non-lexical category (I will henceforth follow a more recent practice to notate the Spec), whereas the subject of a noun phrase is located in [Spec, N″], a lexical category? A related question is: why does the "passive" in a sentence (e.g. *the city was destroyed* (*by the enemy*)) move a maximal projection to the specifier position of a non-lexical category ([Spec, I″]), but the corresponding passive in a noun phrase (e.g. *the city's destruction* (*by the enemy*)) moves a maximal projection to the specifier position of a lexical category ([Spec, N″])? Also, why does V project from V′ to V″, without having Spec? And so on. The structures in (27) are clearly not parallel enough to capture the similarities between noun phrases and sentences.

Two proposals were made in the mid- to late 1980s which played important roles in resolving these problems. They are (i) the "DP-analysis" (Fukui and Speas 1986, Abney 1987; see also Brame 1981, 1982), and (ii) the "Predicate-Internal Subject" Hypothesis (see Hale 1978, Kitagawa 1986, Koopman and Sportiche 1991, Kuroda 1988, among others, for various versions of the "VP-Internal Subject" Hypothesis; see Fukui and Speas 1986 for a generalized form of the hypothesis as it is applied to all predicative categories).

The DP-analysis claims that "noun phrases" are in fact "determiner phrases" (DP) headed by the head D which takes a noun phrase as its complement. (See Longobardi in this volume for much relevant discussion, including detailed (crosslinguistic) analyses of the internal structure of noun phrases under this hypothesis.) According to this analysis, then, the internal structure of a noun phrase should be as follows:

(28)



It was argued in the above-mentioned works that the DP-analysis is in fact supported by various syntactic considerations (see also Bernstein and Longobardi in this volume). Furthermore, the DP-analysis of noun phrases received much justification from the semantics of nominal expressions (a similar analysis had in fact been assumed in Montague semantics before the syntactic DP-analysis was proposed). Thus, this analysis has become more or less a standard analysis of noun phrases and is assumed in much current literature.

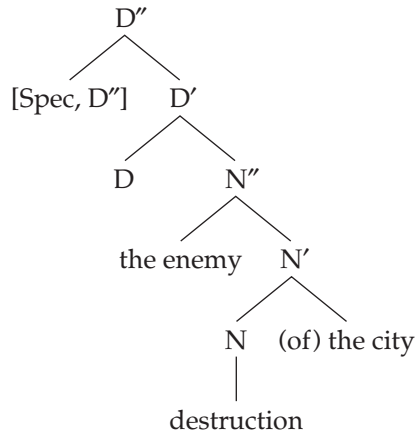
Notice that the DP-analysis provides a "two-story" structure for noun phrases that looks quite similar to the structure of sentences: in both structures, a non-lexical category (I in a sentence, D in a noun phrase) heads the whole phrase, taking a complement headed by a lexical category (V in a sentence, N in a noun phrase). Given the DP-analysis, then, the parallelism between sentences and noun phrases becomes much more visible and easy to capture than in the traditional analysis of noun phrases.

Where, then, is the subject located in these structures? Quite independently of the DP-analysis, it was proposed that the subject of a sentence should be generated in the projection of a verb (see the references cited above). In fact, the analysis that the subject of a sentence should be generated within a verb's projection is a rather traditional one (see, for example, Jackendoff 1977), which has been challenged by various evidence that the subject of a sentence is in a close relationship with I (see the discussion above). In other words, there seem to be two apparently conflicting sets of evidence regarding the status of the subject in a sentence: one type of evidence (most of which has to do with theta-theoretic considerations) indicates that the subject should be inside the verb's projection, while the other type of evidence (having to do with Case, agreement, government, etc.) suggests that the subject must occupy [Spec, I"]. The "VP-Internal Subject" Hypothesis was proposed mainly to reconcile these two types of evidence. The crucial and novel part of this hypothesis is the movement process that raises the subject (which is generated inside the verb's projection) to [Spec, I"]. This movement is driven by the need for Case assignment. Thus, the subject of a sentence is generated in [Spec, V"] (in some versions of the VP-Internal Subject Hypothesis, not in others), and then, is moved to [Spec, I"] in order to receive Case in that position. The D-structure position of the subject accounts for the subject's theta-theoretic status with respect to the verb, whereas its S-structure position (after the movement) accommodates the evidence indicating its close relationship with the inflectional head (I) (note that Case and agreement are S-structure (or at least non-D-structure) phenomena).

