

Bare Phrase Structure

Noam Chomsky
MIT

1. Some Leading Ideas in the Study of Language

This paper is an extension of earlier ones (Chomsky 1991, 1993) that were concerned with two related questions: (1) What conditions on the human language faculty are imposed by considerations of virtual conceptual necessity? (2) To what extent is the language faculty determined by these conditions, that is, how much special structure does it have beyond them? The first question in turn has two aspects: what conditions are imposed on the language faculty by virtue of (A) its place within the array of cognitive systems of the mind/brain, and (B) general considerations of simplicity, elegance, and economy that have some independent plausibility?

Question (B) is not precise, but not without content, as in the natural sciences generally. Question (A) has an exact answer, but only parts of it can be surmised, given what is known about related cognitive systems. To the extent that the answer to question (2) is positive, language is something like a 'perfect system,' meeting external constraints as well as can be done. The 'minimalist' program for linguistic theory seeks to explore these possibilities.

Any progress toward this goal will deepen a problem for the biological sciences that is already far from trivial: how can a system such as human language arise in the mind/brain, or for that matter, in the organic world, in which one seems not to find systems with anything like the basic properties of human language? That problem has sometimes been posed as a crisis for the cognitive sciences. The concerns are appropriate, but their locus is misplaced; they are a problem for biology and the brain sciences, which, as currently understood, do not provide any basis for what appear to be fairly well-established conclusions about language. Much of the broader interest of the detailed study of language lies right here, in my opinion.

The leading questions that guide the minimalist program came into view as the principles-and-parameters (P&P) model took shape. A look at recent history may be helpful in placing these questions in context; needless to say, these remarks are schematic and selective, and benefit from hindsight.

Early generative grammar faced two immediate problems: to find a way to account for the phenomena of particular languages ('descriptive adequacy'), and to explain how knowledge of these facts arises in the mind of the speaker-hearer ('explanatory adequacy'). Though it was scarcely recognized at the time, this research program revived the concerns of a rich tradition, of which perhaps the last major exponent was Otto Jespersen. Jespersen recognized that the structures of language "come into existence in the mind of a speaker" by abstraction from presented experience, yielding a "notion of structure" that is "definite enough to guide him in framing sentences of his own," crucially, "free expressions" that are typically new to speaker and hearer. These properties of language determine the primary goals of linguistic theory: to spell out clearly this "notion of structure" and the procedure by which it yields "free expressions," and to explain how it arises in the mind of the speaker the problems of descriptive and explanatory adequacy, respectively. To attain descriptive adequacy for a particular language L, the theory of L (its grammar) must characterize the state attained by the language faculty. To attain explanatory adequacy, a theory of language must characterize the initial state of the language faculty and show how it maps experience to the state attained. Jespersen held further that it is only "with regard to syntax" that we expect "that there must be something in common to all human speech"; there can be a 'universal (or general) grammar,' though "no one ever dreamed of a universal morphology."

In the modern period these traditional concerns were displaced in part by behaviorist currents and in part by various structuralist approaches, which radically narrowed the domain of inquiry while much expanding the data base for some future inquiry that might return to the traditional and surely valid concerns. To address them required a better understanding of the fact that language involves infinite use of finite means, in one classic formulation. Advances in the formal sciences provided that understanding, making it feasible to deal with the problems constructively. Generative grammar can be regarded

as a kind of confluence of long-forgotten concerns of the study of language and mind and new understanding provided by the formal sciences.

The first efforts to address these problems quickly revealed that traditional grammatical and lexical studies do not begin to describe, let alone explain, the most elementary facts about even the best-studied languages. Rather, they provide hints that can be used by the reader who already has tacit knowledge of language and of particular languages. This is hardly a discovery unique to linguistics. Typically, when questions are more sharply formulated, it is learned that even elementary phenomena had escaped notice and that intuitive accounts that seemed simple and persuasive are entirely inadequate. If we are satisfied that an apple falls to the ground because that is its natural place, there will be no serious science of mechanics. The same is true if one is satisfied with traditional rules for forming questions or with the lexical entries in the most elaborate dictionaries, none of which come close to describing simple properties of these linguistic objects.

Recognition of the unsuspected richness and complexity of the phenomena of language created a tension between the goals of descriptive and explanatory adequacy. It was clear that to achieve explanatory adequacy a theory of the initial state must hold that particular languages are largely known in advance of experience. The options permitted in universal grammar (UG) must be highly restricted; limited experience must suffice to fix them one way or another, yielding a state of the language faculty that determines the varied and complex array of expressions, their sound and meaning, in a uniform and language-independent way. But this goal receded still further into the distance as generative systems were enriched in pursuit of descriptive adequacy in radically different ways for different languages. The problem was exacerbated by the huge range of phenomena discovered when attempts were made to formulate actual rule systems.

This tension defined the research program of early generative grammar at least, the tendency within it that concerns me here. From the early 1960s its central objective was to abstract general principles from the complex rule systems devised for particular languages, leaving rules that are simple, constrained in their operation by these UG principles. Steps in this direction reduce the range of language-specific constraints, thus contributing to explanatory adequacy. They also tend to yield simpler and more natural theories, laying the groundwork for an eventual minimalist approach. These two aspects of inquiry are logically independent: it could turn out that an 'uglier' and richer version of UG reduces permissible variety, thus contributing to the primary empirical goal of explanatory adequacy. In practice, however, the two enterprises have proven to be mutually reinforcing and have progressed side by side.

These efforts culminated in the P&P model, which constituted a radical break from the rich tradition of thousands of years of linguistic inquiry, far more so than early generative grammar, which could be seen as a revival of traditional concerns and ways of addressing them (which is why it was more congenial to traditional grammarians than to modern structural linguists). The basic assumption of the P&P model is that languages have no rules at all in anything like the traditional sense and no grammatical constructions (relative clauses, passives, etc.) except as taxonomic artifacts. There are universal principles and a finite array of options as to how they apply (parameters). Furthermore, it may be that Jespersen's intuition about syntax-morphology can be captured, with parameters limited to the lexicon, indeed to a narrow part of it: functional categories. So I will henceforth assume.

The P&P model is in part a bold speculation rather than a specific hypothesis. Nevertheless, its basic assumptions seem reasonable in the light of what is currently at all well understood and do offer a natural way to resolve the tension between descriptive and explanatory adequacy.

If these ideas prove to be on the right track, there is a single computational system CHL for human language and only limited lexical variety. Variation of language is essentially morphological in character, including the critical question

of which parts of a computation enter the phonological component, a question brought to the fore by Jean-Roger Vergnaud's theory of abstract Case and James Huang's work on wh-constructions.

This account of the P&P approach overstates the case. Languages may vary in parts of the phonology that can be determined by readily available data, as well as in "Saussurean arbitrariness"; that is, the soundmeaning pairing for the substantive part of the lexicon. We put these matters aside, along with many others that appear to be computationally irrelevant; that is, not entering into CHL: among them, variability of semantic fields, selection from the lexical repertoire made available in UG, and nontrivial questions about the relation of lexical items to other cognitive systems.

Like the earliest proposals in generative grammar, formulation of the P&P model led to a huge expansion in empirical materials, by now from a wide variety of typologically different languages. The questions that could be clearly posed and the empirical facts with which they deal are novel in depth and variety, a promising and encouraging development in itself.

Insofar as the tension between descriptive and explanatory adequacy is reduced in this way, the problem of explanation becomes far harder and more interesting. The task is to show that the apparent richness and diversity of linguistic phenomena is illusory and epiphenomenal, the result of interaction of fixed principles under slightly varying conditions. And still further questions arise, namely, those of the minimalist program. How 'perfect' is language? One expects 'imperfections' in the formal part of the lexicon. The question is whether, or to what extent, this component of the language faculty is the repository of departures from virtual conceptual necessity, so that the computational system CHL is not only unique but optimal. Progress toward this further goal places a huge descriptive burden on the answers to the questions (A) and (B): the interface conditions and the specific formulation of general considerations of simplicity. The empirical burden, already severe in any P&P theory, now becomes extraordinary. The problems that arise are therefore extremely interesting. It is, I think, of considerable interest that we can at least formulate such questions today and even approach them in some areas with a degree of success. If the thinking along these lines is anywhere near accurate, a rich and exciting future lies ahead for the study of language and related disciplines.

2. The Minimalist Program

All these investigations have been based on several underlying factual assumptions. One is that there is a component of the human mind/brain dedicated to language the language faculty interacting with other systems. A more specific assumption is that there are just two such interacting systems: an articulatory-perceptual system A-P and a conceptual-intentional system C-I. The particular language L is an instantiation of the language faculty with options specified. L must therefore provide 'instructions' to be interpreted at these two interface levels. L is then to be understood as a generative system that constructs pairs (π, λ) that are interpreted at the A-P and C-I interfaces, respectively. π is a PF representation and λ an LF representation, each consisting of 'legitimate entities' that can receive some interpretation at the relevant level (perhaps interpretation as gibberish). A linguistic expression of L is at least a pair (π, λ) of this sort and under minimalist assumptions, at most such a pair, meaning that there are no 'levels of linguistic structure' apart from the two interface levels PF and LF; specifically, no levels of D-structure or S-structure.

We say that a computation (derivation) converges at one of the interface levels if it forms an interpretable representation in this sense, and converges if it converges at both interface levels, PF and LF; otherwise it crashes. We thus adopt the (nonobvious) hypothesis that there are no PF-LF interactions; similarly, that there are no conditions relating lexical properties and interface levels, such as the Projection Principle. The notion "interpretable" raises nontrivial questions, to some of which we return.

It seems that a linguistic expression of L cannot be defined just as a pair (π, λ) formed by a convergent derivation. Rather, its derivation must also be optimal, satisfying certain natural economy conditions, for example, conditions of locality of movement, no 'superfluous steps' in derivations, and so on. Less economical computations are 'blocked' even if they converge.

Current formulation of such ideas still leaves substantial gaps. It is, furthermore, far from obvious that language should have anything like the character postulated in the minimalist program, which is just that: a research program concerned with filling the gaps and asking how positive an answer we can give to question (2) of the first paragraph: how 'perfect' is language?

Suppose that this approach proves to be more or less correct. What could we then conclude about the specificity of the language faculty (modularity)? Not much. It could be that the language faculty is unique among cognitive systems, or even in the organic world, in that its principles satisfy minimalist assumptions. Furthermore, the morphological parameters could be unique in character. Another source of possible specificity of language is the conditions imposed at the interface, what we may call "bare output conditions." These will naturally reflect properties of the interface systems A-P and C-I (or whatever they turn out to be), but we have no idea in advance how specific to language these properties might be; quite specific, so current understanding suggests.

In brief the question of the specificity of language is not directly addressed in the minimalist program, except to indicate where it should arise: in the nature of the computational procedure CHL, in the properties of the bare output conditions and the functional component of the lexicon, and in the more obscure but quite interesting matter of conceptual elegance of principles and concepts.

It is important to distinguish the topic of inquiry here from a different one: to what (if any) extent are the properties of CHL expressed in terms of output conditions, say filters of the kind discussed in Chomsky and Lasnik (1977) or chain-formation algorithms in the sense of Rizzi (1986) in syntax or conditions of the kind recently investigated for phonology in terms of optimality theory (Prince and Smolensky 1993, McCarthy and Prince 1993)? A related question is whether CHL is derivational or representational in character: does CHL involve successive operations leading to (π, λ) (if it converges), or does CHL select two such representations and then compute to determine whether they are properly paired (or select one and derive the other)? The questions are rather subtle; typically, it is possible to recode one approach in terms of the other. But the questions are nevertheless empirical, turning basically on explanatory adequacy. Thus, filters were justified by the fact that simple output conditions sufficed to limit the variety and complexity of transformational rules, advancing the effort to reduce these to just Move- α (or Affect- α , in the sense of Lasnik and Saito 1984) and thus to move toward explanatory adequacy. Similarly, Rizzi's proposals about chain formation were justified in terms of the possibility of explaining empirical facts about Romance reflexives and other matters.

