DOCUMENT DELIVERY ARTICLE

Journal Title: Step by step : essays on minimalist syntax in honor of Howard Lasnik /

Trans. #: 284181

Article Author: Chomsky, Noam

Call #: E-BOOK

Article Title: Minimalist Inquiries: The Framework

Location:

Volume:

Item #:

Issue:

Month/Year: 2000

Pages: 89-155

Imprint:

CUSTOMER INFORMATION:

HEIDI B HARLEY
PO BOX 210028
TUCSON, AZ 85721-0028
hharley@email.arizona.edu

STATUS: Faculty
DEPT: Linguistics

Chapter 3— Minimalist Inquiries: The Framework

Noam Chomsky

The remarks that follow are "inquiries," a term intended to stress their tentative character. They are "minimalist" in the sense of the "Minimalist Program," itself exploratory as the term indicates, and in its short career already developing in partially conflicting and attractive directions. What appears here is the first part of an unfinished manuscript. Here, I will keep to general considerations, rethinking the issues and concerns that motivate the program and attempting to give a clearer account and further development of them from one point of view, taking as a starting point the final sections of Chomsky 1995b (henceforth *MP*). [1] That collection reflects an evolution over several years, with fairly radical changes along the way. Much like earlier stages, this one reflects a collective effort, incorporating ideas and proposals of many students and colleagues with no serious attempt at attribution, in fact no way to achieve it. I should, however, like to take the occasion to express my very special indebtedness to Howard Lasnik for many years of close collaboration, which has been extremely rewarding for me and is most inadequately recorded in print, though well known to participants in these enterprises.

3.1— Background

Let us begin by reviewing briefly a series of assumptions, discussed and qualified elsewhere.

First, there is a faculty of language FL, a component of the human mind/brain dedicated to language. Given this endowment, a human infant, but not her pet kitten, will reflexively categorize parts of the confusion around her as "linguistic" and develop rich and highly articulated capacities to enter into these peculiar modes of human thought and action. In contrast, the infant and kitten will, it seems, develop along a rather similar path in acquiring capacities to deal with many other aspects of the world. [2] FL can be regarded as a "language organ," in the informal sense in which the visual system, the immune system, and the circulatory system are commonly described as organs of the body: not objects that can be removed leaving the rest intact, but subsystems of a more complex structure that we hope to understand by investigating parts that have distinctive characteristics, and their interactions. Despite explicit denials and apparent controversy, this much seems to be generally assumed, at least tacitly.[3]

Like other organs, FL has an "initial state" S_0 that is an expression of the genes. To a good first approximation, it is uniform for the species, apparently also biologically isolated in essential respects and a very recent evolutionary development.[4] FL undergoes state changes under triggering and shaping influences of the environment. If Jones's FL is in state L, we say that Jones has (speaks, knows, . . .) the (I-)language L.[5] Two immediate tasks of a theory of language are to characterize the languages (states) attained and the shared initial state: the tasks of "descriptive adequacy" and "explanatory adequacy," respectively. We understand Universal Grammar (UG) to be the theory of the initial state, and particular grammars to be theories of attained states.

The language L includes a cognitive system that stores information: roughly, information about sound, meaning, and structural organization. Performance systems access this information and put it to use. Empirical questions arise at once: in particular, to what extent are the performance systems part of FL, that is, language-dedicated, specifically adapted for language?[6] On the "sound side," the answer is unclear and disputed; on the "meaning side," the questions are much harder and more obscure for obvious reasons, and judgments can only be highly tentative. A standard working assumption is that performance systems are external to FL. That is a simplifying assumption, not definitely known to be false, though it may well be, perhaps in important ways.[7] The issues have had little effect on empirical inquiry into questions of descriptive and explanatory adequacy, but come to the fore within the Minimalist Program.

I will adopt usual conventions for present purposes, recognizing that they are not innocent. I therefore take L to be a cognitive system alone.

I will assume further that L provides information to the performance systems in the form of "levels of representation," in the technical sense.[8] The performance systems access these "interface levels." Assume further that performance systems are of two kinds: sensorimotor systems and systems of thought (to give a name to something very poorly understood). Let us take them (tentatively) to be unitary and distinct, in the sense that all sensorimotor systems access one interface level, and all systems of thought access a distinct interface level. On these assumptions we understand L to be a device that generates expressions Exp = <Phon, Sem>, where Phon provides the "instructions" for sensorimotor systems and Sem the "instructions" for systems of thought—information about sound and meaning, respectively, where "sound" and "meaning" are understood in internalist terms, "externalizable" for language use by the performance systems. [9] Theories of PF and LF seek to spell out the nature of Phon and Sem. I will assume some version of standard theories to be adequate for present purposes, using the conventional term *features* for the properties of language that enter into PF, LF, and the computational system that generates them.

Again, the assumptions are not innocent. Thus, Epstein et al. (1998) pursue a strong derivational approach in which performance systems access the computation itself, dispensing with levels of representation. That articulatory and perceptual systems access the same information (PF) is also far from self-evident, corresponding assumptions on the meaning side even less so. And there are many other questions.

To say that phonetic features are "instructions" to sensorimotor systems at the interface is not to say that they have the form "Move the tongue in such-and-such a way" or "Perform such-and-such analysis of signals." Rather, it expresses the hypothesis that the features provide information in the form required for the sensorimotor systems to function in language-independent ways. Similar observations hold on the (far more obscure) meaning side. The framework imposes a distinction between (a) linguistic expressions Exp = <PF, LF> that are internal to the mind/brain, and (b) observable events, utterances, and actions—externalization of (mentally constructed) speech acts. No questions arise about the ontological status of the set of expressions {Exp} generated by L; its status is somewhat like that of potential visual images or plans for limb motions.[10]

Finally, I will assume that the principles-and-parameters (P&P) approach is in important respects on the right track. Within any version of it, the major problem is to discover the principles and parameters, and to show how a particular choice of parameter values and lexicon enters into fixing a language L[11]—and to proceed beyond, to the study of use, acquisition, pathology, cellular mechanisms, and a wide range of other questions having to do with the place of language in the biological and social worlds.

Whatever its ultimate fate, the crystallization of the P&P approach has contributed to substantial progress in several of these areas. The approach also opens some new questions. Concern for descriptive and explanatory adequacy is as old as the study of language. As soon as the two traditional goals were reformulated within modern generative grammar, serious tension arose between them: the search for descriptive adequacy seems to lead to ever greater complexity of rule systems, varying among grammatical constructions and across languages, whereas the search for explanatory adequacy leads to the conclusion that language structure is largely invariant. It is this tension that has driven the research inquiry of generative grammar from its inception. The P&P framework suggests a way to resolve the tension, thus offering some conception of the form that a genuine theory might take.

It therefore becomes possible to consider some new questions about FL. In particular, we may ask the question, how well is FL designed? Suppose that a super-engineer were given design specifications for language: "Here are the conditions that FL must satisfy; your task is to design a device that satisfies these conditions in some optimal manner (the solution might not be unique)." The question is, how close does language come to such optimal design?

If the question is real, and subject to inquiry, then the P&P approach might turn out to be an even more radical break from the tradition than it seemed to be. Not only does it abandon traditional conceptions of "rule of grammar" and "grammatical construction" that were carried over in some form into generative grammar, but it may also set the stage for asking novel questions that have no real counterpart in the earlier study of language. [12]

The Minimalist Program is the attempt to formulate and study such questions. One should bear in mind that it is a *program*, not a theory, even less so than the P&P approach. There are minimalist questions, but no minimalist answers, apart from those found in pursuing the program: perhaps that it makes no sense, or that it makes sense but is premature. The program presupposes the common goal of all inquiry into language—to discover the right theory—and asks further why language is that way. More narrowly, it seeks to discover to what extent minimal conditions of adequacy suffice to determine the nature of the right theory.[13]

Questions of this kind are not often studied and might not be appropriate at the current level of understanding, which is, after all, still quite thin in a young and rapidly changing approach to the study of a central component of the human brain, perhaps the most complex object in the world, and not well understood beyond its most elementary properties.

The program is recent, and it is too early to assess it with any confidence. My own tentative judgment has two aspects, one methodological, the other substantive.

At the methodological level, the program has a certain heuristic and therapeutic value. It brings to light what might be fundamental problems, where empirical evidence and minimalist expectations conflict. And it encourages us to distinguish genuine explanations from "engineering solutions"—a term that I do not mean in any disparaging sense. Problems of descriptive and explanatory adequacy are vast and largely obscure. One tries to overcome them somehow, with special assumptions that are often not independently well motivated, hoping to reformulate the problems in ways that will facilitate further inquiry. Take, say, the study of conditions on extraction of subjects in terms of government and licensing, or attempts to account for the verb-second phenomenon or linear ordering in terms of X-bar theory, with its standard stipulations. Various solutions have been proposed that are useful and enlightening, but we can ask whether they are of roughly the order of complexity of the original problem. If so, it would be wrong to conclude that such proposals lack value; on the contrary, they may and often have opened the way to considerable progress. But we can still ask whether they are genuine solutions. Or consider the target of verb-raising. Evidence has accumulated that the verb can raise to a position higher than T but lower than C, differentiating languages by the position of a functional category α, on current assumptions. [14] But there is reason to doubt that such α can exist; or, to put it differently, if it does, then departures are needed from what appears to be the simplest and most principled form of phrase structure theory. Again questions arise as to whether there is some better way to conceive the matter. The Minimalist Program helps to focus attention on such issues, and perhaps to address them by showing that elimination of descriptive technology yields empirical results that are as good as, possibly even better than, before.

The substantive thesis is that language design may really be optimal in some respects, approaching a "perfect solution" to minimal design specifications. The conclusion would be surprising, hence interesting if true.

To introduce some terminology of *MP*, we say that a computation of an expression Exp *converges at an interface level IL* if Exp is legible at IL, consisting solely of elements that provide instructions to the external systems at IL and arranged so that these systems can make use of them; otherwise, it *crashes* at IL. The computation *converges* if it converges at all interfaces. Call the expression Exp so formed *convergent* as well. As in *MP*, we keep here to a restricted version of the concept of convergence, setting aside the matter of legible arrangement (which raises all sorts of complex issues), and tentatively assuming it to be irrelevant—no slight simplification. Certain features of lexical items are *interpretable*, that is, legible to the external systems at the interface; others are *uninterpretable*. We assume, then, that if an expression contains only features interpretable at IL, it converges at IL. [18]

The property *converges at IL* may hold of an expression formed in the course of a derivation that then proceeds on to IL. If, say, particles or adverbs have only LF-interpretable features, then they converge at LF when extracted from the lexicon and at every subsequent stage of derivation to LF. Similarly, an embedded clause may converge, for example, the bracketed subpart of *John thinks [it is raining]*. The phrase "converge at an interface" should not mislead: convergence is an internal property of an expression, detectable by inspection.[19]

Suppose that in state L, FL generates expressions Exp = <PF, LF>. Then L determines sound-meaning associations: the sounds and meanings determined by PF and LF, respectively, are associated in Exp. These are matters of fact that lie well beyond legibility conditions.[20] Take such standard examples as these:

(1) a. PF_1: John is impossible to forgive
 b. PF_2: John is impossible to be forgiven

Suppose L assigns to both a semantic representation LF on the model of *John is unlikely to forgive*. Then the generated expressions converge but with the wrong pairings. Suppose L assigns to both a representation LF' corresponding (closely enough) to that of *it is impossible to forgive John*. Then the association is right for Exp_1 = <PF_1, LF'> and wrong for Exp_2 = <PF_2, LF'>, though both converge. The assignment is wrong because it does not indicate the deviance of Exp_2, a crucial property distinguishing it from Exp_1. The conclusion holds even if the language user in state L assigns LF′ to PF_2 by some interpretive mechanism, using L but presumably going beyond.

Suppose all "best ways" to satisfy legibility conditions yield incorrect associations. Then departure from optimal design is required. If there are some that consistently yield the right sound-meaning relations, then we have reason to believe that language design is optimal in nontrivial respects.

Suppose that FL satisfying legibility conditions in an optimal way satisfies all other empirical conditions too: acquisition, processing, neurology, language change, and so on. Then the language organ is a perfect solution to minimal design specifications. That is, a system that satisfies a very narrow subset of empirical conditions in an optimal way—those it must satisfy to be usable at all—turns out to satisfy all empirical conditions. Whatever is learned about other matters will not change the conclusions about FL. That would be a strange and surprising result, therefore interesting to whatever extent it might be true. The Minimalist Program explores the possibility that language approaches "good design" in this sense. The strongest minimalist thesis would be this:

(2) Language is an optimal solution to legibility conditions.

Insofar as the thesis is true, information about other matters (sound-meaning connections, neurophysiology, etc.) may be helpful in practice—even indispensable—for discovering the nature of FL and its states. But it is irrelevant in principle. The tasks of biology of language remain as before, but become even more intriguing and difficult, because a new problem arises: how did the structure of the brain and the course of evolution happen to yield the outcome (2)? The internalist study of language—syntax in the broad sense—becomes much harder, hence more interesting and significant, because a standard of explanation is set that is very difficult to meet: descriptive machinery must satisfy stringent conditions, imposed by (2). Issues relating to the interface become of central concern. The problem of discovering whether, and if so how, considerations of economy enter into language design also gains new prominence, along with questions about their role in language acquisition. In general, all questions become harder, hence more interesting and significant—insofar as there is some truth to the strong thesis.

In these terms we might also be able to devise an interpretation for a thesis about language and psychology that seems to make little sense, but that has been so widely held that one might suspect that some significant intuition may lie behind it. The thesis is that linguists are to study "linguistic evidence" and "linguistic intuitions," but the results of their work, however revealing and far-reaching, do not bear on "reality," sometimes called "psychological reality." Other kinds of evidence are required to find out about reality. To take a classic example, Sapir provided rich "linguistic evidence" for phonological analyses he proposed, and went on to adduce much weaker "psychological evidence" to demonstrate their "psychological reality"; this was considered an audacious and controversial move, mainly on grounds that even the psychological evidence doesn't bear on reality.

In the recent period such ideas have appeared often in critical discussion of the program of generative grammar. Similarly, it is sometimes held that conclusions based on linguistic evidence must be confirmed by "converging" evidence from other sources, though conclusions based on these sources stand on their own. Another variant, thankfully put to rest after too many years, is that it is the task of psychologists to test (verify, refute) the theories of linguists, which are based on "linguistic evidence," not to contribute directly to these theories.

The linguistic evidence is generally understood to consist of informant judgments about sound and meaning and their relations. [21] The proposal is odd, as has often been pointed out: evidence does not come with a mark saying "I do or do not bear on reality." Judgments about (la,b) have no different status than other kinds of evidence with regard to the nature of FL, as part of the (physical) world. As in the study of vision and other domains, these judgments are discovered by experiment, typically informal in this case, though they can be carried out as carefully as is necessary to advance understanding. Uncontroversially, one seeks the widest possible range of relevant evidence, converging or conflicting. But there is no principled asymmetry between categories of evidence in this regard.

In terms of the preceding discussion, we might replace these proposals by a substantive (but extraordinarily strong) empirical hypothesis, namely, the thesis (2): an optimal solution to legibility conditions satisfies all other empirical tests as well. The reformulated thesis replaces the obscure notion of "linguistic evidence" by the meaningful notion: satisfaction of interface conditions. Relevant evidence is very limited. The thesis is even stranger than the requirement that "linguistic evidence" suffices to determine grammars, but it has empirical content. One might perhaps suggest it as a kind of rational reconstruction of dubious ideas about linguistic evidence and psychological reality that have appeared in one or another form.

Suppose we understood external systems well enough to have clear ideas about the legibility conditions they impose. Then the task at hand would be fairly straightforward at least to formulate: construct an optimal device to satisfy just these conditions, and see how well it satisfies other empirical conditions. If all such efforts fail, then add "imperfections" as required. But life is never that simple. The external systems are not well understood. Progress in understanding them goes hand in hand with progress in discovering the language systems that interact with them. So the task is simultaneously to set the conditions of the problem and to try to satisfy them, with the conditions changing as we learn more about how to do so. That is not surprising. It is much what we expect when trying to understand some complex system. We proceed with tentative proposals that seem reasonably firm, expecting the ground to shift as more is learned.

3.3— Architecture

We are taking L to be the recursive definition of a set of expressions Exp = <PF, LF>. We can now raise a question—at least, an apparent question—about the interpretation of the recursive definition.