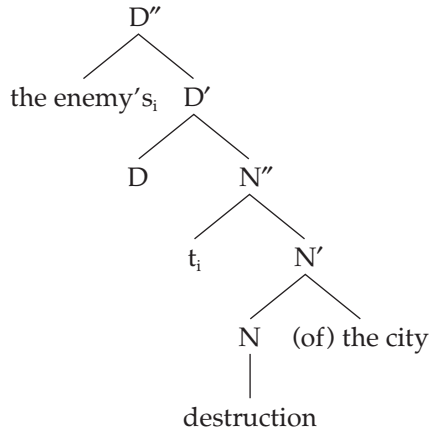
Combining the DP-analysis and the VP-Internal Subject Hypothesis (thus making the latter the "Predicate-Internal Subject" Hypothesis), we have the following completely parallel structures for noun phrases and clauses/sentences (Fukui and Speas 1986):

(29) a. **Noun phrases**

D-structure:

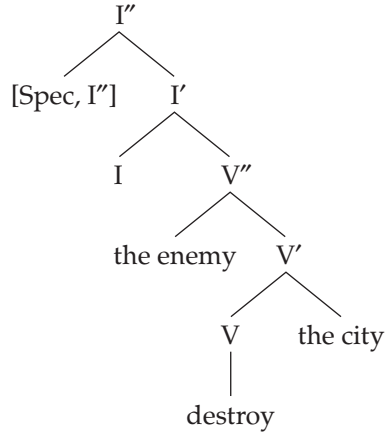


S-structure:

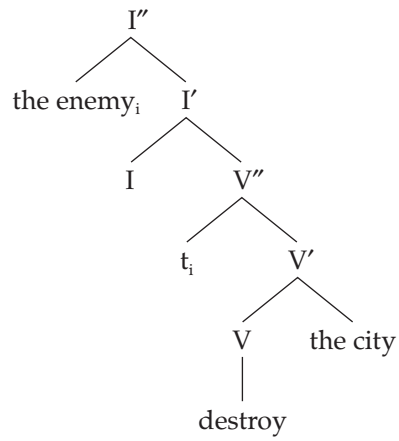


b. **Clauses**

D-structure:



S-structure:



The subjects in both noun phrases and sentences are generated within the projection of the lexical category (N in a noun phrase and V in a sentence), receiving a theta-role in their original positions, and then are raised to the Spec positions of associated non-lexical categories (D in the case of noun phrases, I in sentences) to receive Case (genitive in noun phrases, nominative in sentences).[8] Passives in noun phrases (e.g. *the city's destruction* (*by the enemy*)) and those in sentences (e.g. *the city was destroyed* (*by the enemy*)) can be analyzed uniformly as a process involving movement of an object from its base position (the complement position of a predicate N/V) to the Spec of an associated non-lexical category ([Spec, D″] in noun phrases and [Spec, I″] in sentences).

The integration of the DP-analysis and the Predicate-Internal Subject Hypothesis was based on the following ideas about the lexicon as it relates to syntactic computation. (See Fukui and Speas 1986, Abney 1987, for some preliminary discussions; see also Fukui 1986 for further discussion on this and related issues.) Items of the lexicon are divided into two major subtypes: lexical categories and "functional" categories. The latter types of category roughly correspond to the traditional non-lexical categories, renamed in consideration of their nature. Lexical categories have substantive content, and include nouns, verbs, adjectives, etc. They typically enter into theta-marking. Functional categories do not have substantive content, and do not enter into theta-marking (although they do have other feature structures, including categorial features, agreement features, etc.). Lexical categories play an important role in interpretation of linguistic expressions, and indeed, most of the items in the lexicon belong to this type. Functional categories, on the other hand, do not play a comparable role in interpretation of linguistic expressions; their role is largely restricted to "grammatical" (or "computational") aspects of linguistic structure (although some of the proposed functional categories, e.g. I and D, may sometimes function as operators, bearing some "semantic" import). These categories constitute a small (and often closed) set, which include C, I, D (assuming the DP-analysis), and a few others.

Thus, the general view on the nature of these categories is the following division of labor for constructing linguistic expressions:

(30) (i) Lexical categories: the "conceptual" aspects of linguistic structure.
(ii) Functional categories: the "computational" aspects of linguistic structure.

Lexical categories bear semantic features, including, in particular, features having to do with theta-roles ("theta-grids'' in the sense of Stowell 1981). They assign (or "discharge") theta-roles/features associated with them to other phrases, thereby forming larger structures that embed them. Functional categories do not bear theta-roles. Their role is largely restricted to purely formal and computational aspects of linguistic structure such as marking grammatical structures (nominals and clauses) or triggering movement operations. More specifically, some functional categories (functional heads) bear "agreement features," and these agreement features attract a maximal projection to their neighborhoods (their specifier positions), in order for the latter to agree with the former. Thus, functional categories are indeed the "drive" for syntactic movement operations; lexical categories lack agreement features of this kind, and hence do not induce movement.

The idea of functional categories as the major driving force for movement opened up a new way of looking at crosslinguistic variation, and facilitated much subsequent work on comparative syntax in terms of properties of functional elements in languages. Given the nature and role of functional categories, it was proposed that language variation be restricted (apart from ordering restrictions) to the functional domain of the lexicon (Fukui 1986, 1988; see also Borer 1984), and this proposal contributed to constructing a more restrictive theory of comparative syntax. At the same time, numerous "new" functional categories were proposed in the late 1980s, achieving tremendous descriptive success, although from an explanatory point of view, it was clear that the class of possible functional categories has to be severely restricted in a principled way (Fukui 1988, 1995; see also Chomsky 1995a for a "Minimalist" critique of functional elements). See Belletti and Zanuttini (both this volume) and references there for much relevant discussion.