My own judgment is that a derivational approach is nonetheless correct, and the particular version of a minimalist program I am considering assigns it even greater prominence. There are certain properties of language, which appear to be fundamental, that suggest this conclusion. Under a derivational approach computation typically involves simple steps expressible in terms of natural relations and properties, with the context that makes them natural 'wiped out' by later operations and not visible in the representations to which the derivation converges. Thus, in syntax, head-movement is narrowly 'local,' but several such operations may leave a head separated from its trace by an intervening head, as when N incorporates to V leaving the trace t_N , and the $[V \text{ V-N}]$ complex then raises to I leaving the trace t_V , so that the chain (N, t_N) at the output level violates the locality property satisfied by each individual step. In segmental phonology, such phenomena are pervasive. Thus the rules deriving the alternants *decide-decisive-decision* from an invariant underlying lexical form are straightforward and natural at each step, but the relevant contexts do not appear at all in the output; given only output conditions, it is hard to see why decision, for example, should not rhyme with Poseidon on the simplest assumptions about lexical representations and optimal output conditions. Similarly, intervocalic spirantization and vowel reduction are natural and simple processes that derive, say, Hebrew *ganvu* 'they stole' from underlying *g-n-B*, but the context for spirantization is gone after reduction applies; the underlying form might even all but disappear in the output, as in *hitu* 'they extended', in which only the /t/ remains from the underlying root /ntC/ (C a 'weak' consonant).

In all such cases it is possible to formulate the desired result in terms of outputs. For example, in the head-movement case, one can appeal to the (plausible) assumption that the trace is a copy, so the intermediate V trace includes within it a record of the local $N \rightarrow V$ raising. But surely this is the wrong move. The relevant chains at LF are (N, t_N) and (V, t_V) , and in these the locality relation

observed by successive raising has been lost. Similar artifice could be used in the phonological examples, again improperly, it appears. These seem to be fundamental properties of language, which should be captured, not obscured. A fully derivational approach captures them and indeed suggests that they should be pervasive, as seems to be the case.

I will continue to assume that the computational system CHL is strictly derivational and that the only output conditions are the bare output conditions determined externally at the interface.

We hope to be able to show that for a particular language L, determined by fixing options in the functional part of the lexicon, the phenomena of sound and meaning for L are determined by pairs (π, λ) formed by maximally economical convergent derivations that satisfy output conditions. The computation CHL that derives (π, λ) must, furthermore, keep to natural computational principles (e.g., locality of movement) and others that are also minimalist in spirit. A natural version of this requirement is that the principles of UG should involve only elements that function at the interface levels; specifically, lexical elements and their features, and local relations among them. Let's adopt this proposal, sharpening it as we proceed.

In pursuing a minimalist program we want to make sure that we are not inadvertently 'sneaking in' improper concepts, entities, relations, and conventions. I assume that an item in the lexicon is nothing other than a set of lexical features, or perhaps a further set-theoretic construction from them (e.g., a set of sets of features), and that output conditions allow nothing beyond such elements. The point of the occasional forays into formalism below is to ensure that CHL keeps to these conditions, introducing no further elements, expressing only local relations, and deriving stipulated conventions where valid. Naturally the more spare the assumptions, the more intricate will be the argument.

3. The Computational System

A linguistic expression (π, λ) of L satisfies output conditions at the PF and LF interfaces. Beyond that, π and λ must be compatible: it is not the case that any sound can mean anything. In particular they must be based on the same lexical choices. We can then think of CHL as mapping some array of lexical choices A to the pair (π, λ) . What is A? It must at least indicate what the lexical choices are and how many times each is selected by C_{HL} in forming (π, λ) . Let us take a numeration to be a set of pairs (l, n) , where l is an item of the lexicon and n is its index, understood to be the number of times that l is selected. Take A to be (at least) a numeration N; C_{HL} maps N to (π, λ) . The procedure C_{HL} selects an item from N and reduces its index by 1, then performing permissible computations. C_{HL} does not converge unless all indices are zero.

If an item is selected from the lexicon several times by CHL, the choices must be distinguished; for example, two occurrences of the pronoun he may have entirely different properties at LF. l and l' are thus marked as distinct for CHL if these are two selections by CHL of a single lexical item.

We want the initial array A not only to express the compatibility relation between p and l, but also to fix the reference set for determining whether a derivation from A to (π, λ) is optimal, that is, not 'blocked' by a more economical derivation. Selection of the reference set is a delicate problem, as are considerations of economy of derivation generally. For the moment let us take N to determine the reference set, meaning that in evaluating derivations for economy we consider only alternatives with the same numeration. At least this much structure seems to be required; whether more is needed is a hard question. We return to the matter in section 7.

Given N, C_{HL} computes until it converges (if it does) at PF and LF with the pair (π, λ) . In a perfect language, any structure Σ formed by the computation hence π and λ is constituted of elements already present in the lexical elements selected for N; no new objects are added in the course of computation (in particular, no indices, bar-levels in the sense of X' theory, etc.). Let us assume that this is true at least of the computation to LF; standard theories take it to be radically false for the computation to PF.

Output conditions show that π and λ are differently constituted. Elements interpretable at the A-P interface are not interpretable at C-I, or conversely. At

some point, then, the computation splits into two parts, one forming π and the other forming λ . The simplest assumptions are (i) that there is no further interaction between these computations and (ii) that computational procedures are uniform throughout. We adopt (i) and assume (ii) for the computation from N to λ , though not for the computation from N to π ; the latter modifies structures (including the internal structure of lexical entries) by processes sharply different from those that take place before entry into the phonological component. Investigation of output conditions should suffice to establish these asymmetries, which I will simply take for granted here.

We assume, then, that at some point in the (uniform) computation to LF there is an operation Spell-Out that applies to the structure Σ already formed. Spell-Out strips away from Σ those elements relevant only to π , forming Σ_π and leaving Σ_λ , which is mapped to λ by operations of the kind used to form Σ . Σ_π is then mapped to π by operations unlike those of the $N \rightarrow \lambda$ mapping.

We then assume that each lexical entry is of the form $\{P, S, F\}$, where components of P serve only to yield π (phonological features), components of S serve only to yield λ (semantic features), and components of F (formal features, e.g., the categorial features $[\pm N, \pm V]$) may enter into computations but must be eliminated (at least by PF) for convergence. Since we take computation to LF to be uniform, there is no way to stipulate that elements of F are eliminable only after Spell-Out. But the mapping to PF has different properties and may contain rules that eliminate F-features in ways not permitted in the $N \rightarrow \lambda$ computation.

The lexical entry for "book," for example, might contain the phonological feature [begins with stop] stripped away by Spell-Out and mapped to PF, the semantic feature [artifact] that is left behind by Spell-Out, and the formal feature [nominal] that is both carried over by Spell-Out and left behind, interpreted at LF, and eliminated in the course of computation to PF, though relevant to its operation.

Let us assume further that Spell-Out delivers Σ to the module Morphology, which constructs wordlike units that are then subjected to further processes that map it finally to π . To fix terminology, let us continue to call the subsystem that maps the output of Morphology to PF the phonological component and the subsystem that continues the computation to LF after Spell-Out the covert component. Other terms are familiar but have had misleading connotations. I will have little to say about the phonological component here, except with regard to the matter of ordering of elements in the output (see section 6).

The simplest assumption is that Spell-Out can apply anywhere, the derivation crashing if a 'wrong choice' is made. After Spell-Out neither the phonological nor covert component can have any further access to the lexicon, a requirement for any theory, on the weakest empirical assumptions (otherwise sound-meaning relations would collapse). It is unnecessary to add stipulations to this effect. Because of the way C_{HL} is constructed, to which we return, an element selected from the lexicon cannot be embedded; hence the issue is narrow, arising only at the root of a phrase marker. If the phonological component adds a lexical item at the root, it will introduce semantic features, and the derivation will crash at PF. If the covert component does the same, it will introduce phonological features, and the derivation will therefore crash at LF.

Questions remain about lexical items lacking semantic or phonological features: can these be added at the root by the phonological or covert components, respectively? Empirical consequences seem to arise only in connection with functional elements that have 'strong features' in the sense of Chomsky (1993), that is, those that must be 'satisfied' before Spell-Out. Suppose that root C (complementizer) has a strong feature that requires overt wh-movement. We now want to say that unless this feature is checked before Spell-Out it will cause the derivation to crash at LF to avoid the possibility of accessing C after Spell-Out in the covert component. Slightly adjusting the account in Chomsky (1993), we now say that a checked strong feature will be stripped away by Spell-Out, but is otherwise ineliminable.

4. Phrase Markers in a Minimalist Framework

The development of X' theory in the late 1960s was an early stage in the effort to resolve the tension between explanatory and descriptive adequacy. A first step was to separate the lexicon from the computations, thus eliminating a serious

redundancy between lexical properties and phrase structure rules and allowing the latter to be reduced to the simplest (context-free) form. X' theory sought to eliminate such rules altogether, leaving only the general X'-theoretic format of UG. The problem addressed in subsequent work was to determine that format, but it was assumed that phrase structure rules themselves should be eliminable.

In the papers on economy and minimalism cited earlier, I took X' theory to be given, with specific stipulated properties. Let's now subject these assumptions to critical analysis, asking what the theory of phrase structure should look like on minimalist assumptions and what the consequences are for the theory of movement.

At the LF interface lexical items and their constituent features must be accessed. Accordingly, such items and their (semantic and formal) features should be available for C_{HL} . It is also apparent that some larger units constructed of these items are accessible, along with their types; for example, noun phrases are interpreted differently from verb phrases. Of the larger units only maximal projections seem to be relevant to LF-interpretation. If so, output conditions make the concepts minimal and maximal projection available to C_{HL} , and on minimalist assumptions, nothing else apart from lexical features. Minimal and maximal projections must be determined from the structure in which they appear without any specific marking; as proposed by Muysken (1982) they are relational properties of categories, not inherent to them. There are no such entities as XP (X^{max}) or X^0 (X^{min} , terminal element) in the structures formed by C_{HL} , though we may use these as informal notations, along with X' (X-bar) for any other category. Given a phrase marker, a category that does not project any further is a maximal projection XP and one that is not a projection at all is a minimal projection X^0 ; any other is an X', invisible at the interface and for computation.

We also hope to show that computation keeps to local relations of XP to terminal head. All principles of UG should be formulated in these terms, and only such relations should be relevant at the interface for the modules that operate there.

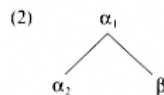
Given the numeration N, C_{HL} may select an item from N (reducing its index) or perform some permitted operation on the structures it has already formed. One such operation is necessary on conceptual grounds alone: an operation that forms larger units out of those already constructed; call it Merge. Applied to two objects α and β , Merge forms the new object γ . What is γ ? γ must be constituted somehow from the two items α and β ; the only alternatives are that γ is fixed for all α, β , or that it is randomly selected; neither is worth considering. The simplest object constructed from α and β is the set $\{\alpha, \beta\}$, so we take γ to be at least this set, where α and β are the constituents of γ . Does that suffice? Output conditions dictate otherwise; thus verbal and nominal elements are interpreted differently at LF and behave differently in the phonological component. γ must therefore at least (and, we assume, at most) be of the form $\{\delta, \{\alpha, \beta\}\}$, where δ identifies the relevant properties of γ ; call δ the label of γ .

The label must be constructed from the two constituents α and β . Suppose these are lexical items, each a set of features. Then the simplest assumption would be that the label is either (a), (b), or (c):

- (1) a. the intersection of α and β
 b. the union of α and β
 c. one or the other of α, β

The options (a) and (b) are immediately excluded: the intersection of α, β will generally be irrelevant to output conditions, often null, and the union will be not only irrelevant but contradictory if α, β differ in value for some feature, the normal case. We are left with (c): the label δ is either α or β ; one or the other projects and is the head of γ . If α projects, then $\gamma = \{\alpha, \{\alpha, \beta\}\}$.

For expository convenience we can map γ to a more complex object constructed from additional elements such as nodes, bars (primes, XP, etc.), subscripts, and other indices. Thus we might represent γ informally as (2) (assuming no order), where the diagram is constructed from nodes paired with labels and pairs of such labeled nodes, and labels are distinguished by subscripts:



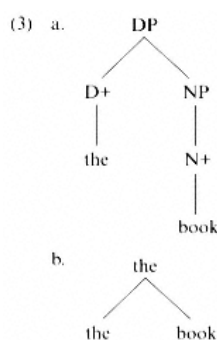
This, however, is informal notation only: empirical evidence would be required for postulation of the additional elements that enter into (2) beyond lexical features, and the extra sets.