One might construe L as a step-by-step procedure for constructing Exps, suggesting that this is how things work as a real property of the brain, not temporally but as part of its structural design. Assumptions of this nature constitute a *derivational approach* to L. The *strong* derivational approach dispenses with the expression altogether, assuming that information is provided to interface systems "dynamically" (see p. 91). A *weak* derivational approach assumes that interface levels exist, allowing "postcyclic" operations that apply to them in whole or in part (deleting the tail of a chain, imposing metrical structure, determining ellipsis and scope, etc.). There are many options.

With richer set-theoretic assumptions, a recursive definition can be restated as a direct definition, in this case, of the following form: E is an expression of L iff . . . E . . . , where . . .— . . . is some condition on E. One might, then, take L to be a direct definition of the set {Exp}, adopting a *representational approach*. Again there are weaker varieties, for example, the assumption that the set of LF representations is given (universally, or by L), PF being derived from LF by some computational procedure.

The issue is reminiscent of old questions about morphological processes ("item-and-process" vs. "item-and-arrangement," etc.) and grammatical transformations. Thus, does a transformation map an input structure to an output structure, or is it an operation on the "output" that expresses properties of the "input"? It is unclear whether these are real questions; on the surface they look like the question whether $25 = 5^2$ or $5 = \sqrt{25}$. If the questions are real, they are subtle. They have elicited no little passion over the years, but it is out of place. [22] The apparent alternatives seem to be mostly intertranslatable, and it is not easy to tease out empirical differences, if there are any.

Surprisingly, there is reason to believe that the questions may be real.[23] The evidence that has been adduced is far from conclusive and often conflicting. I will adopt the derivational approach as an expository device, though I suspect it may be more than that. If so, that would be a curious and puzzling fact about the nature of the mind/brain.[24]

Suppose that the issue is real, and the derivational approach in fact correct. Then further questions arise. For example, we might inquire into the complexity of the generative procedure. Such questions have arisen over the years, in one or another form. One category concerns "least effort" conditions, which seek to eliminate anything unnecessary: (a) superfluous elements in representations, (b) superfluous steps in derivations. The tacit assumption is that failure to meet these conditions imposes deviant interpretations, in principle an empirical issue though often not an easy one to resolve. Subcategory (a) involves legibility conditions and convergence ("full interpretation"); (b) holds that operations are allowed only if there is some reason for them. In the terms we are exploring, reasons are reduced to effects at the interface. Possibilities that have been investigated (if not in these terms) include constraints that bar PF-vacuous overt movement and others that seek to limit effects on PF (Procrastinate). An LF counterpart is that covert operations are allowed only if they have an effect on interpretation at LF. Another category seeks to reduce "search space" for computation: "Shortest Movement/Attract." successive-cyclic movement (Relativized Minimality, Subjacency), restriction of search to c-command or minimal domains, and so on. Yet another imposes "local determinability" conditions (barring "look-ahead," "backtracking," or comparison of alternatives). I will assume these ideas to be generally on the right track and pursue them further below.[25]

Some of these notions have analogues in formal complexity theory. Most are the kinds of intuitive ideas about "operative complexity" that enter commonly into the cognitive sciences[26] and design considerations generally. Suppose automobiles lacked fuel storage, so that each one had to carry along a petroleum-processing plant. That would add only bounded "complexity," but would be considered rather poor design. Something similar might well be true for language.

Let's consider a few such proposals, beginning with conventional ones and proceeding to others that are more controversial; it's worth bearing in mind, however, that the logic is similar throughout.

UG makes available a set ***F*** of features (linguistic properties) and operations C_{HL} (the computational procedure for human language) that access **F** to generate expressions. The language L maps **F** to a particular set of expressions ***Exp***. Operative complexity is reduced if L makes a one-time selection of a subset [F] of **F**, dispensing with further access to **F**. It is reduced further if L includes a one-time operation that assembles elements of [F] into a lexicon *Lex*, with no new assembly as computation proceeds. On these (fairly conventional) assumptions, acquiring a language involves at least selection of the features [F], construction of lexical items Lex, and refinement of C_{HL} in one of the possible ways—parameter setting. [27] One could offer a conceptual argument that conventional assumptions are mistaken, on the grounds that a theory lacking certain concepts (here [F], Lex, and the operations forming [F] and Lex) is better than an otherwise identical one that employs them. But if operative complexity matters, the argument loses force. Conceptual arguments can be given either way. The issues are empirical and can be settled only by investigating consequences of alternative conceptions, considered so obvious in this case that the question has not arisen.

We assume, then, that a language L maps ([F], Lex) to **Exp**. The next natural simplification would be to reduce access to the domain ([F], Lex) of L. Consider [F]. In the computation of LF—what we may call *narrow syntax*—it seems that [F] is not accessed, only Lex (and features of its items). The restriction does not extend to phonology, however: features are introduced in the course of computation, and in different ways for different languages, whatever approach one takes to computation of PF. Keeping to narrow syntax, then, we may take C_{HL} to be a mapping of Lex to the LF representations of **Exp**.

Is it also possible to reduce access to Lex, the second component of the domain of L? The obvious proposal is that derivations make a one-time selection of a *lexical array* LA from Lex,[28] then map LA to expressions, dispensing with further access to Lex. That simplifies computation far more than the preceding steps. If the derivation accesses the lexicon at every point, it must carry along this huge beast, rather like cars that constantly have to replenish their fuel supply. [29] Derivations that map LA to expressions require lexical access only once, thus reducing operative complexity in a way that might well matter for optimal design.

Again, conceptual arguments can be given either way, but they carry little weight. The questions are empirical. Investigating them, we can hope to discover whether (and if so how) what might reasonably be considered complexity/economy considerations enter into language design.

If FL operates with the economy principles just reviewed, then a language L follows procedures (3a) and (3b) to specify the language (apart from parameter setting), then applies (3c) and (3d) to derive a particular Exp.

(3) a. Select [F] from the universal feature set **F**.
b. Select Lex, assembling features from [F].
c. Select LA from Lex.
d. Map LA to Exp, with no recourse to [F] for narrow syntax.

We will return to further steps along the same path, but let us first look more closely at general properties of Lex and narrow syntax (the recursive part of L).

First, what operations enter into this component of C_{HL}? One is indispensable in some form for any language-like system: the operation *Merge*, which takes two syntactic objects (α, β) and forms K(α, β) from them. A second is an operation we can call *Agree*, which establishes a relation (agreement, Case checking) between an LI α and a feature F in some restricted search space (its *domain*). Unlike Merge, this operation is language-specific, never built into special-purpose symbolic systems and apparently without significant analogue elsewhere. We are therefore led to speculate that it relates to the design conditions for human language. A third operation is *Move*, combining Merge and Agree. The operation Move establishes agreement between α and F and merges P(F) to αP, where P(F) is a phrase determined by F (perhaps but not necessarily its maximal projection) and αP is a projection headed by α. P(F) becomes the specifier (Spec) of α ([Spec, α]). Let us refer to Move of P to [Spec, ϕ] as A-*movement*, where ϕ is an agreement feature (ϕ-feature); other cases of Move are $\bar{A}$-movement.

Plainly Move is more complex than its subcomponents Merge and Agree, or even the combination of the two, since it involves the extra step of determining P(F) (generalized "pied-piping"). Good design conditions would lead us to expect that simpler operations are preferred to more complex ones, so that Merge or Agree (or their combination) preempts Move, which is a "last resort," chosen when nothing else is possible. Preference for Agree over Move yields much of the empirical basis for Procrastinate and has other consequences, as do the other preferences. [30]

Let us turn next to the lexicon Lex, adopting some fairly common assumptions along with more controversial ones, and keeping to simple cases.

LIs fall into two main categories, substantive and functional; we are concerned now mainly with the latter. Take the *core functional categories* CFCs to be C (expressing force/mood), T (tense/event structure), and v (the "light verb" head of transitive constructions). All CFCs may have ϕ-features (obligatory for T, v). These are uninterpretable, constituting the core of the systems of (structural) Case agreement and "dislocation" (Move). Neither T nor v assigns inherent Case; other light verbs may, as may substantive categories.[31]

Consider the selectional properties of CFCs, beginning with semantic (s-)selection.[32] Assume that C can be unselected (root), whereas v and T cannot. C is selected by substantive categories, v only by a functional category. T is selected by C or V. If selected by C, it has a full complement of ϕ-features; if by V, it is *defective* (T_{def}). C selects T; T and v select verbal elements. v may also select a nominal phrase NP/DP as its *external argument* EA = [Spec, v].[33]

Each CFC also allows an extra Spec beyond its s-selection: for C, a raised *wh*-phrase; for T, the surface subject; for v, the phrase raised by object shift (OS). For T, the property of allowing an extra Spec is the Extended Projection Principle (EPP). By analogy, we can call the corresponding properties of C and v *EPP-features*, determining positions not forced by the Projection Principle. I will restrict attention to XP positions, though a fuller picture might add X^0 as another case of the EPP (see note 92). EPP-features are uninterpretable (nonsemantic, hence the name), though the configuration they establish has effects for interpretation.

Basic structural properties of CFCs are illustrated in the configuration (4), where H is the CFC, XP is the extra Spec selected by its EPP-feature, and EA is the external argument selected by H = v.

(4) α = [XP [(EA) H YP]]

The following properties hold among these elements:

(5) a. If H is v/C, XP is not introduced by pure Merge.
b. In the configuration $[_\beta\ T_\beta \ldots \alpha]$, β minimal,

i. if H is C, T_β is independent of α;

ii. if H is v, T_β agrees with EA, which may raise to [Spec, T_β] though XP cannot;

iii. if H is T_{def}, XP raises to [Spec, T_β] if there is no closer candidate γ for raising. [34]

Pure Merge is Merge that is not part of Move. The relevant properties of T_β have to do with Case/agreement and the EPP. In (5bii), if EA does not raise, [Spec, T_β] is introduced by pure Merge to satisfy the EPP. The case of H = nondefective T is omitted in (5b): if (5bi) holds for C, it holds for (nondefective) T selected by C. In fuller generality, β in (5b) should be taken to be the minimal β containing α headed by any CFC H_β, which would therefore be either T or v. The relations of T_β to α extend partially to v as well—specifically, exceptional case marking in (5biii) as compared to raising to [Spec, T]. I put this aside temporarily, for ease of exposition and because of some differences to which I will return.

Property (5a) follows in part from the θ-theoretic principle (6). which is implicit in the conception of θ-roles as a relation between two syntactic objects, a configuration and an expression selected by its head.[35]

(6) Pure Merge in θ-position is required of (and restricted to) arguments.

We can therefore restrict attention in (5a) to expletives (*Expl*): for arguments it follows from (6). (6) also applies to (5b), restricting pure Merge in [Spec, T_β] to Expl.

The property (5bi) is illustrated in (7), (5bii) in (8), and (5biii) in (9).

(7) a. there are questions about [$_\alpha$ what C [$_{TP}$ John read *t*]]

b. there is a possibility [$_\alpha$ that proofs will be discovered]

(8) T_β [$_\alpha$[$_{DO}$ the book] [[$_{Subj}$ many students] [read t_{DO}]]]

(9) a. T_β-is likely [$_\alpha$ there to be a proof discovered]

b. T_β [$_{vP}$ I expected [$_\alpha$ there to be a proof discovered]]

In (7) α = CP with H = C, and the relevant properties of matrix T (T_β) are independent of α in accord with (5bi). α is a closed system with regard to Case/agreement properties, determined internally without outside effect.

In (8) α is an OS construction with H = v and DO = XP of (4)/(5bii). Under Holmberg's Generalization V raises to T_β, which agrees with EA = Subj. EA can then raise to [Spec, T_β] (*many students read-pl the book (never)*) or remain in situ with Merge of Expl to satisfy the EPP (*there read-pl the book (never) any students*), illustrating the two options for (5bii). The positions of DO and Subj can be determined by left-edge adverbs. [36]

In (9) α = TP with H = T_{def} and XP = *there*. XP raises to [Spec, T_β] in (9a), but not in (9b) with intervening $\gamma = I$, yielding (10a) and (10b), respectively, which illustrate (5biii).

(10) a. there is likely to be a proof discovered
b. I expected there to be a proof discovered

Properties of matrix T (= T_β) depend on α in (9a) but not in (9b), where it is the head of vP (not T_β) that is related to α. The relation of T_β to α in (9a) is analogous to that of v to α in (9b), but there are differences to which we will return.

The order object-subject in (8) follows from the assumption that Merge preempts the more complex operation Move. When the derivation has reached the stage (11), two operations have to take place for an OS construction, each creating Spec: Merge of EA and raising of DO (Move).

(11) [v[V DO]]

If Merge applies first, the order is as in (8).
The examples in (9)/(10) contrast with those in (12).

(12) a. *there is likely [$_\alpha$ a proof to be discovered]
b. *I expected [$_\alpha$ *t* to be a proof discovered]
c. I expected [$_\alpha$ a proof to be discovered]

Suppose the derivation has reached reached the stage (13), analogous to (11), with $T_\alpha = T_{def}$.

(13) [T_α [be a proof discovered]]

The EPP requires that something occupy [Spec, T_α]. Two options are available: merge *there* or move *a proof.* Preference of Merge over Move selects the former. Accordingly, (9a)/(10a) is permitted and (12a) is barred. But Merge of an argument in [Spec, T_α] violates the θ-theoretic condition (6). Therefore, (12b) is barred. Either an expletive is merged, yielding (10b), or Move applies, yielding (12c). The choice depends on whether or not an expletive is available in the initial lexical array—the first time step (3c) enters.

Four kinds of complexity considerations enter into this account:

(14) a. Simple operations preempt more complex ones.
b. Search space is limited (locality).

c. Access to the feature set **F** is restricted by (3).
d. Computation is locally determined (no look-ahead).

Conclusion (14d) follows from (14a–c) and (6), all plausible principles of some generality; in the background is the derivational approach (cyclicity). Alternative analyses have been proposed, but to my knowledge they are not locally determinable or they introduce special or dubious assumptions.

By the reasoning just reviewed, other constructions similar on the surface to (12c) could be legitimate—for example, (15), where V is a *seem*-type verb with an optional Spec [37] that is raised to the matrix Spec in (15a) and missing in (15b).

(15) a. me(dat) V [$_{TP}$ several people to be in the room]
b. there V [$_{TP}$ several people to be in the room]

(15a) is similar to (12c): Move applies in the embedded TP in preference to Merge of *me*, in accord with (6); *me* then merges as Spec of the *seem* construction and raises to matrix subject. Case (15b) could arise if the embedded clause is a multiple-subject construction, in which both Merge (of Expl) and Move (of *several people*) have applied, Expl then raising to the matrix Spec. Both possibilities appear to be legitimate in languages with the relevant properties.[38]

Control infinitivals, I have assumed, fall together with finite clauses, headed by C selecting nondefective T (with tense-modal structure and a full complement of ϕ-features). Like other CPs, they generally undergo movement and clefting and can appear as root expressions (typically with *wh*-phrase Spec or as discourse fragments), and structural Case is assigned to the subject of T. These properties are common to CPs and distinguish them from raising/ECM infinitivals headed by T_{def}, lacking C and tense structure and assigning no Case to subject, and lacking the distributional freedom of CP.[39] We also find convergent constructions like those in (16), analogous to (7) and contrasting with (12a).

(16) a. it's fun [$_{\alpha}$ PRO to [*t* go to the beach]]
b. it's about time [$_{\alpha}$ PRO to [leave *t*]]
c. it was decided [$_{\alpha}$ PRO to be executed *t* at dawn]

Raising is possible throughout in the closed system α (as in (7)). In (16c), at least, PRO is controlled by an implicit argument: it can mean that the prisoners decided that they would be executed at dawn, but not that we decided that they would be.

A problem throughout the whole account is why raising is *ever* possible, if Agree and Merge preempt Move. The question is answered in part by the θ-theoretic principle (6), which bars pure Merge of arguments in non-θ-positions and correspondingly restricts Move to such positions. Choice of Move over Agree follows from presence of EPP-features, where pure Merge is inapplicable. The remaining question, then, is why Merge of Expl does not always bar Move. That question is partly answered by the initial choice of lexical array: it may or may not make Expl available. But that cannot be the whole story, as illustrated in (7) and (16), where Expl is available in the lexical array but Move takes place in the embedded phrase α. [40]

A straightforward solution would be to take the derivational approach still more seriously and further extend the procedures in (3) that reduce access to the domain of L. Suppose we select LA as before, under (3c); the computation need no longer access the lexicon. Suppose further that at each stage of the derivation a subset LA_i is extracted, placed in active memory (the "workspace"), and submitted to the procedure L. When LA_i is exhausted, the computation may proceed if possible; or it may return to LA and extract LA_j, proceeding as before. The process continues until it terminates. Operative complexity in some natural sense is reduced, with each stage of the derivation accessing only part of LA. If the subarray in active memory does not contain Expl, then Move can take place in the corresponding stage; if it does, Merge of Expl preempts Move.[41]

The next step is to characterize the subarrays LA_i that can be selected for active memory. LA_i should determine a natural syntactic object SO, an object that is relatively independent in terms of interface properties. On the "meaning side," perhaps the simplest and most principled choice is to take SO to be the closest syntactic counterpart to a proposition: either a verb phrase in which all θ-roles are assigned or a full clause including tense and force. Call these objects *propositional*. Considerations on the "sound side" support the choice, given properties of the kind mentioned earlier distinguishing CP from TP, which extend to vP (fronting, extraposition, pseudoclefting, response fragments, etc.).