Explicit recognition of the division of labor between lexical and functional categories, as well as increasing emphasis on the importance of features in phrase structure composition, naturally led to a theory of phrase structure called "Relativized X'-theory," which is an attempt to minimize the role of X'-theory, while maintaining its basic insights.[9] The fundamental idea of Relativized X'-theory, inspired by categorial grammars, can be summarized as follows:

(31) Phrase structure composition is driven by feature discharge.

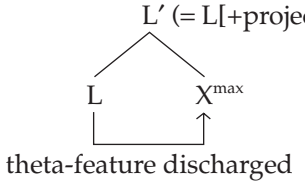
Recall that lexical items have always been assumed, at least since Chomsky (1970), to be feature complexes (see the discussion in section 1). Given the

fundamental difference between lexical and functional categories noted above, we can roughly assume the following feature specifications of these categories (see Chomsky 1995a for recent and much more elaborated discussion on features):

(32) (i) Lexical categories = {categorial features, theta-features (theta-roles/theta-grids), subcategorization features, phonological features, etc.}.
(ii) Functional categories = {categorial features, agreement features, subcategorization features, phonological features, etc.}.

The crucial difference, then, is that lexical categories bear theta-features but not agreement features, whereas functional categories lack theta-features but are associated with agreement features. And this crucial difference is directly reflected in their modes of projection in Relativized X'-theory. Thus, lexical categories project as they discharge their theta-features in the following manner:[10]

(33)

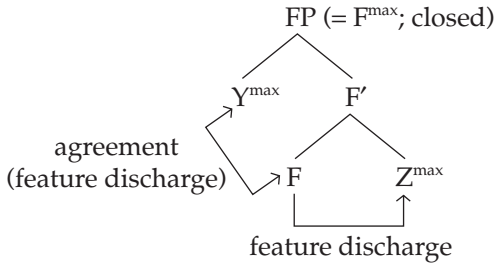


Lexical categories continue to project, forming larger structures, as they discharge their theta-features, until all the features have been discharged. In other words, the structure created in this process is recursive, and in this sense, the projection of a lexical category is never "closed." Note that in this system, the notion of "maximal projection" can no longer be defined in terms of "bar-levels," as in the standard X'-theory. Thus, maximal projection is defined as follows, in a way that is "relativized" to each head and configuration (see Muysken 1982 for an original proposal of this kind; see also Baltin 1989 for a similar approach):

(34) The "maximal projection" of X is a category X that does not project further in a given configuration.

The mode of projection of functional categories, although also governed by feature discharge, is different from that of lexical categories, since functional categories do not bear theta-features but instead have agreement features to discharge, and it is claimed that agreement is typically a one-to-one relation (Fukui 1986, Kuroda 1988). Thus, if a functional head F takes a maximal projection, discharging its subcategorization feature to the latter, and then takes another maximal projection for the purpose of agreement, its projection is "closed" at that point, due to the one-to-one nature of agreement:

(35)



Once agreement occurs, therefore, the projection is closed (the closed projection of X is notated as "XP" in this theory, as in (35)), and no further projection is possible. As a closed category cannot project further, it is also a maximal projection. Notice that the reverse is not true. While a closed projection is always a maximal projection, being a maximal projection does not imply it is closed (by agreement). In fact, projections of lexical categories do have maximal projections, but they never have closed projections, simply because lexical heads do not have agreement features (therefore, there is no independent "LP," a closed lexical projection).

Recall that agreement features are the driving force for movement in syntax. Since only functional categories bear these features, it follows that only functional categories induce movement; lexical categories never trigger movement. And this is well in accord with the facts: NP-movement (passive and raising) moves a noun phrase to [Spec, I^{max}]; passive in a noun phrase is the process of moving a noun phrase to [Spec, D^{max}]; and *wh*-movement moves a *wh*-phrase to [Spec, C^{max}] (note that in Relativized X'-theory, maximal projections are not inherently related to bar-levels, even if the latter notion exists at all). It looks as though all typical movements (at least in English) are to the specifier position of a functional category. To sharpen the notion of "Spec" still further, Relativized X'-theory proposes that Spec be defined in terms of agreement:

(36) The specifier of X is a maximal projection that agrees with X.

Thus, a moved *wh*-phrase, the subject that agrees with I, etc. are all Specs, but lexical categories do not have Specs, since they do not have agreement features to license Specs. The definition of Spec in (36) clarifies the nature of Spec more than ever: Spec is the landing site for movement.