The terms "complement" and "specifier" can be defined in the usual way, in terms of γ . The head-complement relation is the 'most local' relation of an XP to a terminal head Y, all others within YP being head-specifier (apart from adjunction, to which we return); in principle there might be a series of specifiers, a possibility that seems to be realized (see section 7). The principles of UG, we assume, crucially involve these local relations.

Further projections satisfy (1c) for the same reasons. We call these projections of the head from which they ultimately project, restricting the term "head" to terminal elements and taking complement and specifier to be relations to a head.

If constituents α, β of γ have been formed in the course of computation, one of the two must project, say α . At the LF interface γ (if maximal) is interpreted as a phrase of the type α (e.g., as a nominal phrase if its head k is nominal), and it behaves in the same manner in the course of computation. It is natural, then, to take the label of γ to be not a itself but rather k , the head of the constituent that projects, a decision that also leads to technical simplification. Assuming so, we take $\gamma = \{k, \{\alpha, \beta\}\}$, where k is the head of α , and its label as well, in the cases so far discussed. We keep to the assumption that the head determines the label, though not always through strict identity.

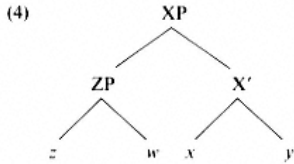
The operation Merge, then, is asymmetric, projecting one of the objects to which it applies, its head becoming the label of the complex formed. There can be no nonbranching projection. In particular there is no way to project from a lexical item a subelement $H(\alpha)$ consisting of the category of α and whatever else enters into further computation, $H(\alpha)$ being the actual 'head' and α the lexical element itself; nor can there be such 'partial projections' from larger elements. We thus dispense with such structures as (3a) with the usual interpretation: the, book taken to be terminal lexical items and D+, N+ standing for whatever properties of these items are relevant to further computation (perhaps the categorial information D, N; Case; etc.). In place of (3a) we have only (3b):



Standard X' theory is thus largely eliminated in favor of bare essentials.

Suppose that the label for $\{\alpha, \beta\}$ happens to be determined uniquely for α, β in language L; we would then want to deduce that fact from properties of α, β, L ; or, if it is true for α, β in language generally, from properties of the language faculty, and similarly if the label is uniquely determined for arbitrary α, β, L , as may be the case.

Suppose that we have the structure that we represent informally as (4), with x, y, z, w terminals:



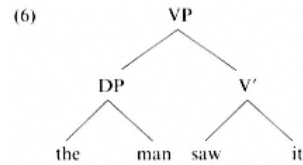
Here $ZP = \{z, \{z, w'\}\}$, $X' = \{x, \{x, y'\}\}$, $XP = \{x, \{ZP, X'\}\}$. Note that w and y are both minimal and maximal; z and x are minimal only.

The functioning elements in (4) are at most the nodes of the informal representation: that is, the lexical terminals z, w, x, y ; the intermediate element X' and its sister ZP ; and the 'root element' XP standing for the full structure formed. In the formal representation the corresponding elements are z, w, x, y ; $\{x, \{x, y'\}\} = P$ and its sister $\{z, \{z, w'\}\} = Q$; and the root element $\{x, \{P, Q\}\}$. These alone can be functioning elements; call them the terms of XP . More explicitly, for any structure K ,

- (5)
- K is a term of K
 - if L is a term of K , then the members of the members of L are terms of K

Terms correspond to nodes of the informal representations, where each node is understood to stand for the subtree of which it is the root.

In (4), x is the head of the construction, y its complement, and ZP its specifier. Thus (4) could be, say, the structure VP with the head *saw*, the complement *it*, and the specifier *the man* with the label *the*, as in (6):



Here $V' = VP = \text{saw}$ and $DP = \text{the}$.

Note that this very spare system fails to distinguish unaccusatives from unergatives, a distinction that seems necessary in the light of differences of behavior that have been discovered. The simplest solution to the problem would be to adopt the proposal of Hale and Keyser (1993) that unergatives are transitives; I will assume so.

The structure (6) will yield the sentence "the man saw it" when further inflectional elements are added by Merge and the specifier of the VP is raised (assuming this particular form of the predicate-internal subject hypothesis, as I will throughout). This example involves the second operation that forms phrase markers, the operation Move (Move- α). Given the phrase marker Σ with terms K and α , Move targets K , raises α , and merges α with K to form the new category with the constituents α, K . This is the operation of substitution, a term borrowed from earlier theory that is now somewhat misleading, though we continue to use it. As matters now stand, either K or α could project. The operation forms the chain (α, t) ; α c-commands t , which is a copy of α . We return to adjunction. The only other operation is Delete (Delete- α), which leaves the structure unaffected apart from some indication that α is not 'visible' at the interface.

In the case of (6) *the man* raises overtly (pre-Spell-Out), targeting the highest Agr phrase, while *it* raises covertly, targeting the lowest Agr phrase, each of the raised elements becoming the specifier of the targeted category. As noted, there is another option: the raised phrase itself might have projected, so that the targeted Agr phrase would become the specifier of the raised nominal phrase, which would now be a D' , not a DP , since it projects to a higher phrase. In preminimalist work this obviously unwanted option was excluded by conditions on transformations and stipulated properties of X' theory. But we no longer can – or wish to – make recourse to these, so we hope to show that the conventional assumption is in fact derivable on principled grounds – that it is impossible for Move to target K raising α , then projecting α rather than K . For the case of substitution the result is immediate within the minimalist framework under the strong formulation of the principle Greed of Chomsky (1993), which licenses

movement of α only as a step toward satisfying one of its own properties. If α raises and merges with K , then projecting, α is now an X' category, not X^{\max} . Therefore it can neither enter into a checking relation nor be moved further, being invisible to the computational system C_{HL} . Accordingly, the raising cannot satisfy Greed, and the unwanted option is excluded.

I will assume that Greed holds in this strong form:

- (7) Move raises α to a position β only if morphological properties of α itself would not otherwise be satisfied in the derivation.

Thus Greed cannot be overridden for convergence. We cannot, for example, derive (8a) by raising, violating Greed to satisfy EPP (the strong DP-feature of Infi), and (8b) cannot be interpreted as something like (c), with covert raising:

- (8)
- *It is believed [a man to seem to t that ...]
 - *There seem to a lot of us that ...
 - It seems to a lot of us that ...

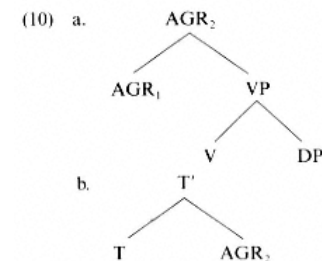
Similarly, DP cannot raise to $[Spec, VP]$ to assume an otherwise unassigned Θ -role. There can be no words "hit" or "believe" with the Θ -structure of hit, believe but no Case features, with John raising as in (9) to pick up the Θ -role, then moving on to $[Spec, Infi]$ to check Case and agreement features:

- (9)
- John [_{VP} t' [HIT t]]
 - John [_{VP} t' [BELIEVE [t to be intelligent]]]

The only possibility is direct raising to $[Spec, Infi]$ so that the resulting sentences "John HIT" and "John BELIEVES to be intelligent" are deviant, lacking the external argument required by the verb. We thus have a partial analogue to the P&P principle that there is no raising to a Θ -position. And we can assume that, for substitution at least, it is the target that projects. We will see that there are independent reasons for the same conclusion.

Raising of a targeting K is barred by (7) unless some property of α is satisfied by its moving to, or through, this position, and that property would not have been satisfied had this operation not applied; there is no movement to or through a position unless that operation is necessary in this sense. Consistent with Greed, such movement would be permitted if there were no other way for α to reach a position where its features would eventually be satisfied. Suppose (as we assume) that movement is constrained by a 'minimal link condition' (MLC), meaning that α must make the 'shortest move' in a sense that must be defined. That could in principle require movement through a position that satisfies no properties of α . The situation should not arise in the kind of case just discussed: a substitution operation that 'creates' a new position, $[Spec, K]$, by raising of α . It might well occur, however, in the case of adjunction satisfying the MLC.

Suppose that the structure K targeted by Move is a proper substructure of Σ . Thus suppose the operation is covert raising of the object to $[Spec, AgrO]$ for Case and object agreement. Prior to this operation we have (in informal notation) the structure (10a) embedded in the larger structure (10b):



Here T' is $\{T, \{T, K\}\}$, where K (namely (a)) is $\{Agr, \{Agr, VP\}\}$, $VP = \{V, \{V, DP\}\}$. If we target K , merging DP and K and projecting Agr as intended, we form (11) with the raised DP the specifier of $AgrP$ (Agr^{\max}):



Here AgrP is {Agr, {DP, K}} = L, and the term T' immediately dominating it is {T, {T, L}}, not {T, {T, K}} as it was before Move raised DP. Note that labels do not change, only constituents, if it is the target that projects, not the raised category.

It remains to extend the discussion to adjunction, forming a two-segment category. That adjunction and substitution both exist is not uncontroversial; thus, Lasnik and Saito (1992) adopt only the latter option while Kayne (1994) adopts (virtually) only the former. Nevertheless, I will assume here that the distinction is necessary; that is, that specifiers are distinct in properties from adjuncts and, generally, A- from A'-positions.

We have so far considered operations that form $Q = \{k, \{\alpha, K\}\}$, where k is the head (= the label) of the projected element K . But we now have a second option, in which Q is a two-segment category, not a new category. Plainly we need a new object constructed from K but not identical with its head k . The minimal choice is the ordered pair $\langle k, k \rangle$. We thus take $Q = \{\langle k, k \rangle, \{\alpha, K\}\}$. Note that $\langle k, k \rangle$, the label of Q , is not a term of the structure formed. It is not identical to the head of K , as before, though it is constructed from it in a trivial way.

Suppose that we adjoin α to K where K is embedded. For substitution we were able to derive the conventional assumption that the target projects, not the raised element. The argument given does not carry over to adjunction, but let us adopt the convention for the moment, returning to the matter. Thus when α is adjoined to K , the resulting structure is necessarily $[K, K] = \{\langle k, k \rangle, \{\alpha, K\}\}$, which replaces K in a structure containing K . Recall that it is the head that projects; the head either is the label or, under adjunction, determines it.

Adjunction differs from substitution, then, only in that it forms a two-segment category rather than a new category. Along these lines the usual properties of segments vs. categories, adjuncts versus specifiers, are readily formulated.

The bare theory outlined here departs from conventional assumptions in several respects. One is that an item can be both an X^0 and an XP. Does this cause problems? Are there examples that illustrate this possibility? I see no particular problems, and one case comes to mind as a possible illustration: clitics. Under the DP hypothesis clitics are Ds. Assume further that a clitic raises from its theta-position and attaches to an inflectional head. In its theta-position, the clitic is an XP; attachment to a head requires that it be an X^0 (on fairly standard assumptions). Furthermore, the movement violates the head-movement constraint (HMC, a tentative posit that we hope to derive), indicating again that it is an XP, raising by XP-adjunction until the final step of X^0 -adjunction. Clitics appear to share XP and X^0 properties, as we would expect on minimalist assumptions.

If the reasoning sketched so far is correct, phrase structure theory is essentially a given on grounds of virtual conceptual necessity in the sense indicated earlier. The structures stipulated in earlier versions are either missing or reformulated in elementary terms satisfying minimalist conditions, with no objects beyond lexical features. Stipulated conventions are derived (with a gap yet to be filled, for adjunction). Substitution and adjunction are straightforward. We return to further consequences.