LA_i can then be selected straightforwardly: LA_i contains an occurrence of C or of v, determining clause or verb phrase—exactly one occurrence if it is restricted as narrowly as possible, in accordance with the guiding intuitions. Take a *phase* of a derivation to be an SO derived in this way by choice of LA_i. A phase is CP or vP, but not TP or a verbal phrase headed by H lacking ϕ-features and therefore not entering into Case/agreement checking: neither finite TP nor unaccusative/passive verbal phrase is a phase. Suppose phases satisfy a still stronger cyclicity condition:

(17) The head of a phase is "inert" after the phase is completed, triggering no further operations.

A phase head cannot trigger Merge or Attract in a later phase, and we can restrict attention to phases in which all selectional requirements are satisfied, including the EPP for T (by virtue of cyclicity) and for v/C, and selection of EA for v if required; otherwise, the derivation crashes at the phase level.

Derivations proceed phase by phase: (18), for example, has the four phases shown in brackets.

(18) [John [*t* thinks [Tom will [*t* win the prize]]]]

An alternative that has been suggested (class lectures, MIT, 1995, and various talks and papers) is to define phases in terms of convergence. The two options are then as follows:

(19) a. Phases are propositional.
b. Phases are convergent.

Under (19a) LA_i is determined by a single choice of C or v. Under (19b) local determination is not possible. Complexity considerations therefore favor option (19a), and again the empirical evidence supports the same conclusion.

The two options have similar (though not identical) consequences in such cases as (18), but are clearly distinguished elsewhere. One case is $\bar{A}$-movement, as in (20). [42]

(20) which article is there some hope [$_\alpha$ that John will read t_{wh}]

For reasons to which we will return, assume that the *wh*-phrase has an uninterpretable feature analogous to structural Case for nouns, which requires it to move to its final position in an appropriate C. Then α is a phase under option (19a) but not option (19b) (it does not converge, containing an uninterpretable feature). The only phase is (20) itself; merger of *there* blocks raising of *John* to [Spec, TP] within α, so (20) is underivable without look-ahead. That is unnecessary under option (19a), where α is derived from LA_i lacking Expl.

The descriptive typology of movement, a leading research topic for years,[43] offers other reasons to suspect that phases are real, understood under option (19a). There are several categories: movement can be feature-driven or not, and in the former case can be directly or indirectly feature-driven. Typical cases include raising to subject (directly feature-driven), the nonfinal stages of successive-cyclic movement (indirectly feature-driven), and QR and "stylistic movement" (perhaps not feature-driven). [44]

Indirect feature-driven movement (IFM) subdivides into types depending on the attracting head H in the final stage: (a) A-movement when H has ϕ-features (yielding the Case/agreement system), or (b) $\bar{A}$-movement when H has *P-features* of the peripheral system (force, topic, focus, etc.).[45] The intuitive argument for IFM has always been that locality conditions require "short movement" in successive stages, leading to convergence in the final stage. We can express a version of this idea as a "phase-impenetrability condition," strengthening further the notion of cyclic derivation. Given HP = [α [H β]], take β to be the *domain* of H and α (a hierarchy of one or more Specs) to be its *edge*. The thesis under consideration is (21).

(21) *Phase-Impenetrability Condition*
In phase α with head H, the domain of H is not accessible to operations outside α, only H and its edge are accessible to such operations.

The cycle is so strict that operations cannot "look into" a phase α below its head H. H itself must be visible for selection and head movement; hence, its Specs must be as well. The Phase-Impenetrability Condition yields a strong form of Subjacency.[46] For A-movement, it should follow from the theories of Case/agreement and locality.[47] The stipulation is for clausal $\bar{A}$-movement, the basic question from the earliest study of these topics. We will return to some speculation about reducibility to economy conditions.

The intended consequences do not follow if phases are determined by convergence, but the propositional alternative (option (19a)) accommodates them, verbal phases being limited to transitive v with ϕ-features and EA. The Phase-Impenetrability Condition requires that $\bar{A}$-movement, target the edge of every phase, CP and vP. There is evidence from reconstruction effects and parasitic gap constructions that this may be true.[48]

The idea that IFM applies only if needed to guarantee eventual convergence appears to raise questions of look-ahead. These are obviated if the Phase-Impenetrability Condition holds. Local determination is straightforward, and an uninterpretable feature in the domain determines at the phase level that the derivation will crash.[49]

Let us return to the basic configuration (4) for CFCs, repeated as (22), where XP is the extra Spec determined by the EPP-feature of the attracting head H.

(22) α = [XP [(EA) H YP]]

Typical examples of (22) are raising to subject (yielding (23a)), OS (yielding (23b), with XP = DO and *t* its trace), and overt $\bar{A}$-movement (yielding (23c), with H = C and XP a *wh*-phrase; see note 31).

(23) a. XP – [T YP]
b. XP – [Subj [v [V *t*]]]
c. XP – [C YP]

The EPP-feature of T might be universal. For the phase heads v/C, it varies parametrically among languages and if available is optional.

Cases (23b) and (23c) are alike, and unlike (23a), in other respects. One is that the EPP-feature can be satisfied by Merge of an Expl in (23a), but not in (23b) and (23c) (see (5a)). This fact might be unexpected for (23b), because (23b) and (23a) enter into the Case/agreement system in much the same way; and the relation of [Spec, H], H, and a related phrase β in the complement of H appears to be similar for T and v.

Phases are determined by a choice of C/v, not T, which suggests a basis for the similarities and asymmetry. The fact that the EPP-feature when available is optional for C/v suggests that it is a property of the phase Ph.

(24) The head H of phase Ph may be assigned an EPP-feature. [50]

Once Ph is completed, exhausting the lexical subarray from which it is derived, (24) may optionally apply, assigning an EPP-feature to H. From the strong cyclicity condition that renders H inert beyond the phase itself (see (17)), it follows that the EPP must be satisfied by raising within Ph: pure Merge from outside Ph is barred. In (23b,c) XP is raised from within the domain of H = v/C, completing the account for case (5a).[51] The remaining properties of (5) should follow from the theories of Case/agreement and locality, to which we will turn shortly.

The picture might be extended to incorporate QR, if alongside the ϕ-and P-features that drive movement, there are also QU-features, attracting quantifiers that pied-pipe an appropriate phrase. One might also explore a generalization of the idea that operations can apply only if they have an effect on outcome; see text at note 25. This would translate here into restricting (24) to the condition in which it permits IFM or specific interpretations associated with peripheral positions (e.g., specificity and informational conditions on OS).[52] There are many problems and ramifications, which I will leave to the side.

(24) yields A- or $\bar{A}$-movement depending on whether the phase head has ϕ- or P-features. It might have both. Suppose that in the construction (25) all four phase heads are assigned an extra Spec by (24), associated with P-features for C and v_2 but not for v_1.

(25) [Spec, C_2] . . . [Spec, v_2] . . . [Spec, C_1] . . . [Spec, v_1] . . . XP

XP raises through the Specs in succession, landing finally in [Spec, C_2]. The result is the four-membered $\bar{A}$-chain ([Spec, C_2], [Spec, v_2], [Spec, C_1], [Spec, v_1]) and the two-membered A-chain ([Spec, v_1], XP) (formed by OS). [Spec, v_2] is an $\bar{A}$-position, by virtue of the P-feature associated with the extra Spec introduced by (24); v_2 also had ϕ-features involved in object Case/agreement but these would have been deleted phase-internally before (24) assigns the extra Spec.[53]

There are more complex cases to consider, along with a variety of other issues.[54] But it seems that the cyclic approach to accessing lexical arrays is plausible on conceptual and empirical grounds, along with option (19a) and the condition (24) on extra Specs.

Nothing has been said yet about the fact that C_{HL} forms syntactic objects in parallel, according to the derivational approach adapted from *MP*. Consider, for example, the expression (26).

(26) the demonstration that glaciers are receding showed that global warming must be taken seriously

The prefinal phases of the derivation are the syntactic objects corresponding to (27a–c).[55]

(27) a. P_1 = [$_{CP}$ that global warming must be taken seriously]
b. P_2 = [$_{CP}$ that glaciers are receding]
c. P_3 = [$_{vP}$ [the demonstration P_2 [show P_1]]]

For each new phase, a subarray provides the lexical material required and the operations proceed in the manner already sketched, with P_1/P_2 unordered. Step (27c), for example, is formed from the subarray {*the, demonstration, show*}; repeated Merge yields DP = *[the [demonstration* P_2]] and then P_3 = [DP [*show* P_1]], with θ-roles assigned. The next subarray adds T and C, permitting the derivation to continue to P_4 = (26).

We have found evidence that computational complexity enters into language design, but one might ask whether the argument can be strengthened on purely conceptual grounds. Consider theses (28a–d), where (28b–d) presuppose a positive answer to (28a).

(28) a. Computational complexity matters for a cognitive system.
b. The solution must be comprehensive, with a guarantee of "quick decision" for all derivations.
c. Complexity should not be allowed to grow "too fast."
d. Decisions in computation attend only to principles of UG.

In recent discussion such theses are sometimes adopted as virtual conceptual necessity. [56] That is a questionable move, however.

One reason is that the theses seem to presuppose that the derivational approach is correct, which is hardly obvious (if the question is even meaningful). That aside, it is not clear why thesis (28a) should be true. Thesis (28b) raises further questions. Languages L have expressions that satisfy all conditions of UG and L but are "unusable" for some reason; see note 17. It could turn out that among these are expressions that cannot be derived efficiently, a result that would be interesting if true (see Chomsky 1991).

Interpretation of thesis (28c) in the technical sense of complexity theory also calls for justification. Why should we expect such properties to be relevant to natural language? Thesis (28d) is intended to bar look-ahead by requiring that at each choice point in a derivation, an irrevocable decision must be made in terms of principles of UG. The intuitive idea is that only "fundamental principles" of UG can be invoked, not consequences of these principles (however easily determined). Again, that is not an obvious conclusion.[57]

Theses (28c) and (28d) are often held to undermine the account of (9)–(10)/(12)/(15)/(16) reviewed earlier on the grounds that it involves look-ahead. If that were true, the validity of the theses would be called into question, not the analysis, unless some justification can be found for them. Notice how difficult it is even to give a clear formulation of thesis (28d), hence of the argument that the proposed analysis of these constructions even raises look-ahead issues.[58] The assumption is that thesis (28d) is violated by resort to the θ-theoretic principle (6), though apparently not by preference of Merge over Move. Why is unclear: selection of Merge over Move or conversely is determined (trivially) by principles of UG. Furthermore, how do we know that (6) is a "theorem," not a principle of language design (part of θ-theory), so that resorting to it in fact satisfies the intuition that lies behind this application of thesis (28d)?

Again, it seems that we should seek a resolution of the issues on empirical grounds. My suspicion is that thesis (28a) might be correct and perhaps (28b), along with versions of (28c,d) that restrict choices in derivations to elementary principles of UG and bar even narrowly bounded look-ahead. But the questions are empirical. If the theses of (28) hold in some form, that would be a surprising empirical fact about language design; evidence is required to establish it. The best evidence I know is twofold: observed cases seem to support thesis (28a) and to be consistent with (28b). Inquiry seems to reveal further that postulated properties of language that induce computational complexity were incorrectly formulated and that when they are improved, undue complexity is overcome and computation is "local," suggesting that thesis (28c) may be valid. [59] Furthermore, there seems no need to invoke principles beyond the most elementary ones at any point. If such results are consistently found, they would provide confirming evidence for versions of (28), and for the derivational approach, which provides the framework within which they are formulated.

3.4— The Strong Thesis

Let us now consider the strongest minimalist thesis (2) and ask where it fails. The research strategy is to seek "imperfections" of language, properties that language should not have, assuming (2). Apparent imperfections come in several varieties. Given some apparent property P of language, the following outcomes are possible:

(29) a. P is real, and an imperfection.
b. P is not real, contrary to what had been supposed.
c. P is real, but not an imperfection; it is part of a "best way" (perhaps not unique) to meet design specifications.

The most likely—hence least interesting—outcome is (29a). Conclusion (29b) is less likely, therefore more interesting; efforts within the Minimalist Program have sought to establish it in particular cases. Conclusion (29c) is the most interesting possibility. The question is novel, perhaps premature—or unreasonable, because the strongest minimalist thesis (2), or even weaker versions, is not correct, as one might expect for some complex biological system.

Adopting (2) as a point of departure, assume that FL provides no machinery beyond what is needed to satisfy minimal requirements of legibility and that it functions in as simple a way as possible. We would like to establish such conclusions as these: [60]

(30) a. The only linguistically significant levels are the interface levels.
b. The *Interpretability Condition*: LIs have no features other than those interpreted at the interface, properties of sound and meaning.
c. The *Inclusiveness Condition*: No new features are introduced by C_{HL}.
d. Relations that enter into C_{HL} either (i) are imposed by legibility conditions or (ii) fall out in some natural way from the computational process.[61]

Condition (30a) requires that there is no Deep or Surface Structure, or other levels that have been proposed. It holds that everything accounted for in these terms has been misdescribed and is better understood in terms of legibility conditions at the interface: that includes the Projection Principle, binding theory, Case theory, the Chain Condition, and so on.

Condition (30b) is transparently false.

Condition (30c) permits rearrangement of LIs and of elements constructed in the course of derivation, and deletion of features of LI—but optimally, nothing more.

Condition (30d) has to be spelled out. Properties induced by legibility conditions might include adjacency, argument structure, scope, and the like. Those of category (ii) should include at least (perhaps at most) the relations provided directly by the indispensable computational operation Merge. But there should be no government, no stipulated properties of chains, no binding relations internal to language, no interactions of other kinds. It is hardly necessary to observe that all of this is highly unlikely. There is substantial empirical evidence supporting the opposite conclusion at every point. Furthermore, a basic assumption of the work in the P&P framework, with its impressive achievements, is that everything just suggested is false: that language is highly "imperfect" in these respects. It would be no small task, then, to show that this apparatus is a kind of descriptive technology, and that if we abandon it, we can maintain or even extend descriptive and explanatory force. Nevertheless, recent work suggests that such conclusions, which seemed out of the question a few years ago, are at least plausible, possibly correct in nontrivial respects.

Much of *MP* is devoted to establishing parts of (30) in terms of bare phrase structure, with a cyclic notion of generalized transformations and reinterpretations of binding and Case theory. I will assume that these are the right directions, though many serious problems remain.

Consider condition (30c). It requires that there be no phrasal categories or bar levels, hence no X-bar theory or other theory of phrase structure, apart from bare phrase structure, which will be simplified further below. It also rules out introduction of traces, indices, λ-operators, and other new elements in the course of operation of C_{HL}. Recourse to such devices could be innocuous (e.g., if used for convenience to annotate properties that can be determined by inspection at LF), but questions arise if they enter into interpretation and function significantly within the computation—for example, percolation of indices, or operations that apply specifically to trace.

In *MP* indices are introduced, but not within C_{HL} itself—rather, in extending lexical arrays to numerations (see note 28). This narrow departure from condition (30c) could be eliminated by weakening the requirement that an item of a lexical array be removed when accessed in computation, leaving this as an option. Whether that is the correct move depends on the consequences. A minor matter is that it would require modification of interpretive procedures at the LF interface. More important, it would require a new notion of chain. That seems to be the only significant consequence, and it merits some thought.

If α in the syntactic object SO is merged somewhere else (by the operation Move) to form SO', then the two occurrences of α constitute a chain, the original occurrence called the *trace or copy* of the new one. The terminology is misleading, for several reasons. First, each of the elements is a "copy" of the other. Second, copy theory is the simplest version of transformational grammar, making use only of Merge, not Merge followed by an operation that deletes the original—and, under trace theory, a further operation that creates a new kind of element, trace, a serious violation of the Inclusiveness Condition. These are "imperfections," to be avoided unless shown to be necessary. I will continue to use the terminology, but only for expository convenience, adopting the copy theory as the null hypothesis. [62]

If LAs are extended to numerations Num, a chain can be defined as a sequence of identical α's—more accurately, a sequence of *occurrences* of a single α. That seems to be the simplest way to characterize chains; hence, it is to be adopted unless shown to be empirically inadequate. Thus, in (12c), repeated as (31), there are two occurrences of *I* and two occurrences of *a proof* (one where the terms are spelled out and another at the trace).