Relativized X'-theory has an important implication for the theory of movement. The issue, again, has to do with the notion of Spec. In the standard theory of movement, an empty category Δ is generated in the position of a specifier at D-structure (or in the course of a derivation; see Chomsky 1993) as a target for substitution. However, it is impossible to justify this empty category in Relativized X'-theory. It is not licensed by theta-assignment, since a functional head never assigns a theta-role, and it is not licensed by agreement because it is the moved maximal projection, not an empty category, that agrees with the functional head. Also, the empty category never appears on the surface;

it is there only to be replaced by a moved category, and it always has to be "erased" before the derivation ends. Therefore, such a superfluous empty category is eliminated in Relativized X'-theory, and hence, substitution operations no longer exist as operations that "substitute for" some existing element. Formally, then, the operation that is involved in "substitution" is no different from Adjunction, in the sense that it does not substitute for anything.[11] Thus, the traditional notion of "substitution" transformation, with the dubious empty category Δ, is eliminated from the theory of grammar (but see n. 11). And if phrase structure composition is also carried out by a formal operation Adjunction (whose application is driven by feature discharge), as we saw above, then it seems that there is one uniform operation which is responsible for both phrase structure building and movement, namely, Adjunction.[12]

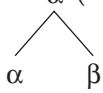
Relativized X'-theory minimized, in fact virtually eliminated, the need for an X'-schema, which had been assumed throughout the previous development of X'-theory. It takes seriously the notions (i) projection and (ii) feature discharge, and claims that every position in phrase structure must be licensed in terms of these notions. Since lexical and functional categories have different feature specifications – in particular, only the latter bears agreement features – the modes of projection of these two types of category must reflect the difference. Thus, only functional categories have Specs as a landing site for movement, triggered by agreement features associated with the functional heads, whereas lexical categories never have Specs and their projections are thus never closed. From this, it immediately follows that if a language lacks functional categories (or if its functional system is inert), then the phrase structure of the language is essentially based on the lexical system, phrasal projections in the language are never closed, and no syntactic movement is triggered. Fukui (1986, 1988) argues that this is indeed the case in languages like Japanese, and demonstrates that a variety of typological properties of Japanese, e.g. the lack of *wh*-movement, the existence of multiple-nominative/genitive constructions, scrambling, and many others, are derived from this fundamental parametric property of the language.

The total elimination of X'-theory was proposed and carried out by Chomsky's (1994) "bare phrase structure" theory (see also Kayne 1994 for a different approach). The bare theory is couched within the "Minimalist program" (Chomsky 1993), according to which all the principles and entities of grammar must be motivated and justified either by the properties of two "interface representations," LF and PF, or by considerations of economy (see Chomsky 1993 for details; see also Collins in this volume, and other relevant chapters). Most of the basic claims of Relativized X'-theory carry over to the bare theory, except for a particular characterization of Spec in the former as an X^{max} agreeing with a head (see (36); see also the next section for some relevant discussion).

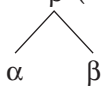
Chomsky argues that (the standard) X'-theory specifies much redundant information, while the only structural information needed is that a "head" and a "non-head" combine to create a unit. He then proposes that a phrase structure

is constructed in a bottom-up fashion by a uniform operation called "Merge," which combines two elements, say α and β, and projects one of them as the head. This is illustrated in (37), where the prime simply means the category is projected (see n. 10):

(37) α β $\xrightarrow[\text{Merge}]{}$ if α projects α′ (= α[+projected])



if β projects β′ (= β[+projected])



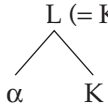
Since Merge does not specify the linear order of α and β, the tree structures in (37) can be more formally, and more accurately, represented as in (38):

(38) $K = \{\gamma, \{\alpha, \beta\}\}$, where $\gamma \in \{\alpha, \beta\}$

(38) states that Merge forms a new object K by combining two objects α and β, and specifies one of them as the projecting element (hence the head of K). Merge applies recursively to form a new structure.

Chomsky further argues that Merge is involved in both phrase structure composition and movement processes. Suppose that Merge is to apply to α and K, to form a new unit L, with K projecting:

(39) α K $\xrightarrow[\text{Merge}]{}$ L (= K′)



The only difference between simple phrase structure building and movement is whether α in (39) comes from the lexicon (or from the Numeration, in current terms), as in the case of phrase structure building, or from within K (leaving its copy in the original place), as in the case of movement. Thus, the bare theory unifies phrase structure composition and movement in terms of the single operation Merge (which is somewhat reminiscent of Adjunction in Relativized X′-theory).

A "maximal projection" is also defined relationally in the bare theory: a category that does not project any further in a given configuration is a maximal projection. The terms "complement" and "specifier' are defined in the usual way. Note that the definition of the latter concept (Spec) in the bare theory is different from that of Spec in Relativized X′-theory. In Relativized X′-theory, Spec is defined in terms of agreement (cf. (36)), with the consequence that only functional categories have Specs. In the bare theory, on the other hand, agreement does not play any significant role in defining Spec, and hence Spec is

defined in the traditional way as a phrase that is immediately dominated by a maximal projection. This (and the associated distinction between X^{max} (a simple maximal projection) and XP (a closed maximal projection)) seems to be the only substantive difference, apart from details, between Relativized X'-theory and the bare theory. See Fukui (1991), Fukui and Saito (1992), and Saito and Fukui (1998) for some arguments for the necessity of X^{max}/XP distinction. See also the next section for some relevant discussion.

With Chomsky's bare theory, X'-theory is now completely eliminated as an independent module of grammar. The basic insights of X'-theory, in particular, the insight that every phrase is headed in human language (cf. (18a)), is straightforwardly expressed as a fundamental property of the operation Merge, without postulating an additional "principle."