5. Properties of the Transformational Component

We have so far considered two operations, Merge and Move, the latter with two cases, substitution and adjunction. The operation Merge is inescapable on the weakest interface conditions, but why should the computational system C_{HL} in human language not be restricted to it? Plainly, it is not. The most casual inspection of output conditions reveals that items commonly appear overtly 'displaced' from the position in which they are interpreted at the LF interface. This is an irreducible fact of human language, expressed somehow in every theory of language, however the displacement may be concealed in notation; it is also a central part of traditional grammar, descriptive and theoretical, at least back to the Port Royal Logic and Grammar. The only question is: what is the

nature of these transformational devices (whether one chooses to call them that or not)? On minimalist assumptions we want nothing more than an indication at LF of the position in which the displaced item is interpreted; that is, chains are legitimate objects at LF. Since chains are not introduced by selection from the lexicon or by Merge, there must be another operation to form them: the operation Move.

In the early days of generative grammar speculation about this matter invoked parsing and semantic considerations improved parsability on certain assumptions, the separation of theme-rheme structures from base-determined semantic (theta) relations, etc. (see Miller and Chomsky 1963, Chomsky 1965 for review). The minimalist framework, with the disparity it imposes between θ -role assignment and feature-checking, requires such an operation, as Kenneth Wexler has observed. Our concern here is to ask how spare an account of Move the facts of language allow.

This question was a second focus of the effort to resolve the tension between descriptive and explanatory adequacy alongside the steps that led to X' theory. A central concern was to show that the operation Move- α is independent of α ; another was to restrict the variety of structural conditions for transformational rules. These efforts were motivated by the dual concerns discussed earlier: the empirical demands posed by the problems of descriptive and explanatory adequacy, and the conceptual demands of simplicity and naturalness. Proposals motivated by these concerns inevitably raise the new leading problem that replaces the old: to show that restriction of the resources of linguistic theory preserves (and we hope, even enhances) descriptive adequacy while explanation deepens. The efforts have met with a good deal of success, though minimalist assumptions would lead us to expect more.

Consider first the independence of Move- α from choice of α . While this currently seems a reasonable supposition, it has so far been necessary to distinguish various kinds of movement: XP-movement from X^0 -movement; and among XPs, A-movement from A'-movement. Various kinds of 'improper movement' are ruled out, essentially by stipulation; for example, head-raising to an A'-position followed by raising to Spec. A further goal would be to eliminate any such distinctions, demonstrating on general grounds, without any special assumptions, that the 'wrong kinds' of movement crash; not an easy problem.

Some of the general constraints introduced to reduce the richness of descriptive apparatus (hence the variety of transformations) also have problematic aspects. Consider Emonds's structure-preserving hypothesis (SPH) for substitution operations (Emonds 1969). As has been stressed particularly by Jan Koster, it introduces an unwanted redundancy in that the target of movement is somehow 'there' before the operation takes place; that observation provides one motive for nonderivational theories that construct chains by computation on LF (or S-structure) representations. The minimalist approach overcomes the redundancy by eliminating the SPH: with D-structure gone, it is unformulable, its consequences derived –we hope to show– by the general properties of Merge and Move.

It has also been proposed that something like the SPH holds of adjunction: thus, heads adjoin to heads and XPs to XPs. This extended SPH introduces no redundancy and is not affected by the minimalist program, though we would like to deduce it from more elementary considerations.

The descriptive facts are not entirely clear, but they might be something like this. Suppose that only YP can adjoin to XP, and that pre-Spell-Out only Y^0 can adjoin to X^0 , though covert operations may adjoin YP to X^0 ; for example, VP-adjunction to causative V. We then have two problems: (i) to explain why the SPH holds at all, and (ii), if it does, why it differs before and after Spell-Out, apparently violating the (optimal) uniformity assumption on C_{HL} .

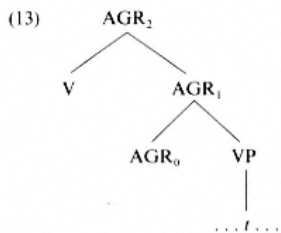
The answer to the second problem may lie in the nature of the Morphological component. Recall that at Spell-Out, the structure Σ already formed enters Morphology, a system that presumably deals only with wordlike elements, which we may take to be X^0 's –that is, either an item selected from the lexicon (hence with no constituents) or such an item with an element adjoined to it (hence $\langle a, a \rangle$, $\{a, b\}$):

- (12) Morphology gives no output (so the derivation crashes) if presented with an element that is not an X^0 .

On this natural assumption the largest phrases entering Morphology are X^0 's; if some larger unit appears within an X^0 , the derivation crashes. The pre- versus post-Spell-Out asymmetry follows.

It remains to explain why Y^0 adjoins only to X^0 ; for example, why can a verb not raise from VP and adjoin to an XP, escaping HMC (a tentative posit, which we hope to derive), and then move on to adjoin to an Infl element? Recall that we assume that one case of Y^0 -adjunction to XP is legitimate, namely, when Y^0 is also a YP, as in the case of clitics. The question arises for nonmaximal Y^0 .

Consider more closely the case we want to exclude, say, extraction of V from VP, adjoining to the higher AgrP:



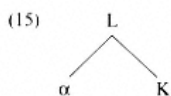
Here Agr₂ and Agr₁ are segments of a category formed by adjunction of V to Agr.

In (13) the V head of the chain (V, t) is V^{\max} (VP), by definition. But its trace is not. A natural requirement is that chains meet the uniformity condition (14), where the 'phrase structure status' of an element is its (relational) property of maximal, minimal, or neither:

(14) A chain is uniform with regard to phrase structure status.

Adjunction of nonmaximal α to XP is therefore blocked, including (13). The argument carries over to substitution: quite generally, we cannot raise nonmaximal α targeting K, where K then projects. For example, D cannot raise from (nontrivial) DP to [Spec,Infl] (subject) position, leaving the residue of the DP. These consequences of the SPH for substitution and adjunction have commonly been stipulated; they follow from (14), within the bare theory.

We have seen that when Move raises α targeting K to form a new category L (substitution), then the target K must project, not α . The uniformity condition provides an independent reason for this conclusion when α is maximal. If α raises and projects, it will be an X' , not an X_{\max} , so the chain will violate (14). There is also an independent reason for this conclusion when α is nonmaximal. Suppose that (15) is formed by Move, raising α to attach to target K, forming L, a projection of α :



Since α is nonmaximal, the operation is head movement. By HMC it cannot have crossed any intervening head, from which it follows that K can only be the projection of α itself. Suppose that such 'self-attachment' is ruled out on principled grounds (we return to the matter directly). If so, it must be the target that projects.

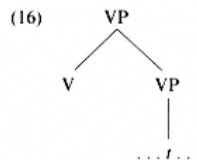
Note that this argument holds both for substitution and adjunction. We therefore have several independent arguments that the target projects under substitution and an argument that the same is true for adjunction when the raised element is nonmaximal.

Consider more closely the general case of adjunction, as in (15), with L a segment. Suppose that L is projected from α , the case we wish to exclude. We have to determine what the head of the chain formed by the adjunction operation is: is it α , or the two-segment category $[\alpha, \alpha]$? The latter choice is ruled out by (14). But the former leaves us with a category $[\alpha, \alpha]$ that has no interpretation at LF, violating Full Interpretation (FI) (the same problem would

have arisen, this time for α , had we taken the head of the chain to be $[\alpha, \alpha]$). Again, we conclude that the target must have projected.

The asymmetry of projection after movement thus seems to have solid grounds: it is only the target that can project, whether movement is substitution or adjunction.

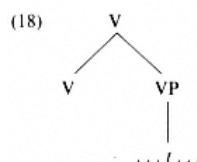
One strand of the argument was based on the assumption that self-attachment is impermissible, as in (16):



Thus suppose we have the VP "read the book" and we adjoin to it the head "read," forming the two-segment category [read [t the book]]. Under the intended interpretation of (16), with the target projected, we have formed the object (17), where γ is the target VP = {read, read, the book} (omitting further analysis):

(17) {<read,read>, {read, γ }}

Suppose, however, that we had projected the adjunct V ("read") in (16), yielding (18):

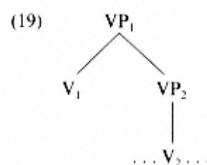


But this too is an informal representation of (17), just as (16) is, though the intended interpretations differ: in (16) we have projected the target, in (18) the adjunct. Furthermore, the latter interpretation should be barred.

Note that the problem is not limited to adjunction. Suppose that we raise the head N of NP to [Spec,NP]. Then in exactly the same way we will construct the same formal object whether we think of NP or Spec as projecting.

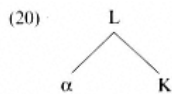
We might conclude that this is exactly the right result, with such ambiguity interpreted as a crashed derivation. Then such operations of 'self-attachment' (whether adjunction or substitution) are barred outright, as appears to be the case, incidentally filling the gap in the argument that the target projects under head raising.

Let's turn now to the case of raising of V (= V₂) in a Larsonian shell, as in (19):



Since self-attachment is ruled out, the operation cannot have targeted VP₂ either as adjunction or substitution. It must be, then, that VP₁ is not a projection of the raised verb V₂ but rather a VP distinct from VP₂, as stipulated in earlier accounts. Thus, V₂ raises to an already filled position occupied by a 'light verb' v that has been selected from the lexicon and heads its own projection, VP₁. V₂ adjoins to v forming [_v V₂ v]; the V₁ position is not 'created' by the raising operation. That conclusion, in fact, is independently imposed by economy conditions, which permit the raising of V₂ only to satisfy its morphological properties (Greed). Raising of V is legitimate only if the operation is necessary for satisfaction of some property of V. That would not be the case if VP₁ were a projection of V₂, but can be if it is a projection of V₁, to which V₂ adjoins, the complex then raising to satisfy morphological properties of V₂.

We now have several conclusions about chains. Consider again the basic structure (20), where α is the head of the chain $CH = (\alpha, t)$:



Whether the operation OP forming CH is substitution or adjunction, L is a projection of K, not α . If OP is substitution, then t is an X^{max} . Suppose OP is adjunction. If K is maximal, then α is maximal; pure heads can only adjoin to pure heads. If K is nonmaximal, then α too must be nonmaximal if OP is overt (pre-Spell-Out), though it can be maximal if OP takes place in the covert component, not entering Morphology. We eliminate the SPH for adjunction as well as for substitution. Furthermore, a number of the cases of 'improper movement' are eliminated, carrying us a step forward toward a bare principle Move with no breakdown into various kinds of movement.

We have so far sidestepped a problem that arises in the case of normal head adjunction. Take K to be a nonmaximal head in (20) and α to be a head. Since K projects to L, α is maximal. Thus α is both maximal and minimal. If that is true of t as well (e.g., the case of clitic-raising), then CH satisfies the uniformity condition (14). But suppose t is nonmaximal, as in the case of V-raising to Infl or to V. Then under a natural (though not necessary) interpretation (14) is violated; CH is not a legitimate object at LF, and the derivation will crash. That is obviously the wrong result. We might therefore assume that at LF, wordlike elements are 'immune' to the algorithm that determines phrase structure status:

- (21) At LF, X^0 is submitted to independent word-interpretation processes WI,

where WI ignores principles of C_{HL} within X^0 . WI is something like a covert analogue of Morphology, except that we expect it to be compositional, unlike Morphology, on the assumption that the $N \rightarrow LF$ mapping is uniform throughout.

Suppose that K in (20) is a maximal head and α a pure (nonmaximal) head; thus this is a case of head adjunction in which the target happens to be a maximal projection. The status of this case depends on the precise interpretation of (21). But the question need not concern us, since the case cannot arise: for reasons of c-command, a pure head α must raise from within the target K.

Suppose that (20) is formed by adjunction, so that L and K are segments, with $L = K$. So far, there are two ways in which (20) could have been formed: by strict merger of α , K (without movement), or by raising of α , forming the chain CH, α then merging with K. In either case, we form the structure $\gamma = \langle k, k \rangle, \{\alpha, K\}$ with the three terms α , K, γ ; k the head of K. Each of these is a category that is 'visible' at the interface, where it must receive some interpretation, satisfying FI. The adjunct α poses no problem. If it heads CH, it receives the interpretation associated with the trace position; if it is added by strict merger, it would presumably be a predicate of K (e.g., an adverbial adjunct to a verb). But there is only one role left at LF for K and γ . Note that the label $\langle k, k \rangle$ is not a term, hence receives no interpretation.