(31) I_i T [t_i expect [[a proof]$_j$ to be discovered t_j]]

Chains so-defined can be formed only by movement, given LA = Num. Basic properties of chains should then follow from elementary derivational principles. That would be a good result, eliminating stipulated properties of chains and explaining why these hold. If achievable, it would also bear on broader questions that arose in section 3.3, providing support for the derivational interpretation of the recursive function C_{HL}, hence for the surprising conclusion that there is even an empirical issue concerning alternative interpretations.

What exactly do we mean by an "occurrence of α"? To ensure that occurrences in the intended sense are distinguished, we can take an occurrence of α in K to be the full context of α in K. [63] In *MP* a simpler notion is proposed: an occurrence of α is a sister of α. Then in (31) the occurrence of matrix subject *I* is I_1 = *T . . . discovered* t_j and the occurrence of its embedded copy is I_2 = *expect . . . discovered* t_j (actually, the syntactic objects corresponding to them). The occurrences of *a proof* and its trace are P_1 = *to be discovered* t_j and P_2 = *discovered*, respectively.

In (31) two chains are defined: $C_I = \langle I_1, I_2 \rangle$ and $C_P = \langle P_1, P_2 \rangle$. In informal description C_I is the chain $\langle I, t_i \rangle$ and C_P the chain <*a proof,* t_j>, where *I* and *a proof* are called the *heads* of the respective chains.

In (31) occurrences are properly identified if taken to be sisters, but that might not always be the case.[64] The simplified definition in terms of sisterhood is based on the assumption that other properties of C_{HL} guarantee that no problems arise. I will assume that to be true.

We can simplify chains from sequences to sets, relying on the fact that a "higher" occurrence of α properly contains lower ones. Thus, in C_I of (31) I_1 properly contains I_2 (*I* c-commands its trace). Chains formed by successive-cyclic movement fall into place directly. Consider (32).

(32) a. a proof is likely [$_\alpha$ *t* to be [discovered *t*]]

did you say . . . the proof, $Q_2 = R_1$ = *has . . . the proof*, and R_2 = *discovered the proof.* This is one of the permissible analyses into chains; it need only be permitted, not forced. In (32a) there is a three-membered chain; in the (interpretable) analysis of (32b) there are two two-membered chains, sharing one occurrence of *who*. [66]

A chain, then, is a set of occurrences of an object α in a constructed syntactic object K. It would make sense to rethink all notions involving chains in these terms, including their interpretation at the interface. If we do, some apparent problems disappear. Suppose that raising of DP to [Spec, T] checks and deletes its uninterpretable Case feature. We want DP and its trace *t* to be identical, so the feature must also delete in the trace. But what guarantees that the feature is deleted throughout the chain? The question does not arise if we think of a chain, more properly, as a set of occurrences of α in K: the feature is deleted in the single element α, unproblematically.

Should the notions of occurrence and chain be extended to proper subparts of LI that are not elements of the lexical array: features and sets of features? Let's put the question aside for the moment and turn to the relations permissible under condition (30dii). Consider the operation Merge (indispensable in some form). Merge takes two objects α and β and forms a new object K(α, β). The operation provides two relations directly: *sisterhood*, which holds of (α, β), and *immediately contain*, which holds of (K, α), (K, β), and (K, K) (taking it to be reflexive). Suppose we permit ourselves the elementary operation of composition of relations. Applying it in all possible ways, we derive three new relations: the transitive closure *contain of immediately contain, identity* (= (sister(sister)), and *c-command* (= sister(contain)). Thus, K contains α if K immediately contains α or immediately contains L that contains α; conversely, α is a *term* of K if K contains α. And α c-commands β if α is the sister of K that contains β.[67]

The relation of c-command is available, and expected, on very weak assumptions. The relation has played a large role in syntactic theory, though it may be that it does not function within narrow syntax but only in interpretation of the information it provides—that is, in mapping it to syntactic objects that belong to mental systems external to the language faculty itself (see note 44). That might be expected if external systems access representations (PF and LF) to which the notions "sister" and "contain" apply.

The sisterhood relation is significant primarily (perhaps only) for heads, that is, LIs and modified LIs (MLIs) formed from them. Furthermore, sisterhood relations presumably remain if LI is modified to MLI: if VP is sister of T, for example, it should remain so even if uninterpretable features are deleted from T. More generally, LI and its modifications are not distinguished with regard to the fundamental relations defined in terms of Merge. For α an LI, then, we extend the relations defined for (α, β) to (MLI(α), β), MLI constructed from α (typically—perhaps only—by deletion of uninterpretable features). We extend the same convention to a feature F(α) of a head α. Thus, if LI α is the sister of β or c-commands β, then MLI(α) and F(α) do as well.

Questions have arisen about the interpretation of these notions for adjunction, particularly head adjunction of H' to H. With no further elaboration, H and H' would be sisters and neither would c-command outside. Whether this matters is unclear. Failure of H' to c-command its trace seems to have no significance in the present framework; failure of H to c-command into its former c-command domain would be problematic only if H functions after adjunction in implementing agreement and movement, but that does not seem necessary. Pending some good reason to sharpen the sisterhood relation for this case, I will leave it as is. [68]

Though varieties of government would be "imperfections," to be avoided if possible, the closer-to-primitive notion of L-marking should pass muster, hence also notions of barrier that are based on nothing more than L-marking. Here numerous questions arise about island conditions within a minimalist framework, about which I have nothing useful to say.[69]

3.5— Imperfections

There are some respects in which the strong thesis seems untenable, and we find what appear to be "design flaws" that are not necessary for language-like systems. The most obvious involve the phonological component, which takes syntactic objects constructed by the computational operations C_{HL} and converts them to representations at the PF interface. Here there are radical violations of the Interpretability and Inclusiveness Conditions (30b,c). The Inclusiveness Condition is violated by operations that introduce such new elements as prosodic structure and narrow phonetics. The Interpretability Condition is violated by the discrepancy between the phonological properties of LIs ("morphophonemes," "phonological units," etc., within various frameworks and terminologies) and the narrow phonetic instantiations of combinations of such elements. It may be that phonological features of LIs do not even appear at the level PF, that the "input" and "output" in different "vocabularies." In that case the Interpretability Condition would be maximally violated by th component, and the Inclusiveness Condition is clearly inoperative.

We therefore turn to the alternatives of (29). Option (29b) seems implausible; the properties appear to be expression. Of the two remaining possibilities, the more interesting by far is (29c). The properties are rea engineer, called upon to map independently motivated syntactic objects to PF, would hit upon the phono solution. I have no idea whether this can be formulated as a sensible research task. As for option (29a), it from "good design" are not surprising in this domain. Direct evidence about sound systems is available f degree of complexity and variation. And the subsystem reflects special properties of the sensorimotor sy "extraneous" to language, relating to externalization by systems with nonlinguistic properties and capabl fixed, as in sign languages. Symbolic systems designed for special purposes (metamathematics, compute phonological component, not facing the need to meet the legibility conditions for human language at the

The strongest sustainable inclusiveness/interpretability requirement, then, is (33).

(33) Inclusiveness holds of narrow syntax, and each feature is interpreted at the level LF or associated with phonetic

The phonological component is generally assumed to be isolated in even stronger respects: there are *true* only to the phonological component and form a separate subsystem of FL, with its own special propertie course of construction of LF, an operation Spell-Out delivers the structure already formed to the phonolc to PF. If LIs express Saussurean arbitrariness in the conventional way, then Spell-Out "strips away" the t derivation can converge at LF; it will crash if later operations introduce

LIs with phonological features. On the assumptions of Distributed Morphology, the phonological features are introduced after Spell-Out by phonological operations applying to LIs lacking them. I will assume some instantiation of this array of options to be correct.

Narrow syntax also involves devices that are imperfections unless shown to be unreal (option (29b), which again seems implausible) or to be motivated by design specifications (option (29c), the most interesting possibility). Consider two striking examples:

(34) a. Uninterpretable features of lexical items
b. The "dislocation" property

Under (34a) we find features that receive no interpretation at LF and need receive none at PF, hence violating any reasonable version of the Interpretability Condition (30b). The example that has played the most important role ever since Jean-Roger Vergnaud's famous unpublished letter written twenty years ago is structural Case. The picture is more complex for agreement features: semantically interpretable for nouns, but not for verbs/adjectives, and phonetically optional throughout. [70] External manifestation of inflectional features appears to be the locus of much of the variety languages display, a topic that has gained prominence within the P&P framework.

These observations presuppose that occurrences of features can be distinguished, a question raised earlier but put off: occurrences of agreement features are distinguished by category, some interpretable, some not. That falls well short of truly identifying occurrences. To do so would be necessary if feature chains exist.[71] In the absence of clear evidence to the contrary, I will assume that feature chains do not exist, hence that features cannot move or be attracted.[72]

The dislocation property (34b) is another apparent imperfection. In (35), for example, the phrase *an unpopular candidate* is in the natural position to be interpreted as object of *elect* in (35a,b) but not in (35c,d), though the interpretation is in relevant respects the same.

(35) a. they [elected an unpopular candidate]
b. there was [elected an unpopular candidate]
c. an unpopular candidate was elected
d. there was an unpopular candidate elected

In (35c,d) the surface phonetic relations are dissociated from the semantic ones.[73] Such phenomena are pervasive. They have to be accommodated by some device in any adequate theory of language, whether it is called "transformational" or something else.

Dislocation of α yields a chain (α, *t*)—more accurately, a chain {X, Y}, where X and Y are occurrences of α. The raised element typically c-commands its trace in the original position, but where true, that follows from independent properties of C_{HL}. Further operations might lead to violation of c-command and of locality relations between the two positions, as in multiple head raising or independent XP-dislocation.

(36) a. $[[V_j\text{-}T]_i\text{-}C]\ [\ldots t_i \ldots [_{VP} \ldots t_j \ldots]]$
b. [written t_j for children]$_i$, [those books]$_j$ couldn't possibly be t_i]

In (36) the indices are redundant, the chains being determined by constitution of the trace. That need not be so, however, as in (37).

(37) whom$_i$ did everyone talk to whom$_j$ about whom$_k$

Chains can be formed with i = j or i = k, both consistent with locality conditions; and the full range of interpretations seems to be available, either way.

In the approach we are pursuing here, the chains at LF are determined by identity throughout, the ambiguity of (37) being resolved by the derivation, given the initial numeration. Other means would be required if we were to eliminate this device along lines discussed earlier. In a strict derivational approach semantic interpretation is cyclic and the problem of interpreting (36) and (37) arises in a somewhat different form. In representational approaches chains are determined by an algorithm A operating in a "search space" of possible LFs; the burden of accounting for locality and other conditions on chains then rests on A. Here we return to issues of general architecture raised in section 3.3.

However these matters are resolved, we have two "imperfections" to consider: uninterpretable features and the dislocation property. These properties (in fact, morphology altogether) are never built into special-purpose symbolic systems. We might suspect, then, that they have to do with externally imposed legibility conditions. With regard to dislocation, that has been suggested from the earliest days of modern generative grammar, with speculations about facilitation of processing (on the sound side) and the dissociation of "deep" and "surface" interpretive principles (on the meaning side). The boundaries are not clear, [74] nor are the mechanisms to express them. One approach to the array of problems was to distinguish the role of deep and surface structure (D- and S-Structure) in semantic interpretation: the former enters into determining quasi-logical properties such as entailment and θ-structure; the latter such properties such as topic-comment, presupposition, focus, specificity, new/old information, agentive force, and others that are often considered more discourse-oriented and appear to involve the "edge" of constructions. Theories of LF and other approaches sought to capture the distinctions in other ways. The "deep" (LF) properties are of the general kind found in language-like systems; the "surface" properties appear to be specific to human language. If the distinction is real, we would expect to find that language design marks it in some systematic way—perhaps by the dislocation property, at least in part.

To the extent that such ideas can be given substance, it would follow that the dislocation property is required; it falls within the design specifications given to the super-engineer seeking an optimal solution to conditions imposed by the external systems.

This line of argument might provide motivation for the dislocation property, but it would remain to find the mechanisms employed to implement it. The distinction is familiar. We may say that the function of the eye is to see, but it remains to determine the implementation (a particular protein in the lens that refracts light, etc.). Similarly, certain semantic properties may involve dislocated structures, but we want to discover the mechanisms that force dislocation. Minimalist intuitions lead us to look at the other major imperfection, the uninterpretable inflectional features. Perhaps these devices are used to yield the dislocation property. If so, then the two imperfections might reduce to one, the dislocation property. But the latter might itself be required by design specifications. That would be an optimal conclusion, falling under option (29c).

To establish any such conclusion is no simple matter. We are entering terrain that is mostly unexplored. One approach is suggested by the observation that for convergence, uninterpretable features must be deleted in the course of computation of LF. Consider the dislocated example (35c), repeated here.

(38) an unpopular candidate T-was elected *t*

There are three kinds of uninterpretable features in this structure: (a) the agreement features of T (taking them as a unit, the set of ϕ-features), (b) the EPP-feature of T that requires "second Merge," and (c) the structural Case feature of *an unpopular candidate*. The ϕ-set (a) identifies T as a target of dislocation; the EPP-feature (b) requires that something be merged in this position; the Case feature (c) identifies *an unpopular candidate* as a candidate for such merger (hence dislocation). Successful implementation of the operation erases all of the uninterpretable features, forming MLIs with a reduced set of features. The approach is optimal, if indeed uninterpretable features are the mechanism for dislocation. [75]

Assuming so, let us look more closely. Suppose that the derivation has constructed the SO (39), having merged T with the copula-headed phrase.

(39) T be elected an unpopular candidate

The new element T has uninterpretable features of two types: its ϕ-set and its selectional feature EPP. Like other selectional features, EPP seeks an XP to merge with the category it heads. The ϕ-set we can think of as a *probe* that seeks a *goal*, namely, "matching" features that establish agreement. The relation of the probe of T to its goal is the *T-associate* relation.

For the ϕ-set of T in (39), there is only one choice of matching features: the ϕ-set of *candidate*. Locating this goal, the probe erases under matching. Taking structural Case to be a reflex of an uninterpretable ϕ-set,[76] it too erases under matching with the probe. The erasure of uninterpretable features of probe and goal is the operation we called *Agree*. But the EPP-feature of T must also be satisfied—in this case by "pied-piping" of a phrase P(G) determined by the goal of T's probe, which merges with (39), becoming [Spec, T]. The combination of selection of P(G), Merge of P(G), and feature deletion under matching (Agree) is the composite operation Move, which dislocates *an unpopular candidate*, eliminating all uninterpretable features.

Matching is a relation that holds of a probe P and a goal G. Not every matching pair induces Agree. To do so, G must (at least) be in the *domain* D(P) of P and satisfy locality conditions. The simplest assumptions for the probe-goal system are shown in (40).

(40) a. Matching is feature identity.
b. D(P) is the sister of P.
c. Locality reduces to "closest c-command."

Thus, D(P) is the c-command domain of P, and a matching feature G is *closest to* P if there is no G' in D (P) matching P such that G is in D(G').

In the absence of evidence to the contrary, we adopt (40), with a qualification taken over from earlier work.[77]

(41) Terms of the same minimal domain are "equidistant" to probes.

The minimal domain of a head H is the set of terms immediately contained in projections of H.

With matching restricted to identity, Case and lexical category cannot enter into Agree or Move, since the probes do not manifest these features. And OS must be an implementation of (here invisible) object agreement, with ancillary Case checking.

If uninterpretable features serve to implement operations, we expect that it is structural Case that enables the closest goal G to select P(G) to satisfy the EPP by Merge. Thus, if structural Case has already been checked (deleted), the phrase P(G) is "frozen in place," unable to move further to satisfy the EPP in a higher position. More generally, uninterpretable features render the goal *active*, able to implement an operation: to select a phrase for Merge (pied-piping) or to delete the probe. The operations Agree and Move require a goal that is both local and active. [78]

We therefore have the possibility of *defective intervention constraints* in a structure (42), where > is c-command, β and γ match the probe α, but β is inactive so that the effects of matching are blocked.