However, of the three kinds of information about syntactic representations listed in (2), i.e., (2i) Dominance, (2ii) Labeling, and (2iii) linear order (Precedence), the last kind of information is not encoded at all in Chomsky's formulation of Merge given above. In fact, whether or not the theory of phrase structure should specify the linear order of elements still remains open in current research, to which we now turn.

3 Linear Order in Phrase Structure

The concept of linear order in a phrase marker was never questioned in an earlier framework of generative grammar. In fact, it was, as stated in (2), one of the few crucial primitive concepts in the theory of phrase structure, and a variety of grammatical rules was formulated with a crucial reference to linear order (see, for example, "pronominalization" transformation in the 1960s). However, it has been increasingly less obvious that linear order plays a role at all in language computation, apart from phonology. Thus, virtually all the principles and conditions assumed in the principles-and-parameters theory in the 1980s are formulated purely in hierarchical terms (in terms of domination and c-command), without referring to linear order. The "head parameter" (and its variants) seems to be the only notion in linguistic theory which crucially refers to linear order.

Kayne (1994) challenges this notion of head parameter. He proposes a universal principle, the Linear Correspondence Axiom (LCA), which states essentially that asymmetric c-command imposes a linear ordering of terminal elements. More specifically, the LCA dictates that if a non-terminal X asymmetrically c-commands a non-terminal Y in a given phrase marker P, then all terminals dominated by X must precede or follow all terminals dominated by Y in P. Kayne takes the relevant ordering to be precedence, rather than subsequence (following), based on his assumptions about the relation between terminals and "time slots" (see Kayne 1994 for more details). Thus, within Kayne's theory, asymmetric c-command relations uniquely map into precedence

relations: all terminals dominated by X precede all terminals dominated by Y, in the configuration stated above. It then follows, given Kayne's formulation, that there is a universal S(pecifier)–H(ead)–C(omplement) order (in particular, S(ubject)–V(erb)–O(bject)), with other orders (S-C-H/S-O-V, for example) being derived via movement. With the universal S-H-C order, the head parameter is entirely eliminated.

Note that in Kayne's theory, linear order still plays a role in the core computation of language, though redundantly, because it is entirely determined by asymmetric c-command relations. In other words, Kayne proposes that linear order is not parametrized and that it is uniquely determined by asymmetric c-command relations, given his LCA, which he claims to apply at every syntactic level. But linear order is still defined and remains visible throughout the derivation and could conceivably play a role in the core computation of language.

Chomsky (1994, 1995a), adopting and incorporating the basic insights of Kayne's LCA into his bare theory, makes a step further toward complete elimination of linear order from the core of language computation. As we saw in the preceding section, Chomsky's bare theory, the recursive procedure Merge in particular, does not encode any information regarding linear order of syntactic elements. This is based on his understanding that there is no clear evidence that linear order plays a role at LF or in the core computation of human language.[13] Thus, he assumes that linear order is not defined and hence does not play a role in the core computation of language, and suggests that ordering is a property of the phonological component, a proposal that has been occasionally made in various forms in the literature. Specifically, he claims that a modified version of the LCA applies as a principle of the phonological component to the output of Morphology, a subcomponent of the phonological component (see Chomsky 1995a for detailed discussion). Thus, under Chomsky's proposal, phrase structure is defined without reference to linear order in the core computational part of human language, and will later be assigned linear order by (a modified version of) the LCA in the phonological component.

By contrast, Saito and Fukui (1998) (see also Fukui 1993, Fukui and Saito 1992) claims that linear order indeed plays an important role in the core computational part of human language, and argues that the head parameter, or more precisely a modified version of it, should be maintained. One way, proposed in Saito and Fukui (1998), to incorporate the head parameter into the bare theory is to replace the set notation $\{\alpha, \beta\}$ in (38), reproduced here as (40), by an ordered pair $<\alpha, \beta>$, thereby specifying which of the two elements projects in a given language. Thus, we have (41) instead of (40):

(40) **Chomsky's Merge:** $K = \{\gamma, \{\alpha, \beta\}\}$, where $\gamma \in \{\alpha, \beta\}$

(41) **Fukui and Saito's parametrized Merge:** $K = \{\gamma, <\alpha, \beta>\}$, where $\gamma \in \{\alpha, \beta\}$

If γ takes the value "α," we have a "head-initial/left-headed" language such as English, whereas if γ = β, a "head-last/right-headed" language like Japanese is defined. Thus, in left-headed English, elements can be merged only on the *right* side of a head, whereas in right-headed Japanese, Merge occurs only on the *left* side of a head. If something is to be introduced on the opposite side of the structure (i.e., on the left side of a head in English, and on the right side of a head in Japanese), it must be "adjoined" to the target, creating a multi-segment structure (see Chomsky 1986b, 1995a, for relevant discussion on substitution vs. adjunction). A case in point is the status of subjects in these languages. The subject in English is in an adjoined position because it appears on the left side of the head, where projection of the target is prohibited by (41) as it is parametrized for English. The subject in Japanese, on the other hand, is introduced into phrase structure by Merge (i.e., substitution; see below), since it shows up on the left side of the head, where merger is possible (Japanese is a right-headed language). See Saito and Fukui (1998) for more detailed discussion, as well as illustrations of this point.