If γ is nonmaximal, the problem is obviated by (21) under a natural interpretation of WI. This should suffice to account for, say, noun incorporation to verbs or verb incorporation to causatives; the same would extend to VP incorporation to V if the LF interface permits such word structures (unlike Morphology). Furthermore, the target K in such cases often lacks any independent function, for example, an affix lacking a Θ -role. In these cases only α or the chain it heads is interpreted, and FI is satisfied.

Suppose γ is nonminimal. We now have two terms, γ and K, but only one LF role. The structure is still permissible if K lacks a Θ -role, as in the case of covert adjunction to an expletive (independently of (21)). The only other possibility is that the adjunct α is deleted at LF, leaving just K. When would this take place?

One case is when α is the trace of successive-cyclic movement of the type that permits intermediate trace deletion, say, along the lines sketched by Chomsky and Lasnik (1993) in terms of a principle of economy of deletion for non-uniform

chains; for example, wh-movement to [Spec, CP] with intermediate adjunction, as in (22):

- (22) Which pictures of John's brother did he expect that [t' [you would buy t]]

Another case is full reconstruction at LF, eliminating the adjunct entirely, thus a structure of the type (23) interpreted only at the trace:

- (23) [_{YP} XP [_{YP} ... t ...]]

It follows that 'scrambling' is permissible only if it is interpreted by reconstruction, as is argued to be the case by Saito (1989 and subsequent work). Similarly, it would follow that such constructions as (24) must be Condition C violations (under the relevant interpretation), and we predict a difference in status between (25) and (22), the latter escaping the violation because the head of the chain is not an adjunct:

- (24) a. Meet John in England, he doesn't expect that I will
b. Pictures of John, he doesn't expect that I will buy

- (25) Pictures of John's brother, he never expected that you would buy

The conclusions are plausible as a first approximation, though we enter here into a morass of difficult and partially unsolved questions of a kind discussed by Barss (1986), Freidin (1986), Lebeaux (1988), and earlier work; see Chomsky (1993) for some discussion.

On strictly minimalist assumptions these should be the only possibilities for adjunction; namely, (26):

- (26) a. word formation
b. semantically vacuous target (e.g., expletive-adjunction)
c. deletion of adjunct (trace deletion, full reconstruction)

In particular, apart from (c) there will be no adjunction to a phrase that assigns or receives a semantic role (e.g., a Θ -role assigner or an argument, a predicate or the XP of which it is predicated). Since (c) is irrelevant to strict merger, the options for the current counterpart to 'base adjunction' are even narrower. We consider adjoined adverbials further in section 7.

Adjunction is therefore an option, but a limited one with rather special properties, under natural minimalist assumptions.

In these terms we might return to the problem of improper movement. We want to show that the wide variety of such cases is excluded on principled grounds. Some fall into place. Thus, standard cases such as (27) cause no problem in the minimalist framework:

- (27) *John is illegal [_{IP} t² [_{IP} t¹ to leave]]

The complement of illegal permits PRO (it is illegal to leave) so that (null) Case is assigned to the subject of the infinitive and further raising is barred by Greed, under the strong interpretation (7) (see Martin 1992).

Consider cases of the type (28), with t² adjoined to IP in (a) and in [Spec, AgrP] in (b):

- (28) a. *John seems [that [t² [it was told t¹ [that ...]]]]
b. *Why do you wonder whether John said that Bill [t² [left t¹]]

Here we do not want to permit the intermediate (offending) trace t² to delete, unlike (22). The distinction suggests a different approach to intermediate trace deletion: perhaps it is a reflex of the process of reconstruction, understood in minimalist terms as in Chomsky (1993). The basic assumption here is that there is no process of reconstruction; rather, the phenomenon is a consequence of the formation of operator-variable constructions driven by FI, a process that may (or sometimes must) leave part of the trace, a copy of the moved element intact at LF, deleting only its operator part. The reconstruction process would then be restricted to the special case of A'-movement that involves operators. Some other cases of improper movement also can be eliminated along lines considered

here, for example, XP-movement passing through or adjoining to a pure Y^0 position, the trace then deleting. The general topic merits a comprehensive review.

So far, we have kept to the minimalist assumption that the computational procedure C_{HL} is uniform from N to LF; any distinction pre- and post-Spell-Out is a reflex of Morphology within the phonological component. I have said nothing so far about the 'extension condition' of Chomsky (1993), which guarantees cyclicity. The condition is motivated for substitution pre-Spell-Out by relativized minimality effects (in the sense of Rizzi 1990) and others, and does not hold post-Spell-Out if the Case-agreement theory of the minimalist approach is correct. It also cannot hold for adjunction, which commonly (as in head-adjunction) targets an element within a larger projection. We would like to show that these consequences are deducible, not stipulated.

With regard to Merge there is nothing to say; it satisfies the extension condition by definition. Questions arise only in connection with Move. Move targets K, raising α to adjoin to K or to be the specifier of K, K projecting in either case. K may be a substructure of some structure L already formed. That is a necessary option in the covert component but not allowed freely pre-Spell-Out as a result of other conditions, we hope to show.

There are several cases of pre-Spell-Out cyclicity to consider. One is of the type illustrated by such standard examples as (29):

(29) *Who was [α a picture of t_{wh}] taken t_α by Bill

This is a CED violation if passive precedes wh-movement, but it is derivable with no violation (incorrectly) if the operations apply in countercyclic order, with passive following wh-movement. In this case natural economy conditions might make the relevant distinction between the competing derivations. Passive is the same in both; wh-movement is 'longer' in the wrong one in an obvious sense, object being more 'remote' from [Spec,CP] than subject in terms of number of XPs crossed. The distinction should be captured by a proper theory of economy of derivation—though the general problem is nontrivial.

The relativized minimality cases fall into three categories: (a) head movement (the HMC), (b) A-movement, (c) A'-movement. In each case we have two situations to rule out: (I) skipping an already filled position; (II) countercyclic operations, that is, movement that skips a 'potential' position that is later filled. Situation (I) may fall under the Minimal Link Condition (MLC) (no innocuous assumption). As for (II), category (a) is not a problem; as we have seen, head-insertion is necessarily by pure merger, which satisfies the extension condition. The remaining cases to be excluded are countercyclic derivations in which an XP is raised to some Spec crossing the position of a lower Spec that is introduced later by movement. It is not easy to construct examples that are not ruled out in one or another independent way, but a closer look is plainly necessary.

It may be then that there is no need to impose the extension condition of Chomsky (1993) on overt operations. Furthermore, neither the phonological nor covert component can access the lexicon for reasons already discussed. The Morphology module indirectly allows variation before and after Spell-Out, as does strength of features. It seems possible to maintain the preferred conclusion that the computational system C_{HL} is uniform from N to LF, in that no pre- versus post-Spell-Out distinction is stipulated.

6. Order

Nothing has yet been said about ordering of elements. There is no clear evidence that order plays a role at LF or the computation from N to LF. Let us assume not. It must be then that ordering is part of the phonological component, a proposal that has been put forth over the years in various forms. It seems natural to suppose that ordering applies to the output of Morphology, assigning a linear (temporal, left-to-right) order to the elements it forms, all of them X^0 s, though not necessarily lexical elements.

The standard assumption has been that order is determined by the head parameter: languages are head initial (English) or head final (Japanese), with further refinements possible. Fukui (1993) has proposed that the head parameter provides an account of optional movement, which otherwise is excluded under

economy conditions, except in special cases when alternative derivations are equally economical. He argues that movement that maintains the ordering of the head parameter is 'free'; other movement must be motivated by Greed ('last resort'). Thus, in head-final Japanese, leftward movement (scrambling, passive) is optional, while in English such operations must be motivated by feature checking; and in head-initial English rightward extraposition is free, though barred in Japanese.

Kayne (1994) has advanced a radical alternative to the standard assumption, proposing that order reflects structural hierarchy universally. Specifically, he proposes the Linear Correspondence Axiom (LCA), which states that asymmetric c-command imposes a linear ordering of terminal elements; any phrase marker that violates this condition is barred. From his specific formulation of LCA he draws the further conclusions that there is a universal Spec-head-complement (SVO) ordering, and that specifiers are in fact adjuncts. A head-complement structure then is necessarily an XP, which can be extended—exactly once, on Kayne's assumptions—to a two-segment XP. The proposal is very much in the spirit of the minimalist program. Let's consider how it might be incorporated into the bare phrase structure theory just outlined. That is not an entirely straightforward matter, because the bare theory lacks much of the structure of the standard X' theory that Kayne adopts and partially reduces to LCA.

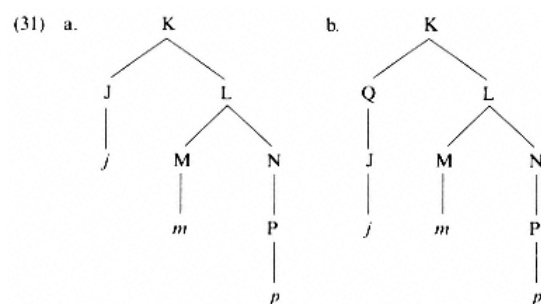
Kayne offers two kinds of arguments for LCA: conceptual and empirical, the latter extended in subsequent work (see particularly Zwart 1993). The conceptual arguments show how certain stipulated properties of X' theory can be derived from LCA. The empirical arguments can largely be carried over to a reformulation of LCA within the bare theory, but the conceptual ones are problematic. First, the derivation of these properties relies crucially not just on LCA, but on features of the standard X' theory that are abandoned in the bare theory. Second, the conclusions are for the most part already derivable in the bare theory without LCA, though in somewhat different form.

Kayne adopts the standard X'-theoretic assumptions (30), illustrated for example in (3), above:

(30) Certain features (categorial features) project from a terminal element to form a head, then on to form higher categories with different bar levels.

The conceptual arguments and the conclusions about ordering crucially rely on these assumptions, which are abandoned in the bare theory.

To illustrate, consider two elementary structures central to Kayne's account (his (4), (13), notations modified):



Putting aside the status of K for the moment, consider first (31a). Here j, m, p are terminals, selected from the lexicon. They project to J, M, P, respectively, the X^0 categories that are the actual heads. P then projects to the X^{\max} category N, which combines with M to form the higher projection L. The categories are J, M, P (heads) and L, N (maximal projections); a tacit assumption is that projection to higher categories in (30) is optional. The c-command relation is stipulated to hold only of categories (not terminals). Asymmetric c-command (ACC) holds between J and M (irrelevantly, also J and N, P) and between M and P. Accordingly, the terminals dominated by these categories are assigned the linear ordering j-m-p, and the structure is admissible under LCA.

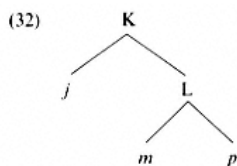
In (31b) there is a new category Q, an Xmax projected from the head J. J does not c-command at all; Q c-commands M (and N, P) and L c-commands J,

asymmetrically in both cases. The ACC relations do not yield a linear ordering: ACC(Q, M) entails that *j* precedes *m*, which precedes *p* as before; ACC(L, J) entails that *m* precedes *j*.

Therefore (31a) is admissible under LCA and (31b) is not. Turning to K, in (31a) the structure remains admissible under LCA whether K is a new category or a segment of the two-segment category [K, L]; ACC holds the same way in either case. The case of (31b) is different, however. Its inadmissibility follows from the fact that L asymmetrically c-commands J. But if c-command is restricted to categories as Kayne proposes, excluding segments of categories along with terminals, then L will no longer c-command J if [K, L] is a two-segment category. Hence, (31b) is inadmissible only if K is a new category; if Q is adjoined to L, (31b) is admissible. The segment-category (adjunct, specifier; A- vs. A'-position) distinction can be maintained in the case of (31a), where *j* projects only to a head, but not in the case of (31b), where *j* projects to an X^{max}; in the latter case we have only adjunction, and the distinctions disappear.

This seems an odd result, for several reasons. On conventional assumptions the admissible structure (31a) should be inadmissible however interpreted; we do not expect nonmaximal heads to be specifiers or to be adjoined to maximal projections. Furthermore, it is strange that the only admissible case of the segment-category distinction (and the related ones) should be for a dubious structure such as (31a). Finally, there are real questions as to whether it is possible to eliminate what seem to be fairly significant distinctions between specifiers and adjuncts, A- and A'-positions.