(42) $\alpha > \beta > \gamma$

We will return to some illustrations.[79]

In *MP* Agree is analyzed in terms of feature movement (Attract) and a concept of matching that is left unclear. Here we take matching to be identity and dispense with Attract, with complications it induces about extended MLIs, feature chains, and other matters. Checking reduces to deletion under matching with an active local goal and ancillary deletion of the uninterpretable feature that rendered the goal active. I will use the terms *checking* and *attract* only for convenience.

Suppose that the EPP-feature of T could be satisfied more simply than by the full operation Move. That is the case in (35b), repeated here.

(43) there [$_\alpha$ T-was elected an unpopular candidate]

Here the lexical array includes the expletive *there*. At stage α of the derivation (= (39)), the independent operations Agree and pure Merge suffice: Agree deletes the ϕ-set of T and the structural Case of *candidate*, and Merge (of *there*) satisfies the EPP-feature of T. The more complex operation Move is preempted; dislocation does not take place, though we have long-distance agreement of T and its goal (its associate).[80]

Manifestation of structural Case depends on interpretable features of the probe: finite T (nominative), v (accusative), control T (null), on our earlier assumptions. We may therefore regard structural Case as a single undifferentiated feature. The same would be expected for the uninterpretable ϕ-set of the probe. Its manifestation depends on interpretable features (namely, ϕ-features) of the goal, so that it too can be taken to be undifferentiated with respect to the value of the individual features of the ϕ-set ([+/–plural], etc.). For both probe and goal, the form of the uninterpretable features is determined by Agree. To rephrase in traditional terms, verbs agree with nouns, not conversely, and Case is assigned.

We therefore understand "feature identity" in (40a) to be identity of the choice of feature, not of value. More important, defective intervention effects are induced whether or not β and γ of (42) are identical in ϕ-feature value (singular blocks plural agreement, etc.). This lends theory-internal support to the earlier observation that ϕ-features are interpretable only for N; their value is specified only in this case. Notice also that only the most underspecified element, PRO, can have null Case, so raising of $\alpha \neq$ PRO to [Spec, T] causes the derivation to crash when T is a control infinitival.

We take deletion to be a "one fell swoop" operation, dealing with the ϕ-set as a unit. Its features cannot selectively delete: either all delete, or none. The ϕ-features of T do not agree with different NPs, for example. In the same spirit, we assume that only a probe with a full complement of ϕ-features is capable of deleting the feature that activates the matched goal. Suppose that the probe for participial (like adjectival) α is a ϕ-set lacking the feature [person] and that G is the closest matching goal in its search space: P(G) = DP may be attracted to [Spec, α], deleting the probe of α (participial agreement), but the operation will not delete structural Case in DP, which can move on to [Spec, T], deleting the probe of T and the Case of DP (subject agreement). v and nondefective T, with a full complement of ϕ-features, delete the uninterpretable feature that activates the matched goal (raised or not). [81]

How would noncontrol infinitivals (T_{def}) and weak expletives Expl of the *there*-type fit into this picture? The former category falls into place if T always has at least a minimal feature complement, perhaps only [person] for T_{def}. If so, Move of α to [Spec, T_{def}] will delete the ϕ-set of T (= uninterpretable [person]) but not the structural Case feature of α, so that α can undergo further movement and agreement. The phase head v/C have no counterpart to T_{def} with a reduced ϕ-set and therefore do not provide an "escape hatch" for successive-cyclic A-movement.

Weak Expl shares the basic movement/attraction properties of nominals. That is expected if Expl has an uninterpretable feature F that activates it until erased and a ϕ-set G that matches a probe in T. But G is uninterpretable for Expl, so a distinct F is unnecessary, obviating the need for structural Case in Expl. The composition of G is determined by two conditions: (a) Expl can raise to [Spec, T_{def}]; (b) Expl cannot delete the probe of nondefective T. Condition (a) requires that G contain a feature to match the probe of T_{def} ([person], if what precedes is correct). From (b) it follows that G must be less than a full ϕ-set, hence optimally just [person]. That (b) holds is shown by long-distance agreement structures such as (44b–d), (44b) surfacing commonly as (44c), or in English more naturally as (44d), as noted.

(44) a. they declared [three men guilty]
b. there were declared [three men guilty]
c. there were declared guilty three men
d. there were three men declared guilty

If the matching feature of the probe were deleted by the operation, it would not be available for associate matching and the nominative Case of the associate would remain unchecked because of the lack of a full complement of features in T (compare participial agreement). The problem does not arise if (b) holds and uninterpretable features delete in an "all or none" fashion, not selectively. In (44b–d) the full complement of ϕ-features of T deletes the uninterpretable feature G of *there*, barring further raising. When Expl raises to [Spec, T_{def}], the probe (a single feature) deletes under matching as before, but G does not, because deletion requires matching with a full complement of ϕ-features of the probe. Therefore, successive-cyclic raising through [Spec, T_{def}] is possible. **82**

Reinterpretation of Attract in terms of Agree eliminates the need to introduce "checking domains." That is a step forward. The notion is complex, and furthermore unnatural in minimalist terms; feature checking should involve features, nothing more, and there is no simpler relation than identity. More important, the notion is irrelevant for the core cases: elements merge in checking domains for reasons independent of feature checking; and feature checking takes place without dislocation to a checking domain.

As discussed, both properties are illustrated in expletive constructions. Much work on the topic has taken long-distance effects to be a property of these constructions, hence of an expletive-associate relation; various ideas have been explored regarding how that relation is established. In *MP*, chap. 4, a different approach is suggested: the long-distance effects are attributed to a T-associate relation that involves features only and is independent of the expletive. The reasons were theory-internal, but a broader range of cases adds empirical support. Long-distance effects are found without expletives in such constructions as (15a), [83] the EPP being satisfied by raising of quirky Case; and expletive subjects are found without T-associate agreement when there is no accessible nominative. We will return to some illustrations. The general conclusions are these:

(45) a. Long-distance agreement is a T-associate (probe-goal) relation.
b. The EPP can be satisfied by
i. Merge of expletive
ii. Merge of associate
iii. Merge of α closer to T than the associate

Case (45bi) is illustrated by T-associate agreement, with the definiteness effect. Case (45bii) exhibits agreement of [Spec, T] and T, but that is ancillary to the T-associate relation. In case (45biii) there is no definiteness effect and long-distance T-associate agreement holds with embedded accessible nominative; or, if such an associate is lacking, T is default.

More generally, we should not expect Spec-head relations to have any special status. Within bare phrase structure, we cannot, for example, take the result of first Merge to α to be sometimes a specifier and sometimes a complement, as in an X-bar-theoretic analysis that takes the object of α to be its complement (*see John, proud [of John]*) but the subject of objectless α to be its specifier (base forms of *John eat, John proud*). The restriction to a single specifier is also questionable: rather, we would expect first Merge, second Merge, and so on, with no stipulated limit.

3.6— Syntactic Objects

We are taking the elements that enter into derivations to be features and objects constructed from them in a restricted way:

(46) a. Lexical items LI
b. Modified lexical items MLI
c. Sets K constructed from given elements α, β

An MLI is an LI with uninterpretable features deleted. In case (46c) K corresponds to the subtree dominated by a node in a standard phrase structure diagram; there are no objects corresponding to the nodes, and no nonbranching projections.

θ-structure and similar semantic roles are based on pure Merge of XP to substantive LIs or their projections. Checking theory, in contrast, involves uninterpretable features of functional categories and is reduced to feature matching under conditions (40a–c). The two theories differ in nature and implementation—plausibly, since they are conceptually quite distinct. Something like θ-theory is a property of any language-like system, whereas checking theory is specific to human language, motivated (we are speculating) by legibility conditions. The Chain Condition is an expression of this duality and should fall out as a descriptive observation, along with other properties of chains. [84] Apart from Merge of selected XP, narrow syntax involves only feature deletion to form reduced MLIs, sometimes associated with Merge in $\bar{\theta}$ (non-θ) positions yielding dislocation.

According to this conception, agreement (hence movement) is driven by uninterpretable features of the probe, which must be deleted for legibility. The operation Greed of *MP*, in contrast, was driven by uninterpretable features of the goal. The principle we are now entertaining is what Lasnik (1995a,b) calls Enlightened Self-Interest, with the further requirement that the matched probe delete: we might call the principle Suicidal Greed. Suicidal Greed does not have the "look-ahead" property of Greed, a complexity reduction that could be significant, as discussed.

With this shift of perspective, structural Case is demoted in significance. The Case Filter still functions indirectly in the manner of Vergnaud's original proposal, to determine the distribution of noun phrases. But what matters primarily are the probes, including ϕ-features of T, v. That reverses much of the recent history of inquiry into these topics and also brings out more clearly the question of why Case exists at all. The question arises still more sharply if matching is just identity, so that Case can never be attracted; operations are not induced by Case-checking requirements. Recall that lexical category also cannot be attracted and does not induce operations, raising the same questions.[85] For Case, a plausible answer is the one already suggested: uninterpretable features activate the goal of a probe, allowing it to implement some operation (Agree or Move).[86] It follows that after structural Case of DP is deleted, the phrase cannot move further to an A position and its ϕ-set cannot induce deletion (though it is still "visible" to a probe, allowing defective intervention effects as in (42)). Suppose quirky Case is (θ-related) inherent Case with an additional structural Case feature, as often suggested in one or another form. Then it too is immobile once it reaches a Case-checking position. [87] If the ϕ-features of T that check the structural Case of raised quirky Case themselves delete, we have default T; if they remain, we have remote agreement with some lower accessible nominative.

The descriptive observations seem generally accurate. In these terms, the visibility thesis and the Chain Condition reduce to LF convergence.

In his detailed review of Icelandic agreement, Sigurðsson (1996) concludes that remote nominative Case allows number agreement but not first/second person agreement. That would follow if the [person] feature of T reduces to [3person] (the default choice) when it attracts quirky Case or Expl to [Spec, T].[88]

Suppose expletive Expl is merged in [Spec, T] without movement. Agreement is manifested for *it*-type Expl with a full complement of ϕ-features, and *there*-type Expl becomes inactive, indicating that its uninterpretable feature deletes ([person], I have assumed). But these are properties of Agree, not Merge.[89] The results are expected if Expl is an X^0 head and its [person] feature is uninterpretable, therefore able to probe its domain T' (= D(Expl)), locating the ϕ-set of T as the closest goal. The uninterpretable probe deletes, and the ϕ-set of T as well if Expl has a full ϕ-set. We therefore have agreement but not via merger: full or partial agreement depending on the ϕ-set of Expl, which becomes inactive.[90]

The head of an A-chain can undergo $\bar{A}$-movement, of course, with different features accessed. Take *wh*-movement. This would be point by point analogous to A-movement if the *wh*-phrase has an uninterpretable feature [wh] and an interpretable feature [Q], which matches the uninterpretable probe [Q] of a complementizer in the final stage; successive cyclicity could then function in the manner discussed.[91] The *wh*-phrase is active until [wh] is checked and deleted. The *Wh*-Island Constraint is then a defective intervention effect of the type (42): [Q] of the already checked *wh*-phrase (β in (42)) bars attraction of lower [Q], but cannot move or check the uninterpretable feature of the probe. A possible analysis of *wh*-in-situ constructions is that [wh] pied-pipes only the head (overtly or covertly).[92]

The reasoning extends to such constructions as (47a–d) ((47b) based on OS).

(47) a. *[John to seem [t^I is intelligent]] (would be surprising)
b. *(we hoped) [PRO to be decided [t^I to be killed at dawn]]

c. *[$_{DO}$ this book] seem [t_{DO} to read [t_{DOI} $_{[never\ [[Subj}$ any students] t_{read}]]]]

d. *there seem [$_{\alpha}$[$_{Subj}$ several people]^{I} are [$_{Pred}$ friends of yours]]

The EPP is satisfied throughout, and if local matching sufficed for agreement, the expressions should converge with uninterpretable features deleted. Appeal to such principles as "maximal checking" would not make the proper distinctions locally (e.g., barring (47c) in favor of subject raising). But in all cases the position superscripted *I* is inactive, hence unable to raise (47a–c) or to delete the features of a matched probe (47d). Case (47d) illustrates a defective intervention effect of type (42): Subj is visible (barring Pred as goal) but inactive, unable to establish agreement with matrix T. 93 The same property holds in (48).

(48) a. *there were decided [$_{\alpha}$ PRO to stay with friends]

b. *XP T-seem that [$_{\alpha}$ it was told friends CP]

PRO and *it* are inactive, their structural Case feature having been checked and deleted in α. But their ϕ-features remain visible and block association of matrix T to *friends*, both of which therefore retain uninterpretable features. Case (48b) is therefore barred with pure Merge of Expl, or raising of *it* or *friends* ("superraising").94

We are now in a position to derive the basic structural properties of CFCs ((5), extended to (50)), in the configuration (49).

(49) α = [XP [(EA) H YP]]

(50) a. If H is v/C, XP is not introduced by pure Merge.

b. In the configuration [$_{\beta}$ H_{β} . . . α]. H_{β} a CFC and β minimal,

i. if H_{α} is C, H_{β} is independent of α;

ii. if H_{α} is v, $H_{\beta} = T_{\beta}$ agrees with EA, which may raise to [Spec, T_{β}] though XP cannot;

iii. if H_{α} is T_{def}, if H_{β} is T then XP raises to [Spec, T_{β}] if there is no closer candidate γ for raising; and if H_{β} is v then XP agrees with v (as may a lower associate if XP = Expl).

(50a) has already been dealt with (see discussion following (24)), so we can turn to (50b).

In case (50bi) α = [XP [C TP]], TP headed by nondefective T_{α}. We can limit attention to T_{α} with its ϕ-set deleted; were the ϕ-set not deleted, the derivation would have crashed at α. Hence, agreement is fully established with the closest associate Assoc that matched T_{α}; Assoc is either raised to

[Spec, T_α] or remains in situ in a long-distance agreement relation, its structural Case feature deleted in either case. Any structural Case feature remaining in α is inaccessible by virtue of the defective intervention effect induced by Assoc, again causing crash detectable at α. Case (50bi) follows: the ϕ-set and EPP-feature of H_β have to be satisfied independently of α.

Consider case (50bii), with α = [XP [EA [v YP]]. XP is raised from within YP, checking and deleting the ϕ-set of v (object agreement) and its own structural Case feature. XP is therefore inactive for A-movement and cannot check the ϕ-features of T_β. These can only be deleted by agreement with EA, deleting its structural Case feature as well, with EA either raising to [Spec, T_β] or remaining in situ. Recall that XP does not induce a defective intervention effect that would bar the T_β-EA relation; see (41). [95]

In case (50biii) α = [XP [T_{def} YP]]. Whether merged (hence Expl) or raised, XP has an uninterpretable feature that renders it active for the Case/agreement system and must be checked outside α by H_β that is either v (ECM) or T (raising). Suppose H_β is v. If XP = Expl, its uninterpretable feature [person] deletes but the ϕ-set of v remains, able to check the Case (accusative) of a lower associate (*I expect [there to be a proof discovered]*); there is no defective intervention effect. If XP is an argument, it is the associate of v and object agreement with (accusative) Case checking proceeds as before. In either case XP raises to [Spec, v] only when v has an EPP-feature (see note 95). If H_β is T, XP raises to [Spec, T_β] unless barred by a closer candidate γ. If T_β is nondefective, it either agrees with XP and checks its Case (nominative) or shifts to default, as already discussed. If T_β too is defective, then XP raised to [Spec, T_β] will have to be associated with a still higher T or v, by raising or agreement in situ.

The basic properties (50) of CFCs therefore follow from simple and plausible assumptions. Without running through cases, it should be clear that the Phase-Impenetrability Condition (21) holds for A-movement for the same reasons (see note 47).

Further insight into these matters should derive from raising constructions with quirky Case moving to matrix subject, as in (51). [96]

(51) a. me(dat) thought(pl) [t_{me} [they(pl,nom) be industrious]]
b. *me(dat) seem(pl) [t_{me} [John(dat) to like horses(pl,nom)]]
c. *John seems(sg) me(dat) [t_{John} to like horses]

The matrix verb agrees with the embedded nominative in (51a), but not in (51b), which requires default inflection because of the defective intervention effect: the ϕ-features of *John* block the T-associate relation between T-*seem* and nominative *horses*. In (51a) as well a phrase with ϕ-features intervenes between matrix T and nominative, namely, the trace of the quirky dative *me*. But the latter is not the head of an A-chain, unlike in (51b) (also (47d) and (48)). Case (51c) is blocked by locality; quirky dative, with structural Case, is accessible. The conclusions are as before, but sharpened: it is only the head of the A-chain that blocks matching under the locality condition (40c). A-movement traces are "invisible" to the probe-associate relation; or from another perspective, the A-chain itself (regarded as a set of occurrences of α) constitutes the barrier.