Saito and Fukui argue that given the parametrized version of Merge (41), it becomes possible to characterize the traditional "adjunction" operations, viz., scrambling in Japanese and heavy NP shift in English, as paradigm cases of Merge (i.e., as substitution, in the sense that they always accompany projection of the target),[14] and hence, given the costless nature of Merge (Chomsky 1995), the optionality of these operations, a matter that has been quite disturbing for the general economy approach to movement (Chomsky 1991), is also straightforwardly accounted for. On the other hand, traditional "substitution" operations (*wh*-movement and NP-movement) are analyzed in this system as genuine adjunction since they never induce projection of the target, creating a multi-segment structure of the target (see Saito and Fukui 1998 for much detail). Further, they point out that the "directionality" of these optional movements correlates with the "directionality" of projection in the language. Thus, head-initial/left-headed English has rightward heavy NP shift, whereas head-last/right-headed Japanese exhibits leftward scrambling, but no other combination is allowed. It is clear that such a correlation can only be captured by a parametrized Merge embodying linear order, as in (41). Saito and Fukui show that a number of other differences between English and Japanese also follow from their theory of phrase structure.

The parametrized Merge has an important implication for the theory of locality on movement. It has been known since Cattell (1976), Kayne (1981c), and Huang (1982) that a non-complement maximal projection forms an island for movement (see also Rizzi, this volume, for some relevant discussion). Thus, extraction out of subjects and adverbial adjuncts results in ungrammaticality, as shown in (42):

(42) a. ?*Who$_i$ did [a picture of t_i] please John
b. ?*Who$_i$ did John go home [because he saw t_i]

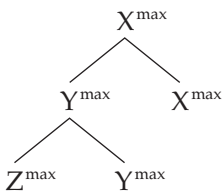
The effects illustrated by (42a) and (42b) are called the Subject Condition effects and the Adjunct Condition effects, respectively. There are two important problems with respect to these effects. One is how to unify them in a natural way. The other problem has to do with the crosslinguistic considerations of these effects. The Adjunct Condition effects are generally assumed to be universal, whereas the Subject Condition effects are known to show crosslinguistic variation. Specifically, it appears that while SVO languages generally exhibit the Subject Condition effects, SOV languages systematically lack the effects (Kayne 1984; see also Aissen 1996 for related discussion). Huang (1982) proposes the Condition on Extraction Domain (CED), which unifies the Subject and Adjunct Condition effects in terms of the notion of "proper government," and suggests a possible way of accounting for the observed crosslinguistic difference with respect to the Subject Condition effects (see Huang 1982 for details). Huang's CED was later incorporated into Chomsky's (1986b) barriers theory as a central ingredient of the latter system.

Takahashi (1994), working under the general "economy" guidelines (see Collins in this volume) and extending Chomsky (1986b) and Chomsky and Lasnik (1993), proposes to derive these effects from the Minimal Link Condition (MLC) and constraints on adjunction sites. The former condition, when interpreted derivationally, requires that movement go through every possible landing site. If any XP dominating the moved element is a potential adjunction site in the case of A'-movement, this implies that the *wh*-phrases in (42) must adjoin to every maximal position that intervenes between their initial positions and the matrix [Spec, CP]. In particular, *who* must adjoin to the subject DP in (42a), and the adverbial CP in (42b). But if adjunction to subjects and adjuncts/modifiers is prohibited, as argued in Chomsky (1986b), then the moved *wh*-phrase must skip a possible landing site in these examples. Hence, (42a–b) both violate the MLC.

The remaining problem is to derive the constraints on adjunction sites. There have been some proposals that treat the subject case and the adjunct (modifier) case separately. (See, for example, Chomsky 1986b, Takahashi 1994a.) However, Saito and Fukui argue that their parametrized Merge approach opens up a refreshingly new way to unify these two cases. Suppose, following a standard assumption, that an adjunct (modifier) appears in a position adjoined to a maximal projection.[15] Then, descriptively, what is prohibited in the adjunct (modifier) case is adjunction to an adjoined phrase. And this extends automatically to the subject case, since a subject in English (or SVO languages generally) is in an adjoined position, as we discussed above. The explanation for the lack of the Subject Condition effects in Japanese (or SOV languages generally) is straightforward: subjects in this language are not in an adjoined position, but rather are introduced into structure by Merge (substitution). The question, thus, reduces to why adjunction to an adjoined position is disallowed.

Saito and Fukui propose that this is due to the indeterminacy of the adjunction site that arises in the relevant case. Consider the following configuration:

(43)



Both X^{max} and Y^{max} neither dominate nor exclude Z^{max} (see Chomsky 1986b for the definitions of these structural notions). Hence, if "adjunction" is defined as in (44), then Z^{max} in (43) is adjoined simultaneously to X^{max} and Y^{max}:

(44) α *is adjoined to* $\beta =_{def}$ neither α nor β dominates the other and β does not exclude α.