Turning to the bare theory, the counterpart to both (31a) and (31b) is (32):



Here L is either *m* or *p*, K is either *j* or L, and K may be either a segment of [K, L] (j an adjunct in an A'-position) or a separate category (j a specifier in an A-position). The heads are the terminal elements *j*, *m*, *p* themselves; there are no head projections. There are no bar levels, only lexical elements and sets constructed from them. Assuming that L is not formed by adjunction, whichever of *m*, *p* is not projected to L is an X^{max} as well as a head. (32) cannot be interpreted as intended for (31a), with [j] a head adjoined to the maximal projection [L M N]. Rather, if L projects to K, then *j* is a single-terminal maximal projection, either specifier or adjunct of the head-complement construction L. And if *j* projects to K, it is either a head with complement L, or L is adjoined to *j* (which is a bare head or both maximal and minimal, depending on higher structure).

The disparity between the systems mounts when we consider richer structures. Thus, Kayne (still adopting standard X' conventions) compares (31) with an alternative in which the head [_M *m*] is replaced by the X^{max} [_{MP} [_M *m*]], inadmissible under LCA. But in the bare theory it again reduces to (32).

Despite the disparity let us ask how a modified LCA might be added to the bare theory. There is no category-terminal distinction, so either may c-command. Turning to (32), suppose that K is a separate category and L projects, so that *j* is a specifier in an A-position. ACC holds of (*j*, *m*), (*j*, *p*), so that *j* must precede *m* and *p*. But it would hold of (*m*, *p*) only if the single-terminal *p* (the complement of the head *m*) were replaced by a complex category. Hence, we have the order Spec-head-complement, though only for nontrivial complement.

Suppose that instead of terminal *j* we had branching J, with constituents α , β . L is an X', neither maximal nor minimal, so it does not c-command. Therefore the ACC relations are unchanged.

Suppose that K is a separate category and *j* projects. ACC holds as before; *j* is now the head of K with complement L.

Suppose that K is a segment, either *j* or L. There is no particular problem, but adjunct-target order (to which we return) will depend on the precise definition of c-command.

In brief LCA can be adopted in the bare theory but with somewhat different consequences. The segment-category distinction (and the related ones) can be maintained throughout. The intended interpretation of (31) is unformulable, correctly it seems. We draw Kayne's basic conclusion about SVO order directly, though only if the complement is more complex than a single terminal.

The conceptual arguments for LCA do not carry over to the bare theory. Thus Kayne shows how it is possible to derive such X'-theoretic stipulations as (33):

- (33)
- a. The complement of a head cannot be a head.
 - b. A category cannot have two heads.
 - c. Every phrase must be headed.

He notes further that (33c) entails that coordination cannot be of the form [XP YP] but must have a coordinate element as head ([α XP [and YP]], etc.), again correctly. But, as noted, derivation of these and other consequences crucially requires not only LCA but also (30), assumptions abandoned in the bare theory, and in the latter the conclusions of (33) follow trivially without LCA, though with a slightly different formulation for (33a) and (33b) ((33c) is unchanged): (a) the complement of a head can be a head, which will be an X^{max}; (b) a category can have two heads, one a bare head that projects, the other an X^{max}.

Let us return now to (32), taking L = *m* with the single terminal complement *p*, both minimal and maximal. Since neither *m* nor *p* asymmetrically c-commands the other, no ordering is assigned to *m*, *p*; the assigned ordering is not total, and the structure fails LCA. That leaves two possibilities. Either we weaken LCA so that non total orderings (but not 'contradictory' orderings) are admissible under certain conditions, or we conclude that the derivation crashes unless the structure $a = [L m p]$ has changed by the time LCA applies so that its internal structure is irrelevant; perhaps *a* is converted by Morphology to a 'phonological word' not subject internally to LCA, assuming that LCA is an operation that applies after Morphology.

Consider the first possibility: is there a natural way to weaken LCA? One obvious choice comes to mind: there is no reason for LCA to order an element that will disappear at PF, for example, a trace. Suppose then that we exempt traces from LCA, so that (32) is legitimate if *p* has overtly raised, leaving a trace that can be ignored by LCA. The second possibility can be realized in essentially the same manner, by assuming that LCA may (but need not) delete traces. Under this interpretation LCA may eliminate the offending trace in (32), if *p* has raised.

In short, if the complement is a single terminal XP, then it must raise overtly. If XP = DP, then its head D is a clitic, either demonstrative or pronominal, which attaches at a higher point (determined either generally or by specific morphological properties, depending on how cliticization works). If XP = NP, then N must incorporate to V (and we must show that other options are blocked). Clitics then are bare Ds without complements, and noun incorporation must be restricted to 'nonreferential NPs' (as noted by Hagit Borer), assuming the referential, indexical character of a noun phrase to be a property of the D head of DP, NP being a kind of predicate. Within DP the N head of NP must raise to D (as argued for different reasons by Longobardi (1990)).

We therefore expect to find two kinds of pronominal (similarly, demonstrative) elements: weak ones that are morphologically marked as affixes and must cliticize, and strong ones with internal structure, which do not cliticize, for example, in French, the determiner D (*le*, *la*, etc.), and the complex element *lui-même*. In Irish the weak element is again D, and the strong one may even be discontinuous, as in an teach sin ('that book,' with determiner an-sin (Andrew Carnie, pc)). A phenomenon that may be related is noted by Esther Torrego. In Spanish, in (34a) the Case-marker *de* can be omitted, but not in (34b):

- (34)
- a. *cerca de la plaza* ('near the plaza')
 - b. *cerca de ella* ('near it')

Deleting *de* in (34a), D = *la* can incorporate into *cerca* satisfying the Case Filter, but that is impossible in (34b) if the strong pronominal *ella* is not D but a more complex word from which the residue of D cannot be extracted.

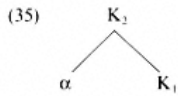
Since the affixal property is lexical, weak pronominals cliticize even if they are not in final position; for example, a pronominal object that is a specifier in a

Larsonian shell. If focus adds more complex structure, then focused (stressed) weak pronominals could behave like complex pronominals. If English-type pronouns are weak, they too must cliticize, though locally, not raising to Infl as in Romance (perhaps as a reflex of the lack of overt verb raising). The barrier to such structures as "I picked up it" might follow. English determiners such as "this," "that" are presumably strong with the initial consonant representing D (as in "the," "there," etc.) and the residue a kind of adjective perhaps. Various consequences are worth exploring.

While apparently not unreasonable, the conclusions are very strong; thus, every right-branching structure must end in a trace on these assumptions.

What about ordering of adjuncts and targets? In Kayne's theory adjuncts necessarily precede their targets. Within the bare theory ordering depends on exactly how the core relations of phrase structure theory, dominate and c-command, are generalized to two-segment categories.

Consider the simplest case, with α attached to K , which projects:



Suppose that K_2 is a new category, a the specifier. Take dominate to be an irreflexive relation with the usual interpretation. Then (35) ($= \{k, \{\alpha, K\}\}$, k the head of K) dominates α and K ; informally, K_2 dominates α and K_1 .

Suppose, however, that the operation was adjunction forming the two-segment category $[K_2 K_1] = \{<k, k>, \{\alpha K\}\}$. Are α and K_1 dominated by the category $[K_2 K_1]$? As for c-command, let us assume that α c-commands outside of this category; thus, if it heads a chain, it c-commands its trace, which need not be in K_1 (as in head raising). But what about further c-command relations, including those within (35) itself?

The core intuition underlying c-command is that

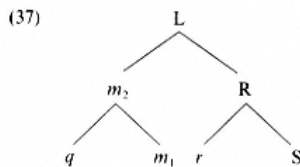
- (36) X c-commands Y if (i) every Z that dominates X dominates Y and (ii) X and Y are disconnected.

For categories we take X and Y to be disconnected if $X \neq Y$ and neither dominates the other. The notions "dominate" and "disconnected" (hence "c-command") could be generalized in various ways for segments.

Let us restrict these relations to terms, in the sense defined earlier: in the case of (35), to α , K ($= K_1$) and the two-segment category $[K_2, K_1]$. K_2 has no independent status. These decisions comport reasonably well with the general condition that elements enter into the computational system C_{HL} if they are 'visible' at the interface. Thus K_1 may assign or receive a semantic role, as may α (perhaps heading a chain), but there is no 'third' role left over for K_2 ; the two-segment category will be interpreted as a word by Morphology and WI (see (21)) if K is an X^{min} , and otherwise it falls under the narrow options discussed earlier.

If that much is correct, we conclude that in (35) $[K_2, K_1]$ dominates its lower segment K_1 , so that the latter does not c-command anything (including a not dominated by $[K_2, K_1]$ but only contained in it).

Turning next to c-command, how should we extend the notion "disconnected" of (36ii) to adjuncts? Take adjunction to a nonmaximal head (Kayne's (16) reduced to its bare counterpart):



Here q is adjoined to the head m to form the two-segment category $[m_2, m_1]$, a nonmaximal X^0 projecting to and heading the category L , which has label m . R is

the complement of m and r its head, and S (which may be complex) is the complement of r . What are the c-command relations for the adjunct structure?

The lowest Z that dominates q and $[m_2, m_1]$ is L ; therefore, q and $[m_2, m_1]$ asymmetrically c-command r and S , however we interpret "disconnected." What are the c-command relations within $[m_2, m_1]$? As noted, m_1 does not c-command anything. The other relations depend on the interpretation of "disconnected" in (36b). Kayne interprets it as "X excludes Y." Then q (asymmetrically) c-commands $[m_2, m_1]$ so that q precedes m_1 ; and in general an adjunct precedes the head to which it is adjoined. If X, Y are taken to be 'disconnected' if no segment of one contains the other, then q c-commands m_1 but not $[m_2, m_1]$, and again q precedes m_1 . If 'disconnected' requires still further dissociation of X, Y —say, that neither is a segment of a category that contains the other—then no ordering is determined for q, m_1 by LCA.

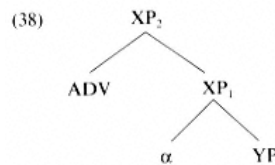
If m_1 is not a head but the complex category $[m, m P]$, so that q is an X^{max} for reasons already discussed, then q c-commands the constituents of m_1 under all interpretations of 'disconnect,' and the adjunct precedes the target (whether q is internally complex or not).

Left open then is the case of adjunction of a head to another head, that is, ordering within words. Whether order should be fixed here depends on questions about inflectional morphology and word formation.

Summarizing, it seems that Kayne's basic intuition can be accommodated in a straightforward way in the bare theory, including the major empirical conclusions, specifically, the universal order SVO and adjunct target (at least for XP-adjuncts). In the bare theory LCA gains no support from conceptual arguments and therefore rests on the empirical consequences. We take LCA to be a principle of the phonological component that applies to the output of Morphology, optionally ignoring or deleting traces. The specifier-adjunct (A-A') distinction can be maintained, and there may be multiple specifiers or adjuncts, though the options for adjunction are very limited for other reasons. There are further consequences with regard to cliticization and other matters, whether correct or not, I do not know.

7. Some Residual Problems

In discussing the options for adjunction we had put aside such structures as (38), with an adverbial adjoined to the two-segment category $[XP, XP]$, projected from α :



The construction is barred if XP has a semantic role at LF; say, if XP is a predicate (AP or VP), as in (39):

- (39) John $[_{VP}$ often $[_{VP}$ reads $\left\{ \begin{array}{l} \text{books} \\ \text{to his children} \end{array} \right\}]]$

Such structures as (38) could have been derived either by Merge or Move. The latter possibility can perhaps be ruled out in principle, under Greed: adverbs seem to have no morphological properties that require movement. The empirical evidence also seems to indicate that they do not form chains. Thus, an adverb in pre-IP position cannot be interpreted as if it had raised from some lower position.

The only option then is Merge. The question is whether we have 'base adjunction' in the EST sense, at least above the level of word formation. So far, it is barred if XP is semantically active as in (39). The sentences themselves are fine, but the structures assigned to them by (39) are not.