This account relies on restricting basic operations to Merge and Agree, based on feature matching (identity) and driven by Suicidal Greed. Suppose that there is a Move α operation dissociated from matching and dependent on properties of α, with locality expressed as "Shortest Movement." In the defective intervention cases, matrix T is the closest target that can check the Case feature of the potential associate α. With Move α available and constrained only by locality, the derivation converges (incorrectly) with raising or agreement of α. Further constraints are required, then, if this option exists. [97]

A crucial property of deletion is that a deleted feature is invisible at LF and inaccessible to C_{HL} (the [±active] property), but accessible to the phonological component. This property poses a problem on the assumption made in *MP* that Spell-Out applies at a single point in a derivation: pre-Spell-Out, the probe must delete when checked yet remain until Spell-Out.[98] The natural conclusion is that Spell-Out is associated with agreement. Deleted features are literally erased, but only after they are sent to the phonological component along with the rest of the structure Σ—possibly at the phase level. Spell-Out therefore applies cyclically in the course of the (narrow syntactic) derivation. I will assume that this approach, apparently the simplest and most principled one, is correct.[99]

The single Spell-Out thesis of *MP* retains the flavor of the Extended Standard Theory model, distinguishing overt from covert operations—pre- and post-Spell-Out, respectively. If both overt and covert operations are cyclic, then there are two independent cycles—and if operations of the phonological component are cyclic, a third cycle as well. With cyclic Spell-Out, contingent on feature-checking operations, these distinctions collapse. There is a single cycle; all operations are cyclic. Within narrow syntax, operations that have or lack phonetic effects are interspersed. There is no distinct LF component within narrow syntax, and we can dispense with troublesome questions about its apparently cyclic character. Agree alone, not combined with Merge in the operation Move, can precede overt operations, contrary to the assumptions made in *MP* and related work. Crucial cases are long-distance agreement, *wh*-in-situ, and others. Many questions arise, but they do not seem obviously unanswerable.

One question has to do with identification of chains by identity under cyclic Spell-Out. Suppose the uninterpretable features of a head H are checked at phase Ph with Ph transferred to the phonological component and H reduced to H', which is transferred at a later phase of derivation. We want to take <H', H> to be a chain, so that H will be unpronounced by general principles, but H, H' are not identical. We may therefore take chains to be determined by a relation of nondistinctness holding between (α, β) if they differ only in uninterpretable features. [100]

The principle Procrastinate is no longer formulable (at least, as before), eliminating another case of look-ahead. The concept of strength, introduced to force violation of Procrastinate, appears to have no place. It remains to determine whether the effects can be fully captured in minimalist terms or remain as true imperfections.

To implement the program in a suitably Spartan fashion, we seek to restrict C_{HL} to indispensable operations that satisfy minimalist conditions. There are two candidates: Merge, indispensable in some form, and Agree, which, we speculate, might ultimately be accounted for in terms of design specifications. We have to determine how these operations apply, seeking and questioning departures from optimal design.

The operations Merge and Agree must

(52) a. Find syntactic objects to which they apply,
b. Find a feature F that drives the operation,
c. Perform the operation, constructing a new object K.

An operation Op takes objects already constructed (perhaps in the lexicon) and forms from them a new object. Condition (52a) is optimally satisfied if Op applies to full syntactic objects already constructed, with no search—that is, if C_{HL} operates cyclically. It follows that derivations meet the condition (53).

(53) Properties of the probe/selector α must be satisfied before new elements of the lexical subarray are accessed to drive further operations.

If the properties of α are not satisfied, the derivation crashes because α can no longer be accessed. We continue to assume cyclic application of all operations, returning to some ambiguities. Let us turn now to condition (52c).

The operation Merge forms K from α, β. Minimally, K should consist only of α and β, so K = $\{\alpha, \beta\}$. More information is needed about K, however: its category (its *label*) and the nature of the merger, either substitution or adjunction—the former at least not entering into narrow syntax, on the sparest assumptions, but needed for the phonological component and LF interpretation. [101] To ensure that every category has a label, let us say that label(α) = α, for α an LI.

Adjunction has an inherent asymmetry: X is adjoined to Y. Exploiting that property, let us take the distinction between substitution and adjunction to be the (minimal) distinction between the set $\{\alpha, \beta\}$ and the ordered pair $<\alpha, \beta>$, α adjoined to β. The constructed objects K, then, are of the form $\{\gamma, \{\alpha, \beta\}\}$ (substitution) or $\{\gamma, <\alpha, \beta>\}$ (adjunction), where γ is the label of K.

The term *substitution*, adapted from earlier work, is misleading within this framework. For clarity, let us refer to substitution as *Set-Merge* and adjunction as *Pair-Merge*. I will put aside here a number of questions that arise about adjunction.[102]

On minimal assumptions, the label γ should be the label of either α or β. Hence, no matter how complex the object constructed, its label is an LI, the head selected from the lexicon that has "projected" through the derivation, or a reduced MLI. If the label is determined from α, β by general principles, then the result of merger of α, β is simply $\{\alpha, \beta\}$ or $<\alpha, \beta>$.

Are labels predictable?

Consider pure Merge. There are two cases: set-Merge and Pair-Merge. The latter adjoins α to β to form $<\alpha, \beta>$. Given the asymmetry, it is natural to conclude that the adjoined element α leaves the category type unchanged: the target β projects. Hence, adjunction of α to β forms K = $\{\gamma, <\alpha, \beta>\}$, where γ is the label of β. Eliminating redundancy, the operation forms K = $<\alpha, \beta>$.

As an operation, Set-Merge is symmetric, so one might expect either label to project. If so, the outcome would either be interpretable at LF or not. But here too properties of language design appear to determine the label without look-ahead. Set-Merge typically has an inherent asymmetry. When α, β merge, it is to satisfy (selectional) requirements of one (the *selector*) but not both. Fairly generally, furthermore, the selector is uniquely determined for a pair (α, β), as can be seen by reviewing cases.

Set-Merge of (α, β) has some of the properties of Agree: a feature F of one of the merged elements (say, α) must be satisfied for the operation to take place. Furthermore, F is in the label of α, hence detectable in an optimal way (satisfying condition (52b)). The selector F for Merge is analogous to the probe for Agree. Furthermore, F is the only element of α that enters into the operation, hence the only one available without further complication to determine the label of the merged elements. In this case too, then, the label is predictable and need not be indicated: the label of the selector projects.

The intuitive content of the Projection Principle is that for a substantive category α, the selector F is a semantic property of α, an interpretable feature. Hence, F does not delete (another difference between θ-theory and checking theory). Further specification depends on how θ-theory is understood. Take, say, transitivity of a verb V. If the property is implemented in terms of θ-grids, then a feature of V selects the object. If it is implemented configurationally as a structure $[_v$ v $[_V$ V . . .]], then v too is a relevant selector and the v-VP structure enters into the interpretation. In either case one or another form of deviance (or crash) results from inappropriate merger. These and many other questions come to the fore as the framework is more carefully articulated.

The two cases of pure Merge differ in several respects. The asymmetrical operation Pair-Merge has no selector and is optional; the symmetrical operation Set-Merge has a selector (typically unique) and is obligatory. [103] In these respects language design is close to optimal, providing just the information necessary for an operation Op to project the label L: L is determined by Op itself if Op is asymmetrical, but a selector is needed to determine L if Op is symmetrical. Accordingly, Merge has a selector for Set-Merge but not Pair-Merge, which is therefore optional. The label is determined without look-ahead to check eventual convergence (contrary to *MP*), another case confirming conclusions about complexity suggested earlier.

Pure Merge, then, satisfies the conditions of (52) rather well. Consider the second elementary operation, Agree. We are now assuming cyclicity: α is a candidate for the operation only if it is the full syntactic object under inspection, so search is unnecessary, satisfying (52a). Agree requires a probe F in α. By condition (52b), F has to be readily detectable, hence optimally in the label L(α) of α, its sole designated element. F seeks a matching feature F'. Deletion takes place under conditions already discussed. The syntactic object α is otherwise unchanged. Its label remains as before. The same properties carry over to Move, constructed from Agree and Merge. [104]

In all cases, then, the label is redundant. The syntactic objects are LIs, or sets $\{\alpha, \beta\}$ or $<\alpha, \beta>$ constructed from them. The label is determined and available for operations within C_{HL} or for interpretation at the interface, but is indicated only for convenience.

Computation is driven by a probe/selector of a label, which projects. Hence, no operation can be contingent on application of earlier ones. That seems to be a valid descriptive generalization, which falls into place. Both label determination and operations are "first-order Markovian," requiring no information about earlier stages of derivation.

Conditions (52a) and (52b) are optimally satisfied, as is (52c) for Merge and partially (so far) for Agree. Continuing with Agree, (52c) requires that the matched goal G must be easily located. We want to identify a domain D(P) of the probe P, such that G is within D(P). There are two candidates for D(P): the smallest and the largest of the categories labeled by the label containing P. The former includes only the complement of P; the latter its specifiers as well. Search space is more limited if D(P) is the smallest category, as so far assumed; see (40b). It is restricted further by the "closest match" condition (40c). Natural complexity/economy conditions are again satisfied, along the lines of (3), (14), and their extensions; there is substantial evidence that G must be in the complement of the probe P, not its specifiers, and that locality conditions enter into choice of G.

The basic operations Merge and Agree satisfy reasonable "good design" conditions (52). The conclusions extend to the third operation, Move, insofar as it is constructed from the basic operations.

Move of β, targeting α, has three components.

(54) a. A probe P in the label L of α locates the closest matching G in its domain.
b. A feature G' of the label containing G selects a phrase β as a candidate for "pied-piping."
c. β is merged to a category K.

P and G' are uninterpretable. P deletes if G is active (Suicidal Greed). G' also deletes, but it cannot delete in step (54a) before carrying out its function in step (54b). There may be reasons to suppose that G cannot delete before step (54c), but I will defer the matter.

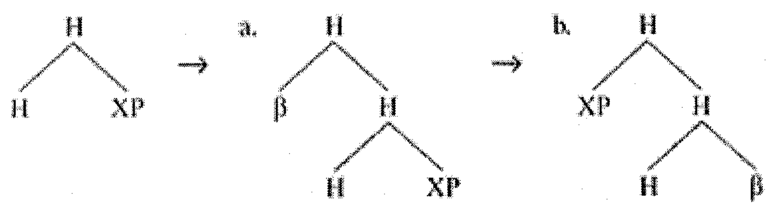
The new object K formed by Merge of β to α retains the label L of α, which projects. There are two reasonable possibilities, illustrating the ambiguity of cyclicity mentioned earlier:

(55) a. α is unchanged.
b. β is as close to L as possible.

Suppose we have the LI H with selectional feature F, and XP satisfying F. Then first Merge yields α = {XP, H}, with label H. Suppose we proceed to second Merge, merging β to α. In this case β is either extracted from XP (Move) or is a distinct syntactic object (pure Merge). There are two possible outcomes, depending on choice of K in (55).

(56) a. {β, {XP, H}} (as in (55a))
b. {XP, {β, H}} (as in (55b))

In tree notation the equivalents are (57a, b), respectively.

(57)


The desired outcome is (57a), not (57b); that has always been assumed without discussion. Thus, if H = T, XP = VP, and β is either an expletive merged to TP or a DP raised from XP, the result of Merge would be (58a), with β the Spec of T and VP remaining its complement, not (58b), with β becoming the complement of T and VP its Spec.

(58) a. $[_T\ \beta\ [T\ VP]]$
b. $[_T\ VP\ [T\ \beta]]$

But the reasons are not entirely obvious. Each outcome satisfies a reasonable condition: (58a) satisfies the familiar *Extension Condition* (55a); (58b) satisfies the condition of *Local Merge* (55b).

One possibility is to stipulate that the Extension Condition always holds: operations preserve existing structure. Weaker assumptions suffice to bar (57b) but still allow Local Merge under other conditions. Suppose that operations do not tamper with the basic relations involving the label that projects: the relations provided by Merge and composition, the relevant ones here being sisterhood and c-command. **105** Derivations then observe the condition (59), a kind of economy condition, where R is a relevant basic relation.

(59) Given a choice of operations applying to α and projecting its label L, select one that preserves R(L, γ).

In the case of (56)–(57) the principle (59) selects (a) over (b). Basic relations of H are preserved in (a), but not in (b), which modifies sisterhood and c-command for H (in fact, the category α = {XP, H} disappears altogether).

One case permitted under (59) but barred by strict adherence to the Extension Condition is head adjunction. The standard assumption is that in this case Local Merge takes precedence over the Extension Condition. Thus, in (56)–(57), (b) is chosen over (a) for β a head H'. Adherence to the Extension Condition would change the status of H' to an XP; V-raising to T, for example, would create a VP-TP structure, with VP the Spec of T, contrary to intention. The uniformity condition for chains (*MP*, sec. 4.1, (17)) bars that choice, leaving as the only option head adjunction yielding (b), in violation of the Extension Condition. (59) is satisfied vacuously. [106] Head adjunction thus provides some reason to weaken the Extension Condition as proposed, permitting Local Merge if (59) is satisfied.

For XP-merger, (59) eliminates the ambiguity of choice of K for second Merge, imposing the Extension Condition (55a) rather than Local Merge (55b). But the choice remains open for third Merge. Suppose we have the outcome (a) of (56)–(57), as required, and now merge γ to it (either by pure Merge, or with γ extracted from XP). The two possible outcomes are (60a) (satisfying the Extension Condition) or (60b) (satisfying Local Merge).

(60) a. {γ, {β, {XP, H}}}
b. {β, {γ, {XP, H}}}

In (60a) the newly merged phrase γ is the outer Spec; in (60b) it is the inner Spec. Either way, sisterhood and c-command are preserved for H, satisfying condition (59). The discussion so far has kept to the Extension Condition (55a) (e.g., (8b)). The question becomes more intricate when we cast a wider net, another issue deferred here.[107]

Both Local Merge and the Extension Condition yield cyclicity and hence satisfy the optimality condition (52a), but in slightly different ways. The distinction brings out an empirical difference between the derivational definition of c-command and the definition given earlier in terms of composition of elementary relations (see note 45; also Collins 1997, 84). Combined with Kayne's Linear Correspondence Axiom (Kayne 1994), as in the work cited, the derivational approach entails that the Extension Condition is inviolable. The compositional approach still allows the option of Local Merge, as in head movement and Merge in inner Spec.

The steps toward minimalist goals discussed above improve the *MP* theory in other respects. It was argued there that Merge is preferred over Move, but on dubious grounds. Now the preference is immediate: Agree and Merge are each components of Move, so it is a simple matter of more versus less. In fact, because of the extra component (b) of (54), Move is more complex than combined application of Agree and Merge. Thus, if a derivation D has reached the stage (61) and an expletive is available, D must proceed to (62a) (Merge combined with Agree, the latter to establish the T-associate relation), not (62b) (Move, then backtracking to the alternative (62a) when the derivation crashes with an unused expletive).

(61) [$_{TP}$ T be [a proof discovered]]

(62) a. there was a proof discovered
b. a proof was discovered

Again, a look-ahead property disappears. This sharpens the analysis of (10a)/(12a): the desired outcome is not just an option, but in fact the only option at the stage (61) of the cycle.

Suppose we have reached the stage (61) and the only unused element of the lexical subarray is the complementizer C. Since Merge is preferred to Move, the *MP* theory wrongly predicts merger of C and TP, barring dislocation of *a proof* to the subject position to yield (62b) (Eduardo Raposo, personal communication). The problem is overcome when we recognize the asymmetry of Merge. T contains no selector that allows C to merge to TP. With selectional features of T satisfied in the strict cyclic derivation of full TP (see (53)), the derivation can move on to C, which selects TP, projecting C.

In *MP* it is speculated that categories lacking interpretable features should be disallowed—specifically Agr, consisting only of uninterpretable ϕ-features (sec. 4.10). That conclusion is forced in this version. **108** Suppose α is an LI that consists of uninterpretable features only and selects β, yielding the syntactic object K = {α, β} with label α. In the course of a convergent derivation, α will disappear, leaving K and higher projections of α without a label. But terms without labels are not well-formed syntactic objects. Accordingly such elements as Agr not only *might* not exist, but *cannot* exist, on rather plausible assumptions. The argument carries over to other cases, among them semantically null determiners D_{null}. If true D relates to referentiality/specificity in some sense, then an indefinite nonspecific nominal phrase (*a lot of people, someone* that enters into scopal interactions, etc.) must be a pure NP, not DP with D_{null} (and the EPP cannot be stated as a D-feature). Notice that the argument holds only for heads α that select β. If α enters a derivation by Merge to a selector, deletion of α leaves a legitimate object (thus expletives might, in principle, delete completely).