Adjunction to adjoined phrases, then, is excluded by the following plausible condition:

(45) An adjunction site must be unique.

Saito and Fukui argue that the condition (45) need not be stipulated as an independent condition on adjunction site, but rather is an instance of the general uniqueness condition on the licensing of (non-root) elements in a phrase marker. (See Saito and Fukui 1998 for a precise formulation of the principle as well as much detailed discussion on relevant points.) Thus, their parametrized Merge, which incorporates the notion of linear order (the head parameter, in particular), unifies, without having recourse to such notions as "proper government" (Huang 1982) and "L-marking" (Chomsky 1986b), the classical cases of CED (the Subject and the Adjunct Condition effects), offering a natural explanation for the parametric variation associated with the Subject Condition effects.

The issue of linear order in phrase structure (and movement) is a complex matter and remains open for future research. It is probably conceptually desirable if we can eliminate the concept of linear order from the core part of human language computation, and locate it in the phonological component, where the importance of linear order is firmly established. On the other hand, if the evidence presented in Saito and Fukui's work is real, then it constitutes a rather strong reason for postulating linear order in the portion of grammar where the theory of economy (cf. their arguments with respect to optionality) and that of locality (recall their reunification of CED) are relevant. See also Fukui and Takano (1998) for related discussion on this issue.

4 Summary and Conclusion

This chapter has discussed the development of the theory of phrase structure in generative grammar. Phrase structure of human language was described in terms of phrase structure rules; context-sensitive phrase structure rules in an

earlier theory of generative grammar, and then context-free phrase structure rules with an enriched lexicon in the Standard Theory. X′-theory was proposed in the late 1960s based on the recognition of the observed deficiencies of phrase structure rules as a means for explaining the nature of phrase structure of human language: (i) phrase structure rules are "too permissive," in that they allow rules generating various structures that are actually never attested, and (ii) phrase structure rules cannot capture certain systematically observed "cross-categorial" generalizations. X′-theory, as an invariant principle of UG, overcomes these problems by claiming (i) that every phrase is "headed" (i.e., has an endocentric structure), with the head X projecting to larger phrases, (ii) that heads (categories) are not atoms, but rather complexes of universal features, and (iii) that projection of heads conforms to the general "X′-schemata" provided by UG. (See (18a–c) in section 1.)

The development of X′-theory from its inception up until the mid-1980s can be characterized as a process of sharpening and elaborating the format of X′-schemata, in such a way as to expand the scope of X′-theory to extensive descriptive material. As we saw in section 1, during this period, the structure of clauses was reanalyzed so as to fall under the scope of X′-theory, and the notion of "Spec" was gradually narrowed down to directly express its nature in phrase markers.

The subsequent development of X′-theory from the mid-1980s to the present can be described, as we discussed in section 2, as an accumulated attempt to minimize the role of X′-schemata, while maintaining the basic insight of X′-theory. Along the way, some novel analyses of particular constructions in phrase structure were proposed (the DP-analysis and the Predicate-Internal Subject Hypothesis), yielding numerous important empirical (crosslinguistic) studies concerning the structure of clauses and noun phrases. Relativized X′-theory makes a fundamental distinction between lexical categories and functional categories, and claims that phrase structure building is essentially feature driven. A "relativized" notion of maximal projection and the further sharpening of the concept of "Spec" in terms of agreement are also major claims of this theory. Motivated by the Minimalist program, the "bare phrase structure" theory completely eliminates the X′-schemata, in terms of the recursive procedure Merge, keeping the major insight of X′-theory almost intact.

Thus, at the current stage of the theory, of the three basic claims of the classical X′-theory (i)–(iii) stated above, (i) and (ii) are preserved in the bare theory, while the claim (iii), i.e., the existence of the X′-schemata, is explicitly denied. In this sense, there seem to be few fundamental differences between the bare theory and other approaches to phrase structure. Various approaches to phrase structure appear to have started converging and fruitfully influencing each other. For instance, given the foremost importance of features in the theory of phrase structure (and in the Minimalist program generally), the explicit mechanisms of feature systems developed in other approaches (e.g. in the GPSG/HPSG traditions) may well have an important impact on further development of the bare theory.

There are of course numerous remaining problems in the theory of phrase structure, many of which, including the influential "shell" structure proposed by Larson (1988), I could not discuss in this chapter. Section 3 briefly discussed one theoretical problem that remains open, i.e., the status of linear order. Various other theoretical questions remain, and as always, vast numbers of descriptive problems keep challenging the current theory of phrase structure. The theory of phrase structure, in my view, will continue to be one of the central topics of linguistic theory for years to come.

NOTES

* Portions of the material contained in this article were presented in my lectures ("Phrase Structure and Movement") at the 1997 Linguistic Society of America's Summer Linguistic Institute (Cornell University). I would like to thank the audience there for many valuable questions and suggestions. I am also grateful to the editors of this volume, an anonymous reviewer for the volume, Takao Gunji, and Heizo Nakajima for useful comments and suggestions. The research reported in this article was supported in part by the University of California's Pacific Rim Research Grant (PI: Naoki Fukui).