Adverbials can, however, be adjoined to such phrases as AgrP or IP or to any X'. Adjunction to X' by merger does not conflict with the conclusion that X' is invisible to C_{HL}; at the point of adjunction the target is an XP, not X'.

Such constructions as (39) have played a considerable role in linguistic theory since Emonds's (1978) study of differences between verb-raising and nonraising languages (French and English). The basic phenomena, alongside (39), are illustrated by (40) (both well-formed in French):

- (40) a. John reads often to his children.
b. *John reads often books.

A proposal sometimes entertained is that V raises from the underlying structure (39) to form (40a), but such raising is barred in (40b) for Case reasons; accusative Case is assigned to books by read under an adjacency condition of the kind proposed by Stowell (1981). French differs in the adjacency property or in some other way.

Apart from the fact that the source construction (39) is barred for the reasons discussed, the general approach is problematic on minimalist assumptions. This framework has no natural place for the assumed condition of adjacency. Furthermore, it takes Case to be assigned by raising to [Spec,AgrS] so that adjacency should be irrelevant in any event. It is also unclear why the verb should raise at all in (40), or where it is raising to. It seems that either the standard analysis is wrong or there is a problem for the minimalist framework.

In fact the empirical grounds for the analysis are dubious. Consider such adverbial phrases as every day or last night, which cannot appear in the position of often in (39):

- (41) *John every day reads to his children.

Nevertheless, we still find the paradigm of (40):

- (42) a. John reads every day to his children.
b. *John reads every day books.

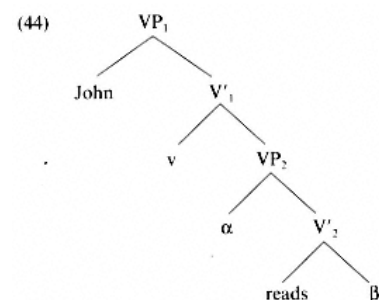
It seems then that the paradigm does not involve verb raising.

Furthermore, similar phenomena appear when raising is not an option at all, as in (43):

- (43) a. John made a decision (last night, suddenly) to leave town.
b. John felt an obligation (last night, suddenly) to leave town.

Here the adverbial may have matrix scope, so that it is not within the infinitival clause. It can appear between the N head and its complement, though the N cannot have raised in the manner under discussion. In general, therefore, it is doubtful that raising has anything to do with the relevant paradigms.

The phenomena suggest a Larsonian solution. Suppose that we exclude (39) from the paradigm entirely, assuming that often appears in some higher position and thus does not exemplify (38) with XP = VP. The structure underlying (40) and (41) is (44):



Here VP₁, V'₁ are projections of the 'light verb' v; and VP₂, V'₂ are projections of *read*. Whether the latter raises or not depends on whether v is selected in the initial numeration. Unless read raises, the derivation crashes because its features

(specifically, its tense and agreement features) are not satisfied; by HMC it must raise to v. The phenomenon is structurally similar to ECM, with a phrase moving so that further movement can satisfy Greed; see below.

Suppose that α in (44) is the adverbial *often*. Then if β = *to the children*, there is no problem. But if β = *books*, the derivation will crash; books cannot raise to [Spec,Agr] to have its Case checked because of the intervening α. The relativized minimality violation cannot be overcome by V-raising, which will not create a large enough minimal domain. Note the crucial assumption that the subject John is in [Spec, VP], the strong version of the VP-internal subject hypothesis that we have been assuming throughout; otherwise that position would be an 'escape hatch' for raising of books.

Under this analysis the basic facts follow with no special assumptions. There is a Case solution, but it does not involve adjacency. The problem of optional raising is eliminated, along with those raised by (42) and (43).

Questions remain about other matters, among them: What is the basis for the French-English distinction? Why do the wh-variants of the adverbials in question behave like adjuncts, not arguments? What about CED effects in the case of adjuncts such as Adj in (45), which is in a complement position if 'base-generated' adjuncts are barred?

- (45) They [read the book [_{Adj} after we left]]

Another question has to do with the scope of adverbials in ECM constructions. Consider the sentences (46):

- (46) a. I tell (urge, implore) my students every year (that they should get their papers in on time, to work hard).
b. I would prefer for my students every year to (get their papers in on time, work hard).
c. I believe my students every year to (work hard, have gotten their papers in on time).

Under the Larsonian analysis just outlined *every year* should have matrix scope in (46a), and (46c) should have the marginal status of (46b) with embedded scope if interpretable at all. The differences seem to be in the expected direction, though they are perhaps not as sharp as they might be. We would incidentally expect the distinction to be obviated in a verb-raising language such as Icelandic, as appears to be the case (Diane Jonas, pc).

Questions also arise about the relevance of the specifier-adjunct distinction to Case and agreement. In Chomsky (1993) I took the checking domain to include adjuncts as well as specifiers in the light of Kayne's (1989) theory of participial agreement. The assumption was that in passive and unaccusative the object passes through the [Spec,AgrO] position (A-movement), checking agreement with the participle, and then raises to subject, driven by Case; and in operator movement, the object adjoins to the Agr phrase (A'-movement), again checking agreement in the checking domain of Agr, then raising ultimately to [Spec,CP], driven by the operator feature. In particular Kayne found dialect differences associated with the two kinds of participial agreement.

Dominique Sportiche and Philip Branigan have observed that the operator-movement case is problematic because of such long-distance movement constructions as (47):

- (47) la lettre [qu'il a [_{AgrP} t' [_{AgrP} dit [que Pierre lui a [envoyé t]]]]]

Raising of the operator from t to t' (perhaps with intermediate steps) and then to [Spec,CP] is legitimate successive-cyclic A'-movement and should yield participial agreement with dit in the higher clause, incorrectly. That suggests that agreement (hence, presumably, Case as well) should be restricted to the specifier position, so that (47) would be ruled out as a case of improper movement if MLC requires lower adjunction on the way to t'. Assuming that the conclusion generalizes properly, we have another reason for the segment-category (specifier-adjunct, A-A') distinction, as well as for a requirement of successive cyclic movement. The dialect differences noted by Kayne remain unexplained, however.

Under either analysis the example illustrates that agreement can be assigned without Case. The same is true of such simple adjectival constructions as (48):

(48) John is [_{AgrP} t' Agr [_{AP} t intelligent]]

Here John raises from the predicate-internal subject position *t* to [Spec, Agr] (*t'*) for agreement with the adjective (raised to Agr), then raises on to the subject position ([Spec, AgrS]) for Case checking, with agreement checked independently by AgrS so that there is double agreement.

The counterpart would be a structure in which Case is assigned without agreement, that is, a structure of the form (49) in which α checks the Case of DP, which then raises to [Spec, Agr] for agreement:

(49) [DP Agr [... [t α ...]]]

This possibility is illustrated by transitive expletives (TEs) as analyzed by Jonas and Bobaljik (1993) and Jonas (to appear). Icelandic has structures of the following type (English words):

(50) [_{AgrS-P} there [_{AgrS} painted [_{TP} a student [_{AgrO-P} the house VP]]]]

The meaning is something like "a student painted the house," or the intelligible but unacceptable English counterpart (51):

(51) There painted the house a student (who traveled all the way from India to do it).

In (50) the expletive is in [Spec, AgrS] (subject) position; painted is the verbal head of VP adjoined to intermediate inflectional nodes and finally to AgrS; a student is raised to [Spec, Tense] with its Case checked by the trace of the raised Tense that heads TP; the house is raised (object raising) to [Spec, AgrO], where its Case and agreement are checked, and the VP contains only traces. Positions are motivated by placement of adverbs and negation in the overt forms. In the covert component, a student adjoins to the expletive for checking of its subject-agreement features, as in expletive constructions generally (we are assuming). The usual definiteness effect holds. Possibly, this reflects the fact that Agr, an inherently 'weak' element, requires a 'strong' element as its specifier –hence a definite (or specific) DP, either a full DP or the expletive; others remain in [Spec, TP] by Procrastinate. A similar argument might bear on the tendency for object raising to prefer definites; Agr, being weak, attracts the definite to its Spec, leaving indefinites behind by Procrastinate.

We thus have the full range of expected cases: agreement and Case are fully dissociated,⁴⁷ and there is good reason to suppose that Agr (i.e., a set of phi-features) appears twice and Tense once in a proliferated Infl system of the type proposed by Pollock (1989), a conclusion strengthened by examples such as (51) that couple object raising with expletives and thus require three pre-VP positions for arguments.

These TE constructions at once raise two questions:

- (52) a. Why do languages differ with regard to TEs, some (Icelandic) allowing them, others (English) not?
b. Are such structures permitted by economy principles?

Question (52a) presupposes some analysis of simple expletive constructions such as (53):

- (53) a. There arrived a man.
b. There is a book missing from the shelf.

Since subject-verb agreement holds between the verb and the postverbal DP (a man, the book), the latter must have raised covertly to [Spec, AgrS] on our assumptions, much as it does overtly in "a man arrived." The fact that the overt raising option exists alongside (53a) raises no problem; they arise from different numerations, so that if reference sets for economy comparisons are determined by the initial numeration, as proposed earlier, the options are not comparable in terms of economy of derivation. Covert adjunction to the expletive must be driven by some unsatisfied feature, either Case or agreement or both. Suppose that partitive Case can be assigned by unaccusatives along lines discussed by

Belletti (1988). Then covert raising in (53) would be motivated by agreement, as we have assumed for TEs.

Assuming that much, why does English lack TEs? Example (51) suggests that the lack of TEs may be a rather superficial phenomenon. As noted, the sentence is unacceptable (as is (53a), to some degree), though intelligible, and with other lexical choices the construction ranges in acceptability, as Kayne has observed, improving as the subject becomes 'heavier':

- (54) a. There entered the room a man from England.
b. There hit the stands a new journal.
c. There visited us last night a large group of people who traveled all the way from India.

Such constructions have been thought to result from an extraposition operation, but that is unformulable in our terms, which allow only one possibility: that they are TEs with the subject in [Spec, T] at LF but on the right overtly. The overt position could be the result of a process in the phonological component, perhaps motivated by properties of theme-rheme structures, which, as often noted, involve 'surface' forms in some manner. Prominence of the theme might require that it be at an 'extreme' position: to the right, since the leftmost position is occupied by the expletive subject. Icelandic might escape this condition as a reflex of its internal V-second property, which requires a method for interpreting internal themes. The lexical restrictions in English presumably reflect the locative character of the expletive. If speculations along these lines prove tenable, question (a) of (52) may not arise: the TE option may be general.

Question (b) of (52) is harder, leading into a thicket of complex and only partly explored issues. Note first that no problem is raised by the fact that TEs alternate with nonexpletive constructions; they arise from different numerations. But it is not obvious that constructions of the form (55), with subject SU in [Spec, T] at Spell-Out, should be allowed at all:

(55) Expletive Agr [SU [T XP]]

If TEs of the form (55) are legitimate, we should expect to find such structures as (56) in Icelandic, with seem raising to AgrS and SU raising from *t* to [Spec, TP] (as demonstrated by placement of adverbials and negation).

(56) There seems [_{TP} [_{SU} a man] [_{IP} t to be in the room]]

Assuming these to be legitimate, we then have to explain why (57) is barred in English:

(57) *There seems (to me, often) [_{IP} a man to be in the room]

Note that we do not have a direct contradiction. Thus, for UG reasons every language might bar (57) while permitting (56), with the associate having raised to [Spec, TP] of the matrix at Spell-Out. Both English and Icelandic would then have only (56), the surface form appearing in English as (58), given the speculative answer to (52a):

(58) There seems to be in the room [a man (who traveled all the way from India)]

Assuming this to be the correct resolution of the problem, we then ask why the structure illustrated in (56) is permitted but that of (57) barred. We cannot appeal to numeration in this case, because it is the same in the two examples.

We also have to explain why the form (59) is permitted, with there raising from the position of *t*, where it satisfies the Extended Projection Principle EPP in the lower (infinitival) clause:

(59) There seems [t to be [a man in the room]]

The problem becomes harder still when we add ECM constructions. In these the embedded subject does raise overtly to a position analogous to *t* in (59), where it is barred (see (57)); and it cannot remain in-situ as it does in (59):

- (60) a. I believe [John to be [t in the room]] (... to have been killed t)

- b. *I believe [a to be [John in the room]] (... to have been killed John)

Within the minimalist framework we expect the answers to these problems to come from invariant UG principles of economy.