I have explored a number of lines of argument in the attempt to refine and improve the Minimalist Program, including (a) restriction of basic operations to Merge and Agree (the latter reducing to deletion of uninterpreted matched features), each satisfying conditions of "good design," each preempting the more complex operation Move; and (b) restriction of access to the features **F** provided by UG in successive steps ((3) and its extension to cyclic derivation based on lexical subarrays). Basic properties of CFCs are then accommodated. These and other cases discussed reduce operative complexity in a natural way, reinforcing the suspicion that there is something to the curious principles in (28) and the intuitions and architectural conceptions on which they are based. Basic relations are restricted to those provided by Merge and composition. A number of other ways of overcoming errors and defects of *MP* and advancing the project of *MP* and related work have been outlined. I have raised, but surely not solved, the problem of whether the most striking apparent imperfections of narrow syntax, violation of the Interpretability Condition and dislocation, are true imperfections or are reasonable ways of satisfying design conditions, perhaps with uninterpretable features serving as the mechanism to induce structural properties required by interpretive systems at the interface. The discussion has been largely theoretical, with only a few indications about how things turn out on "best possible" expectations. Many questions have been left dangling, innumerable others unmentioned. I hope to return to some of these topics in the continuation.

Notes

1. As discussed in the introduction of *MP*, the chapters are largely based on lecture-seminars at MIT, the last of them in fall 1994. What follows draws from discussions during the lecture-seminars of fall 1995, 1997. I will make no attempt to review the impressive range of recent work that bears directly on questions that

arise, a failure that leaves no slight lacunae, as does the failure to consider alternatives that have been developed (see, among others, Abraham et al. 1996, Brody 1995, Collins 1997, Epstein et al. 1998, Frampton and Guttmann 1998, Sportiche 1995, Zwart 1996). For comments on an earlier draft, many of them incorporated here, I am particularly indebted to Zeljko Boskovic, * Chris Collins, Sam Epstein, Howard Lasnik, and Juan Uriagereka.

2. See, for example, Hermer and Spelke 1996. More generally, I assume that mental capacities are "modular" in the sense of Chomsky 1975, with "learning theories" LT(O, D) that may vary for organism O and cognitive domain D. The resulting modules might then have input/output properties of the kind analyzed in Fodor 1983, while belonging to a "central" system more structured than Fodor assumes.

3. For some illustrations and discussion, see Jenkins 1997, 1999, Marcus 1998.

4. Which implies virtually nothing about the novelty of its component elements.

5. I assume here familiar idealizations, abstracting away from real-world interactions that yield complex and widely varying forms of multiple systems. The term *idealization* sometimes misleads; the process is a crucial part of the effort to determine reality.

6. As systems, that is; their components need not be. See note 4.

7. Processing systems vary with languages and language types, even for very young infants, enabling them to sort out distinct languages in the data to which they are exposed. See Bosch and Sebastián-Gallés 1997. Jusczyk 1997, Mehler and Dupoux 1994. Whether these important discoveries (which add new dimensions to "poverty of the stimulus" arguments) are consistent with the simplifying assumption depends on how sensorimotor processing is "modulated" by the target language. Whorfian ideas on the meaning side have a similar flavor. See Phillips 1996 for an intriguing approach that bears on some of these questions.

8. I am assuming here the basic framework of Chomsky 1955, though of course there have been radical changes since. Levels are systems of representations: representations formed in the course of derivation typically do not form part of a level. Note further that the term *representation* is a technical one, with no "representation" relation in the sense of representational theories of ideas, for example.

9. On my own views regarding these issues, see Chomsky 1975, 1995a, 1996, among others.

10. Some disagree, regarding the issues as problematic. See, for example. Carr 1997, and for comment on some related matters, George 1996. On features as "instructions" for vocal gestures, see Halle 1983.

11. L an I-language in the technical sense, here and below. One simplifying assumption is that L is literally deducible from a choice of parameter values and lexicon, so that acquisition is "as if instantaneous." That need not be the case (e.g., in the theory of acquisition proposed in Locke 1997). It therefore becomes interesting to ask how close to true the assumption is. None of this has anything to do with the existence of a "language acquisition device" (LAD). LAD is just S_0,

under a particular construal, including whatever properties of S_0 may manifest themselves in the course of development. Postulation of LAD is often described as questionable or wrong, but that can hardly be so, at least if language is an identifiable component of human cognitive structure in any respect.

12. The question of a perfect language, whether designed by God or humans, is of course an old one, but completely distinct. Note further that the question of optimal design has nothing to do with the issue of "best theory" for FL (however intricate and "imperfect" the design of the system).

13. It is a misunderstanding to contrast "minimalism and X," where X is some theoretical conception (Optimality Theory, Lexicalism, etc.). X may be pursued with minimalist goals, or not.

14. See Belletti 1990 and much subsequent work.

15. Complications can readily be added. Little is known about evolution of higher mental faculties, and it is not clear how much can be learned within the limits of contemporary understanding. For a skeptical appraisal, see Lewontin 1990, 1998; and for critical analysis of recent efforts, see Berwick 1997, Jenkins 1997, 1999, Orr 1997.

16. These are called "bare output conditions" in *MP*, "output" because they are conditions on interface levels (hence "outputs" on a derivational approach), "bare" to distinguish them from filters, ranked constraints, and other devices that are part of the computational system itself.

17. For significance, we might assume further that there is no (nonarbitrary) bound on the number of legible expressions. Note that FL satisfying this minimal condition might—and the real system in fact does—permit generation of expressions that are unusable (structure of memory, garden path, etc.).

18. Interpretability is not to be confused with intelligibility. A convergent expression may be complete gibberish, or unusable by performance systems for various reasons. See note 17. And performance systems typically assign interpretation to nonconvergent expressions.

19. Convergence is defined in terms of properties of the external systems; the concept is clear insofar as these properties are clear. Many questions arise about its role in interpretation of deviance and in economy conditions—specifically, does crash "free up" alternative derivations, as assumed in *MP* but not as these notions are elaborated elsewhere (e.g., Collins 1997, and in what follows)?

20. Also, well beyond the minimal sound-meaning connection given by initial assembly of features in the lexicon, an inescapable step in fixing a language for familiar reasons.

21. I put aside a variant that restricts "linguistic evidence" to identification of "well-formed" ("grammatical") expressions, so that the linguist then faces the alleged problem of selecting among grammars that are extensionally equivalent over these objects. Such demands inherit (and, by the radical restriction of evidence, amplify) the incoherence of the other approaches, adding the further difficulty of deciding what this property might be, for natural language.

22. There are many such debates, often with an oddly one-sided character: criticism of a largely unspecified position, with no defense of it on the part of those who are alleged to hold it but who in fact do not see what the issue is. Examples include the "innateness hypothesis," "autonomy of syntax," and "formalist" approaches. For an effort to find some significance in the "functionalist-formalist controversy," see Lasnik 1999. For similar attempts with regard to the "autonomy" thesis, see Chomsky 1977. Critics of the "innateness hypothesis" may have in mind issues of modularity and species-specificity, though that is unclear, since proposals with any substance are highly modular and (so far as is known) species-specific. See Jenkins 1997, 1999, Marcus 1998.

23. On conceptual and empirical arguments, with varying conclusions, see, among other works, those mentioned in note 1, *MP*, and Chomsky 1998.

24. Any interpretation of L is computational in some sense, raising difficult and obscure questions about what this means for a cognitive system. These are not to be confused with problems of processing (parsing, production).

25. Among many other works, see Chomsky 1986a on vacuous movement, Fox 1995, 1998 and Reinhart 1993 on the LF counterpart, and Collins 1997 on local determination.

26. For example, integrated action/perception models motivated by computational savings over construction of the detailed properties of a presented scene. For a review, see Clark 1998.

27. The properties of features and assembly form a large part of the subject matter of traditional and modern linguistics; I will put these topics aside here, including questions about organization of assembled features within a lexical item LI. Also left to the side is the question whether LI is assembled in a single operation or at several stages of the derivation, as in Distributed Morphology (Halle and Marantz 1993). Rephrasing of the account just given in these terms is straight-forward. Recall that L is a *state* of FL; state changes, of course, may modify the lexicon.

28. Or, if we distinguish independent selections of a single lexical item, a numeration Num (as in *MP*), an extension I will put aside until it becomes relevant.

29. It would not suffice to say that constant memory can be accessed throughout the derivation. The lexicon is a distinct component of memory; for C_{HL}, our beliefs about the stars don't matter, but the lexical properties of *star* do. However hard it may be to make the distinction properly, there is good reason to believe that it is real.

30. Much current work allows Move to raise α to [Spec, V] even when Merge or Agree would be options; see Lasnik 1995c and sources cited there. An interesting question is whether the evidence for this conclusion can be incorporated within the more restricted framework envisaged here under a proper interpretation of the qualification "when possible."

31. I am putting aside many questions concerning the substantive/functional distinction, adopting it only for heuristic purposes. Also omitted is the D head of DP (which seems to belong to a different system), more complex verbal construc-

tions, and the question whether nontransitive ones have a light verb head. Some might—for example, *seem*, which c-commands the experiencer Spec even in nonraising languages like English, either because it selects a light verb with this Spec or because it raises from that position to a light verb. Ignored as well are the "peripheral" systems outside TP; I will use C and T as surrogates for richer systems. On these matters see Rizzi 1997, Cinque 1999, and many other studies on the CFC systems and others. The concepts "inherent/structural Case" are understood as in Chomsky 1981, 1986b: θ-related versus structurally determined.

32. In the sense of Pesetsky (1982), modifying ideas of Grimshaw (1979).

33. Strengthening *may* to *must* stipulates part of Burzio's Generalization; the rest should follow from Case/agreement theory.

34. Irrelevant to our concerns here, for C (5a) might be parametrically contingent on other operations (e.g., partial *wh*-raising as in German). Other possible parameterization is put aside here.

35. In effect, Hale and Keyser's (1993) theory adapted in *MP* (sec. 4.6). One consequence of this conception of θ-theory is that the θ-Criterion cannot be satisfied by raising an argument to a θ-position or by raising of "θ-features" (the existence of such features aside, I will suggest below that feature movement may not be possible). Other conceptions reject these conclusions (see, e.g., Hornstein 1999, Manzini and Roussou (to appear)).

36. The examples in the text are translations of Icelandic examples given by Jonas (1996). It had been assumed that Subj is higher than Obj in these constructions, but Jonas found that the conclusions relied on improper choice of left-edge markers. This eliminates complications in *MP* about the issue.

37. The counterpart of the *to* phrase in *seems to-me [NP to . . .]*. This phrase is sometimes described as the complement of V, but I assume it to be a Spec in a Larsonian shell. See note 31.

38. On (15a), see Sigurðsson 1996. On (15b), see *MP*, sec. 4.9, (168). The structure of subjects in multiple-subject constructions raises a wide range of questions put aside here.

39. See *MP* and Martin 1996. Some of these distinctions have been attributed to trace government, but that mechanism is not available here.

40. Observe that (9), (10), and (12c) appear problematic because of raising of *a proof* from DO of *discover*. There are reasons to suppose that the actual structure at this stage is the expected *there to be [discovered a proof]* as in similar languages, the preferred English construction *a proof [discovered t]* being formed outside the system we are now considering. I will assume so, leaving the issue to the side here.

41. Why not dispense with LA, just selecting subarrays cyclically? Apart from the general considerations about access reduction already discussed, there is a more specific reason: chain properties can be reduced in significant part to identity if lexical arrays are enriched to numerations. To achieve the same result with cyclic choice of successive subarrays requires continual access to the full lexicon and memory of how many times each item has been selected.

42. An island effect, if it exists at all, is very weak with such structures.

43. Early work sought to establish the categories of A- and $\bar{A}$-movement ("Move NP," "Move *Wh*"), later head movement, while parallel inquiries sought commonalities. Important outcomes were Rizzi's (1990) theory of Relativized Minimality and Lasnik and Saito's (1992) Move α theory. The distinctions mentioned here crosscut these categories.

44. "Stylistic" operations might fall within the phonological component (see *MP*, sec. 4.7.3, Kidwai 1996). Operations lacking overt counterparts and apparently not interacting with C_{HL} might be among the principles of interpretation of LF, hence "postcyclic," inspecting a representational level in the manner of many other systems (including binding theory, on the assumptions of *MP*). If so, much of the very enlightening recent work on ellipsis and antecedent-contained deletion (along with event structure and other topics) could be understood as an exploration of the language-external systems at the border of the language faculty, roughly analogous to acoustic and articulatory phonetics on the sound side.

45. The categories might overlap, but unproblematically it seems. System design should preclude unwanted cases of improper movement. That seems attainable, but must be demonstrated. I will continue to restrict attention to raising of XP.

46. It also suggests a new approach to some Empty Category Principle (ECP) issues, such as subject extraction (Idan Landau, personal communication).

47. To clarify this and related conclusions and establish them in full generality requires a far more comprehensive review and analysis than is undertaken below. Similar qualifications hold throughout.

48. See Fox 1998 and Nissenbaum 1998. If adjunction is restricted as suggested in *MP* (sec.4.7.3), then movement to the edge will be to a Spec position for vP as well as CP. Phases might also be the target for QR, if this noncyclic operation targets C', merging the raised quantifier phrase between C and [Spec, C]; see note 107.

49. Convergence is not guaranteed, of course (it can fail in many ways)—only permitted without look-ahead, the desideratum we are exploring. Conditions could be added to restrict crash, but they are redundant, simply restating properties of convergence, unless motivated in some other way. Questions arise about operations that appear to violate Subjacency (see references of note 92, among others). Note the restriction to uninterpretable features *in the domain*. Legitimacy of those at the edge (specifically, EA) will be determined at the next higher phase, a matter that opens interesting questions, put aside here.

50. Parametrically varying properties of H enter into the application of (24), which might be extended to head movement (see note 93 and text). I will call the EPP-feature a *P-feature* (*periphery feature*) if H does have an appropriate EPP-feature by virtue of its inherent properties (e.g., the Case/agreement properties of v, the Q-feature of interrogative C). The device is introduced to extend the general theory of movement beyond A-movement, but should raise warning flags.

51. In *MP* an unsatisfactory argument was required to bar expletives from merging in OS constructions. The analogous problem arose for CP, but was ignored.

52. For some speculations along similar lines, see *MP*, pp. 294 (condition (76)) and 377.

53. On some assumptions, though not here, IFM passing through [Spec, v_2] is improper movement.

54. Among other questions, what is the status of small clauses, or relative clauses and other adjuncts? Possibly the latter are derived "in parallel," in the manner of multidimensional analyses of coordination or parentheticals, with their own LAs and with the ultimate status of the adjunct being determined in the larger structure in which it is inserted (as for other multidimensional structures).

55. See note 31. Many questions about the internal structure of the words are put aside.

56. See, for example, Collins 1997, Johnson and Lappin 1997, Yang 1997.

57. The reference to "fundamental principles" in some (obscure) sense is crucial. Otherwise, for properly selected categories of expressions (which may well exhaust the possibilities), look-ahead properties of computation at stage α might be overcome by resort to whatever aspects of UG determine that the wrong choice eventually crashes.

58. The issue is whether crash "frees up" derivational paths not selected earlier, as assumed in *MP* (incorrectly, I assume here). Computations can crash in endless ways, raising no complexity issue. See note 49.

59. See Chomsky 1998, Collins 1997, Frampton and Gutmann 1999.

60. These are conceptual "good design" conditions, but fairly trivial ones, based on the assumption that less machinery is better than more.

61. The significance of (ii) was brought out in Epstein's (1999) derivational analysis of c-command, which underlies much important work since, including Epstein et al. 1998.

62. That the copy theory is the simplest version is clear in the earliest formulation, in Chomsky 1955. The more complex operation Merge-Delete was adopted there on the assumption that T-markers are mapped to PF. When that picture was modified under the Extended Standard Theory, Delete was abandoned in favor of trace theory, trace being a new kind of element. That was an error (mine), and the copy theory, which restored the simplest case of the original approach, was mistakenly regarded as a further innovation. The divergence of history and logic has caused much confusion.

63. We could, for example, identify this as K' = K with the occurrence of α in question replaced by some designated element Occ distinct from anything in K. In Chomsky 1955 *occurrence* is defined in terms of linear order, adapting a device from Quine 1940; but that mechanism is not available here.