1 It is not implied here that phrase structure rules directly generate phrase markers. In fact, the standard assumption is that phrase structure rules generate "derivations," from which there is an algorithm to construct phrase markers. See Chomsky (1955, 1959) and especially McCawley (1968) for much relevant discussion on the nature of phrase structure rules and their relation to phrase markers.

2 It is now standard to call these elements the specifier of X″, rather than the specifier of X′, and, accordingly, notate them as [Spec, X″].

3 Ideas of this sort were explored and developed in structural linguistics in terms of discovery procedures of constituent analysis (Harris 1946, 1951).

4 The determiners (such as *the*, *a*, etc.) are also analyzed as [Spec, N′]. As it is hard to analyze determiners as maximal projections (noun phrases, in particular), the identification of determiners as Spec elements poses a problem for the uniform characterization of Spec discussed in the text. This problem was later resolved by the "DP-analysis," as we will see in the next section.

5 The X′-scheme in (25) is my interpretation of what is intended in the proposal of Chomsky (1986b). Chomsky's original formulation is as follows (Chomsky 1986b: 3):

X′-schemata:

a. $X' = X\ X''^{*}$

b. $X'' = X'''^{*}\ X'$ (where X^{*} stands for zero or more occurrences of some maximal projection and $X = X^{0}$)

The crucial difference between (25) and Chomsky's original formulation is that the latter allows "flat" and multiple branching structures at

both the single-bar and the double-bar levels, whereas the former (i.e., (25)), while permitting "recursion," never allows flat and multiple branching structures, in accordance with Kayne's (1984) binary branching hypothesis. It seems to me that the schemata in (25) express more properly what was intended by the proposal of Chomsky (1986b).

6 There are of course more complex cases. Whether or not the other linear ordering in the X'-scheme (viz., the Spec–head and head–adjunct order) is subject to parametrization is a complex issue that remains open. See among others Chomsky and Lasnik (1993) and references there for further discussion. We will return to the issue of the head parameter in section 3.

7 The converse is not implied in Chomsky's theory. That is, while X^{max}-movement (substitution) is always to a Spec position, it is not claimed that Spec is always a landing site for X^{max}-movement. Such a claim, which implies further sharpening of the notion of Spec, is in fact put forth in Relativized X'-theory, to be discussed in section 2 below.

8 Details differ in various analyses. For example, we put aside the issue of whether all the subjects of noun phrases are generated within a noun's projection, or some subjects are base generated in [Spec, D"]. There are other problems that remain open. See Longobardi in this volume.

9 Relativized X'-theory was first presented in a preliminary form in Fukui and Speas (1986), and was later developed, in slightly different ways and directions, in Fukui (1986) and Speas (1986, 1990). The following exposition is largely based on Fukui (1986).

10 The formal operation building the structure is assumed to be "Adjunction." Note that Adjunction here is somewhat different from the standard notion of adjunction, which, when applied, creates a multi-segment structure of the target. Adjunction, unlike adjunction, induces a projection of the target element (see the discussion in section 4 of this chapter). Note also that the notion of "bar-level" does not play any significant role in this theory. Thus, X' merely means that X is projected. See also Muysken (1982).

11 To the extent that "substitution" transformations and "adjunction" transformations must be distinguished with respect to their empirical properties, we have to make a distinction somehow, but differently from the traditional definitions. Fukui (1986) attempts to offer appropriate definitions of "substitution" and "adjunction" without having recourse to the empty category Δ, based on the idea that "substitution" is an operation that creates a legitimate structure licensed by (Relativized) X'-theory, whereas "adjunction" creates a structure that is never licensed at the base (in terms of X'-theory). See Fukui (1986: ch. 4). Note incidentally that under these definitions, some instances of Adjunction (see the preceding note), including, for example, scrambling in Japanese, which had been assumed to be adjunction (Saito 1985), should indeed be analyzed as substitution.

12 Notice that this operation includes the traditional "substitution" and the operation that is in charge of building structures, but does not include, perhaps, the traditional "adjunction," which creates a "non-standard" multi-segment structure.

See nn. 10 and 11. See also the discussion in the next section.

13 There are some potentially problematic cases for this claim. The "leftness condition" of Chomsky (1976), carried over into the principles-and-parameters approach in the form of "weak crossover," is one such.

14 Recall the term "substitution" loses its traditional meaning in the bare theory (as well as in Relativized X'-theory), since the dubious empty category Δ is eliminated from the theory of movement. See n. 11.

15 See Ishii (1997) for relevant discussion on this assumption. He proposes that it should be considered a consequence of a general principle on derivation, which he calls the immediate satisfaction principle. It is left open in Saito and Fukui (1998) whether an adjunct (modifier) is directly generated in an adjoined position, or moved there. Under either hypothesis, it follows that an adjunct (modifier) is checked for a feature (the "adverb feature") at the adjoined position, as proposed by Oka (1993) and Lee (1994). (See also Travis 1988 for relevant discussion.) Note finally that since adjunction cannot be a subcase of Merge, it is subject to the Last Resort Principle, as Saito and Fukui argue. See Saito and Fukui (1998) for other details.