The questions have to do with overt movement; hence, the relevant principle should be Procrastinate, which favors covert movement. Recall that Procrastinate selects among convergent derivations. Overt movement is permitted (and forced) to guarantee convergence.

To begin with, let's compare the effects in the contrasting cases (57) and (60). In each case the reference set determined by the initial numeration includes a second derivation; in the case of (57), the one that yields (59); in the case of (60a), the analogous one that yields (60b) with a the trace of raised I. Our goal is to show that in the case of (57) and (59) economy considerations compel raising of *there* from the embedded clause, while in the case of (60), on the contrary, the same considerations block raising of I from the embedded clause, requiring raising of John to satisfy EPP (that is, the strong DP-feature on the embedded Infl). The options available suggest that the difference lies in theta theory.

Consider first (57) and (59). Consider the structure that is common to the two derivations. In each at some stage we construct γ = (61), with the small clause β :

- (61) [γ to be [β a man in the room]]

The next step must fill the specifier position of γ to satisfy EPP. Given the initial numeration, there are two relevant possibilities: we can raise a man to [Spec, α] or we can insert *there* in this position. The former choice violates Procrastinate; the second does not. We therefore choose the second option, forming (62):

- (62) [γ there to be β]

At a later stage in the derivation we reach the structure (63):

- (63) [δ seems [γ there to be β]]

Convergence requires that [Spec, δ] be filled. Only one legitimate option exists: to raise *there*, forming (59). We therefore select this option, not violating Procrastinate, which does not arise.

Why then does the same argument not favor (60b) over (60a)? The common part of the derivations is (64):

- (64) [γ to be [β John in the room]]

Again, we have two ways to fill [Spec, γ], insertion of I being preferred if it yields a convergent derivation. Suppose we insert I, then raising it to form (60b). Recall that we cannot raise I to the VP-internal subject position [Spec,*believe*]; as already discussed, that violates Greed (see (9)). Therefore the LF output violates the Θ -criterion; the argument chain (I, *t*) lacks a Θ -role. Note that we must assume now that the derivation crashes; if not, it will converge as gibberish, blocking the desired derivation of (60a).

We therefore conclude that violation of the Θ -criterion prevents convergence, although the need for a Θ -role is not a formal property, like Case, that permits 'last-resort' movement. The conclusions are delicate. It remains for cases and consequences to be investigated further.

Much the same reasoning applies to such structures as (65):

- (65) β is believed [α to be [DP XP]]

Before Spell-Out both positions α and β must be occupied. Hence, DP must have raised successive-cyclically, yielding (66) (DP = *John*, XP = *in the room*):

- (66) John is believed [*t*' to be [*t* in the room]]

Suppose that the numeration included the expletive *there*. Then (65) would yield the possible outcome (67a) but not (67b) (DP = *a man*, XP = *in the room*; (67a) analogous to (59), (67b) to (57)):

- (67) a. There is believed [*t* to be [*a man* in the room]]
b. *There is believed [*a man* to be in the room]]

Note that (67b) contrasts with (60a), in which overt raising of DP is required.

Suppose that the numeration included expletive *it* instead of *there*. The analogue of (67a) is now impossible because *it*, unlike *there*, is not an LF-affix, so that the features of DP of (65) cannot be checked. Suppose, however, that instead of such DP we had a phrase that required no such feature checking, say, a CP instead of [DP XP] in (65), as in (68):

- (68) β is believed [α to have been proven [_{CP} that ...]]

Then we have the possible outcome (69a) but not (69b):

- (69) a. It is believed [*t* to have been proven [that ...]]
b. *It is believed [[that S] to have been proven *t*]

With a different numeration, lacking *it*, the embedded CP could have raised to matrix subject position, giving (70):

- (70) [that ...] is believed [*t*' to have been proven *t*]

Hence, it is not raising of the CP that is blocked in (68).

Suppose that we have the numeration N that yields (71) with successive-cyclic raising, *t* and *t*' being traces of *John*:

- (71) It seems that [John was believed [*t*' to be [*t* in the room]]]

An alternative derivation with the same numeration yields (72), *t* and *t*' traces of *John*:

- (72) *John seems that [it was believed [*t*' to be [*t* in the room]]]

The two derivations have the common part (73):

- (73) [α to be [β John in the room]]

The next step is to fill [Spec, α] either by raising of *John* (violating Procrastinate unless this is necessary for convergence) or by insertion of *it*. By the earlier reasoning insertion of *it* is required, yielding only the incorrect output (72) and blocking the correct alternative (71). It must be then that derivation of (72) does not converge, even though all formal properties are satisfied. (72) is an example of superraising, violating relativized minimality; in the current framework the Minimal Link Condition (MLC). The conclusion seems to be, then, that violation of this condition –and, presumably, of the conglomerate of properties that fall under ECP generally— causes the derivation to crash. The natural conclusion is that a chain that violates ECP is not a legitimate object at LF; in the framework of Chomsky and Lasnik (1993) a chain with a trace marked * is not a legitimate object. If so, then there is no convergent derivation of (72), and (71) is left as the only convergent derivation, as required.

One aspect of question (52b) still remains unanswered: why is the permitted TE structure (55), repeated here as (74a), not blocked by the alternative (74b) in accord with the reasoning just reviewed?

- (74) a. Expletive Agr [SU [T XP]]
b. Expletive Agr [*t* [T [... SU ...]]]

In other words, why is (50), repeated here as (75a), not blocked by (75b) with *there* inserted in the [Spec,TP] position instead of the student raising to this position in violation of Procrastinate?

- (75) a. [AgrS-P *there* [AgrS painted [TP a student [AgrO-P the house VP]]]]
b. [AgrS-P *there* [AgrS painted [TP *t* [AgrO-P the house [VP a student ...]]]]]

A possible argument is that the [Spec,TP] position must remain open for the subject *a student* after object raising. But that is not compelling. First, the trace of *there*, serving no function at LF, might well be deleted or replaced, so that [Spec,TP] would remain open. Second, the same question arises if there is no object raising.

Consider the structure (76) that is common to the two competing derivations:

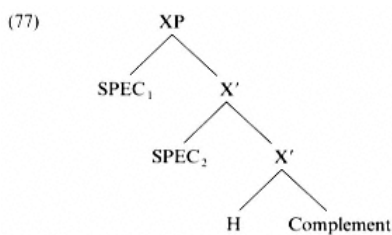
(76) [TP T AgrP]

The object may or may not have raised to Spec of the AgrP complement of T. The next step is to fill [Spec,TP]. Given the numeration, the choice, as before, is between raising of the subject *the student* or insertion of the expletive *there* by Merge. The former violates Procrastinate. Therefore, insertion of *there* is preferred –if it will lead to a convergent derivation. In the cases discussed earlier that was indeed the case: *there* was able to raise to Spec of the matrix clause, satisfying EPP. But in this case the matter is different because of intrinsic properties of *there*. Note that *there* bears Case but lacks intrinsic agreement, the latter determined not by *there* but by its associate, which raises to adjoin to it. Raising of the associate is driven by agreement, not Case, under the assumptions of the previous discussion (following Belletti 1988). Accordingly, if *there* is already in a Case position, it is prevented from raising by Greed; all its features are already satisfied. Hence, if *there* is in [Spec,TP], it cannot raise, and the derivation will crash. Therefore, given (76), the only convergent derivation results from raising of the subject *the students*, overriding Procrastinate, and yielding (75b). Insertion of *there* yielding the alternative (75b) is not an option.

The assumptions here are those of Chomsky (1993), repeated earlier. The principle of Greed (last resort) overrides convergence; Procrastinate selects among convergent derivations. In addition we have several conclusions about expletive constructions, theta theory, economy, and convergence. These are strong assumptions with varied empirical consequences. So far, they seem both conceptually plausible and supported by sometimes rather intricate argument.

The basic assumption about reference sets that underlies the preceding discussion is that they are determined by the initial numeration, but determined stepwise; at a particular point in the derivation we consider the continuations that are permitted, given the initial numeration; the most economical derivation blocks the others. On that assumption we explain the full set of cases although, it must be stressed, it is no trivial matter to generalize the reasoning to more complex cases, and the assumptions to which we have been led are far from solid.

Recall that the bare phrase structure theory allows multiple Specs in principle. Is this option realized? If so, we will have the structure (77):



Here we may assume Spec₁ and Spec₂ to be equidistant targets for movement, being within the same minimal domain.

Such ideas and phenomena related to them have been investigated in recent work (Miyagawa 1993a,b; Koizumi 1993; Ura 1993). If a language permits (77), then it should allow multiple assignment of Case and agreement from the same head, since Spec is available for checking these features. Suppose that the head H has a strong DP-feature. Then Case must be checked pre-Spell-Out at least at Spec₂, satisfying the strong feature of H. But Spec₁ is still an available position and could be filled by covert movement, so that the same Case is checked again by H. Furthermore, Spec₁ allows an escape hatch for superraising. Finally, it permits scrambling with A-position properties (binding, obviating weak crossover effects, etc.), unlike scrambling to an A'-position, which under earlier assumptions involves full reconstruction.

Ura (1993) has found that these properties correlate in a wide range of languages, lending strong empirical support to the conclusion that (77) is indeed an option that a language may have. A language might allow this option for some heads but not others (perhaps C but not Agr), in which case other properties follow, all matters now under investigation.

A major question left unresolved is the exact nature of the Minimal Link Condition (MLC). Specifically, we want MLC to deal with relativized minimality cases without any reference, direct or oblique, to the three categories: (i) HMC, (ii) A-movement, and (iii) A'-movement. Let's review the kinds of problems that arise.

MLC imposes a shortest-move requirement. Exactly how should we interpret that? We might understand MLC to mean that Move- α must attach α to the nearest target in the already formed structure in a way that does not cause the derivation to crash. One natural strategy for implementing this idea would be to rely on the fact that multiple Spec positions are allowed in principle. Whether the 'nearest target' is a head H or an X^{max} HP, the operation of substitution can always create a new equidistant Spec, moving α to occupy it. MLC is therefore satisfied, and no new reason is added for the derivation to crash. If this operation turns out to satisfy Greed in the manner already discussed (and, perhaps, other conditions, if any are relevant), the derivation allows what appears to be a relativized minimality violation, as just discussed. If not, the derivation is illegitimate; whether it crashes depends on how we understand the status of illegitimate multiple Spec positions. Hence, substitution satisfies MLC. We can therefore restrict attention to adjunction; the task is to show that 'improper movement' crashes, bearing in mind that a wide range of cases are admitted in principle when we keep to bare minimalist assumptions.

What is the 'nearest target'? Recall that self-attachment is ruled out for α , a head. The nearest target for α is therefore the head H that immediately c-commands α , or its (equidistant) specifier (or HP itself, if that too is equidistant). These are the nearest targets for nonminimal α as well if adjunction is barred within a minimal domain, as might be the case.

Consider nonmaximal α (a pure head). Recall that substitution (i.e., raising to Spec or to a new head position) is ruled out; only adjunction is an option. Adjunction of α to X^{max} (either [Spec,H] or HP) causes the derivation to violate (14) (uniformity), therefore to crash. That leaves only the possibility that α adjoins to H, yielding HMC in this case. Adjunction of α to H does not cause the derivation to crash on the assumption that α can still raise for feature checking, either by excorporation or pied-piping of H.

For maximal α , MLC requires that α adjoin to H, to HP, or to [Spec,H]. It remains to be shown that all of these are cases of 'improper movement' that crash if α ends up in an A-position.

These remarks are intended only to sketch out the issues that remain when the operation Move is freed from any specific conditions. Whether the facts of the matter are consistent with that goal is at the moment unclear, though parts of the problem seem to be within reach.

To summarize, it seems that we can hope to eliminate the theory of phrase structure almost entirely, deriving its properties on highly principled grounds. Many consequences follow for the theories of movement and economy when these conclusions are combined with other minimalist assumptions. We are left with hard and challenging questions of a new order of depth and prospects for a theory of language with properties that are quite surprising.