64. Suppose, for example, that the operation Move could raise the object of V to become the object of V', forming [V' Obj]. Then if V = V' (by virtue of V-raising), the sisters of Obj would be identical.

65. See *MP*, sec. 3.5, and for improved versions with many consequences, Fox 1998 and Sauerland 1998b. A question that might be raised is where the "reconstruction" operation takes place: within narrow syntax or on the other side of the interface, along with binding theory (I am assuming, following *MP*) and other interpretive systems. Similar questions arise about other covert operations; see note 44.

66. Attempts in *MP* to account for successive-cyclic movement in terms of linked chains and other devices can be eliminated: they were based on failure to take seriously enough the actual notion of chain. They were also in error in not recognizing the role of intermediate traces in interpretation and computation. The conclusions drawn there about trace invisibility no longer hold, though interesting aspects of the question remain.

67. The compositional definition of c-command is suggested by Epstein's (1999) derivational approach (see note 61): α c-commands β if α is merged with K containing β. There are empirical differences between the two approaches, to which I will return. The derivational definition also raises some questions: in particular, why does "containment" enter (i.e., why does X merged with Y c-command terms of Y)? The matter is addressed in Epstein et al. 1998, chap. 6, but inconclusively, as far as I can see. An argument for asymmetry of c-command is also presented there; I am assuming that there is no asymmetry, its effects being derived in other ways.

68. Sisterhood relations would be carried over under head adjunction if the result were taken to be an extended MLI. Many questions dissolve if head adjunction is part of the phonological component. There are, I think, independent reasons to suspect that this may be true, at least over an interesting range, but I will defer the question.

69. See Kitahara 1999, Lasnik 1995a,b, Uriagereka 1999 a.

70. There are important distinctions among these features, which I will largely ignore here. Problems of interpretation are also not trivial, as in (i)–(v).

(i) animal languages (is, are) their main research interest

(ii) three books (is, are) too much to read in a week

(iii) we expected animal languages to be their main research interest

(iv) we found three books too much to read in a week

(v) animal languages raise(s) serious issues/seem(s) to be their main research interest

In (i) and (ii) the difference in meaning appears superficially to lie in the verbal inflection, but it carries over to forms lacking that inflection, as in (iii) and (iv). And though agreement properties of the copula (with surface subject or complement) are a factor, the matter is more convoluted ((iv) and (v)). See Reid 1991 for discussion from a "functionalist" perspective.

71. A possible case is control contingent on long-distance agreement. See *MP*, sec. 4.4.5, (40)–(43), reviewing work of Anna Cardinaletti and Michal Starke. It

was assumed there that both binding and control are subject to these long-distance effects, but closer examination of Italian by Carlo Cecchetto (personal communication) indicates that the effects are limited to control alone (see also Cardinaletti 1997, Den Dikken 1995). That would explain the failure of binding in such examples as *there seem to each other [to be many men in the room]* and fits with other evidence that binding requires an "overt" (possibly PRO, trace, or *pro*) antecedent, not just an implicit argument; see Rizzi 1986. There are additional complications. The phenomena seem less clear for passives than unaccusatives (possibly because of interference from an implicit subject) and become blurred or disappear in "long-distance" cases; see Hornstein 1996, Lasnik 1997.

72. Contrary to *MP*. The extension of these notions to features raises difficulties, not insuperable but better avoided, as seems possible.

73. I continue to disregard (35d), abstracting to the expected form (35b); see note 40.

74. They have also shifted as inquiry has proceeded: thus, from the 1950s through the mid 1970s such matters as quantifier scope were commonly taken to be prototypical examples of "surface" interpretation, whereas more recent work generally takes them to be prototypical properties at LF.

75. The approach renders superfluous the intuitive motivation for pied-piping proposed in *MP*, in terms of PF crash. For argument that some such device nevertheless operates in languages with rich morphology, with consequences for null subject, barriers, and Spell-Out, see Uriagereka 1999 a. For an approach somewhat similar to what is outlined below, from a partially different perspective, see Frampton and Gutmann 1999.

76. If a reflex of an interpretable ϕ-set, it would be erased in situ by the ϕ-set of *candidate* itself.

77. See *MP*, Ura 1996. The condition stated there refers only to Specs of the same head. Whether the generalization (41) is appropriate depends on answers to questions about the structure of more complex constructions (double object verbs, etc.). These and other questions, including parameterization, are put aside here. See Boskovic * 1997 and McGinnis 1998, among others. The notion "feature occurrence" used implicitly here for expository convenience can be eliminated by restatement in terms of the heads to which features belong.

78. Among the problems that arise is the status of scrambling. The logic would suggest that for at least some cases, a scrambling feature induces pied-piping even after Case assignment, the pied-piped element being "attracted" by a higher probe, whereas other cases fall into a category distinct from feature-driven movement. For exploration of alternatives in a comparative study, see Sauerland 1998a.

79. For a similar configuration in phonology, with > = linear order, see the discussion in Halle 1995 of coronal assimilation in Sanskrit, barred by an intervening (nonassimilating) coronal.

80. On the assumptions made in *MP*, sec. 4.10, multiple-subject (including transitive expletive) options are parameterized in terms of deletion of EPP-features. See particularly Ura 1996, and for skepticism about the option, Zwart 1997. Agreement in the sense discussed here is to be distinguished from concord, with different properties.

81. The analysis of structural Case is along the lines proposed by George and Kornfilt (1981). As they observe, structural Case linked to ϕ-features may be dissociated from finiteness. Matters become more complex when we consider ergative/absolutive and mixed systems, and languages in which ϕ-features without finiteness do not suffice for nominative Case assignment (see Iatridou 1993).

82. For a different approach, on the assumption that *there* has structural Case, see Lasnik 1995a. For a different perspective on a wide range of related issues, see Moro 1997.

83. Similar conclusions are supported by locative and quotative inversion (see Collins 1997), though with restrictions and complications, and similarities to other poorly understood constructions (e.g., *still unclear remain (are, seem to be) the answers to those questions*).

84. Among these, the ECP, Subjacency, and other conditions. But many problems remain. See Kitahara 1999, Lasnik 1995a,b, Uriagereka 1999 a.

85. Perhaps substantive lexical categories do not exist, only bare roots. Configurational and morphological properties, along with interpretable noncategorial features of the root, would then determine relevant structural properties, as in Semitic. The possibility is suggested by work since the 1960s on derivational versus transformationally induced morphology. See Marantz 1997 for pertinent discussion.

86. A different motivation, based on the need to overcome ambiguity in the composite verbal element at LF, is developed by Uriagereka (1996). One can also think of various functional arguments: the familiar trade-off between order rigidity and richness of inflection, facilitation of search for attracted elements, and so on.

87. Pure inherent Case I take to be a distinct phenomenon, "invisible" to matching, as if inherent Case inactivates the ϕ-set. See note 31, and for more on the topic, McGinnis 1998 and sources cited there.

88. Reformulation is needed if [3person] is lacking, as has been suggested. Sigurðsson concludes that third person nominative Case requires agreement in the monoclausal construction [Dat V-T [$_{VP}$ *t* Nom]] and allows it as an option, alternating with default T, in the biclausal construction [Dat V-T *t* [$_{TP}$ Nom . . .]] (*t* the trace of Dat; irrelevant properties omitted). These facts could be expressed in terms of optionality of [person] reduction in matrix T, leaving open the reasons.

89. If Merge could induce agreement, vP-internal subject would manifest object agreement and have accusative Case, which would delete, preventing raising to [Spec, T].

90. Questions remain about French-style *il*, with a definiteness effect and other properties. Suppose the [person] feature of T can delete or shift to default with *there*-type Expl, in the manner suggested for quirky Case. That might accommodate "list readings" (as in *there (is/*am, remains/*remain) only me, there (are, remain) only us (John and Bill)*, in response, say, to *who's still here to do the work*), with no person agreement or nominative Case assignment, in the absence of the full complement of ϕ-features. Also to be accounted for is the fact that in ν phases the external argument pronoun does not agree with the ν head, possibly indicating that argument pronouns have true D-N structure, unlike expletives. See Cardinaletti and Starke 1994 and Uriagereka 1988 for some relevant considerations. Needless to say, these remarks barely touch on a rich array of questions.

91. To complete the analogy, C (and ν with its ϕ-set deleted) may have a non-specific P-feature analogous to [person] for T_{def}, perhaps contingent on assignment of the EPP-feature to a phase; see discussion of (24), (25).

92. Following ideas of Watanabe (1992) and Hagstrom (1998). This is not to be confused with the distinction between pied-piping of a full XP and pied-piping of a minimal operator (e.g., interrogatives/raising relatives vs. comparatives/complex adjectivals/nonraising relatives), a choice fixed by context (with various questions about relatives; see Sauerland 1998b). The *wh*-island analysis extends to other constructions if the feature that drives movement shares properties with [wh] (assuming here a hierarchy of features); see Hagstrom 1998 for supporting evidence. Multiple overt *wh*-movement as in some Slavic languages might be analogous to multiple head options for A-movement along with a principle to overcome the Subjacency effect. See Richards 1997, adapting ideas in Brody 1995, and for a general critical review, Boskovic * 1998. As is well known, pied-piping in $\bar{A}$-movement differs from the A-chain analogue, with variation among languages and constructions that is poorly understood.

93. The perennial troublemaker (i) falls into place if the (undeleted) [person] feature of embedded *there* bars association of matrix T to *three men*.

(i) *there seem there to be three men in the room

Groat (1997) points out further complications. Thus, whatever its status, (ii) is more acceptable.

(ii) there look as though there are three men (vs. *[a man]) in the room

That suggests that the [person] barrier may be overridden in some manner.

94. The *MP* analysis of superraising assumed that *it* raises to matrix T by locality and the derivation crashes because *friends* cannot be Case-checked. But covert attraction of the ϕ-features of *friends* by matrix T should still be possible, with convergence (Eduardo Raposo, personal communication). See Raposo and Uriagereka 1996 for a different approach.

95. Furthermore, I am speculating that XP can only appear if forced (barring Expl-raising always, and argument raising except to implement IFM or LF interpretive effects); see discussion of (24), (25). I overlook questions of possible parameterization for OVS languages; see Ura 1996.

96. Translations of Icelandic examples: (51a), with a default variant, from Sigurðsson 1996; (51b,c) from Boeckx (to appear); these and other examples, and discussion, in Schütze 1997. See also (15), (45biii), and note 88.

97. Unwanted chains could be formed without further constraints: for example, given α and β, and γ in α, γ could be merged to β (by Move), which is then merged to α, forming an unwanted $\langle\gamma, \gamma\rangle$ chain violating certain conditions. Motivated by Pollock's (1989) split-inflection ideas, in chapter 2 of *MP* I adopted Attract α in addition to Move α. In section 4.5.6 I sought to eliminate Move α as an independent operation, taking it to be conditional on Attract, revised as feature attraction. Considerations reviewed here seem to me to support this general point of view (now much simplified).

98. The problem is noted in *MP* (p. 385, n. 50) but left unresolved.

99. The conception is similar to the strict derivational interpretation of Spell-Out proposed by Epstein et al. (1998); for similar suggestions, on different grounds, see Uriagereka 1996, 1999 b. The basic architecture resembles that of Bresnan (1971); her results on the interaction of phonological and transformational rules fall into place more directly than in the Extended Standard Theory model. See several papers in Abraham et al. 1996 and Yang 1997 for related discussion.

100. The issue is narrow and might be avoidable in other ways. It would be restricted to subjects in OS constructions if the "reconstruction" operation induced by the final step of $\bar{A}$-movement preempts deletion, so that strict identity holds throughout $\bar{A}$-chains.

101. The discussion of categories and labels here largely follows Collins 1997. Questions about categories raised in note 64 are not relevant here; the syntactic object has some distinctive property that must be recognized.

102. See Kayne 1994 and *MP*. Kayne assimilates adjunct and specifier. *MP* suggests that XP-adjunction may not be part of narrow syntax (sec. 4.7.3)—possibly not a central range of cases of head adjunction either (see notes 32 and 46).

103. On the conventional and reasonable assumption that the selectional feature itself may be optional, as in the case of a verb that may or may not have IO (*send*, etc.).

104. An elaborate argument to guarantee projection of the target in *MP* is superfluous, under this reanalysis. The label is determined by the probe for the Agree component of Move, and by the selector for the Merge component. These are different features, but they yield the same choice of label. It seems a notational question whether we take the label to be the original LI or a reduced MLI.

105. New containment relations are defined whichever choice we make, in incommensurable ways; identity is irrelevant.

106. Whether sisterhood (hence c-command) is preserved depends on how (or if) the notion is defined for head adjunction. See note 46.

107. See Richards 1997 for evidence in support of Local Merge in the important case of multiple Move. See also note 33; the proposal there falls under Local Merge, with the Extension Condition obviated vacuously for "postcyclic" QR.

108. In *MP* it could be avoided only by recourse to the (dubious) distinction between deletion and erasure. Nothing follows about functional categories with semantic properties, as in Rizzi 1997, Cinque 1999.

References

Abraham, Werner, Samuel David Epstein, Höskuldur Thráinsson, and C. JanWouter Zwart, eds. 1996. *Minimal ideas*. Amsterdam: John Benjamins.

Belletti, Adriana. 1990. *Generalized verb movement*. Turin: Rosenberg and Sellier.

Berwick, Robert. 1997. Walking the walk, talking the talk. *Los Angeles Times Book Review*, Sept. 7.

Boeckx, Cédric. To appear. Quirky Agreement. *Studia Linguistica*.

Bosch, Laura, and Nuria Sebastián-Gallés 1997. Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition* 65. 33–69.

Chomsky, Noam. 1995a. Language and nature. *Mind* 104, 1–61.

Chomsky, Noam. 1995b. *The Minimalist Program*. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1996. Internalist Explorations. Ms., MIT, Cambridge, Mass.

Chomsky, Noam. 1998. Some observations on economy in generative grammar. In Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis, and David Pesetsky, eds., *Is the best good enough?* Cambridge, Mass.: MIT Press.

Cinque, Guglielmo. 1999. *Adverbs and functional heads*. Oxford: Oxford University Press.

Clark, Andy. 1998. Where brain, body, and world collide. *Daedalus* 127.2 (special issue, "The Brain"), 257–280.

Collins, Chris. 1997. *Local economy*. Cambridge, Mass.: MIT Press.

Dikken, Marcel den. 1995. Binding, expletives, and levels. *Linguistic Inquiry* 26, 347–354.

Epstein, Samuel David. 1999. The derivation of syntactic relations. In Samuel David Epstein and Norbert Hornstein, eds., *Working Minimalism*. Cambridge, Mass.: MIT Press.

Epstein, Samuel David, Erich Groat, Ruriko Kawashima, and Hisatsugu Kitahara. 1998. *A derivational approach to syntactic relations*. Oxford: Oxford University Press.

Fodor, Jerry. 1983. *The modularity of mind*. Cambridge, Mass.: MIT Press.

Fox, Danny. 1995. Economy and scope. *Natural Language Semantics* 3, 283–341.

Fox, Danny. 1998. Economy and semantic interpretation. Doctoral dissertation, MIT, Cambridge, Mass.

Frampton, John, and Sam Gutmann. 1999. Cyclic computation: A computationally efficient minimalist syntax. *Syntax* 2, 1–27.

George, Alexander. 1996. Katz astray. *Mind & Language* 11, 295–305.

George, Leland, and Jaklin Kornfilt. 1981. Finiteness and boundedness in Turkish. In Frank Heny, ed., *Binding and filtering*. Cambridge, Mass.: MIT Press.

Grimshaw, Jane. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10, 279–326.

Groat, Erich. 1997. A derivational program for syntactic theory. Doctoral dissertation, Harvard University, Cambridge, Mass.

Hagstrom, Paul. 1998. Decomposing questions. Doctoral dissertation, MIT, Cambridge, Mass.

Hale, Kenneth, and Samuel Jay Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In Kenneth Hale and Samuel Jay Keyser, eds., *The view from Building 20*. Cambridge, Mass.: MIT Press.

Halle, Morris. 1983. On distinctive features and their articulatory implementation. *Natural Language &*

Linguistic Theory 1, 91–105.

Halle, Morris. 1995. Feature geometry and feature spreading. *Linguistic Inquiry* 26, 1–46.

Halle, Morris, and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Kenneth Hale and Samuel Jay Keyser, eds., *The view from Building 20*. Cambridge, Mass.: MIT Press.

Hermer, Linda, and Elizabeth Spelke. 1996. Modularity and development: The case of spatial reorientation. *Cognition* 61, 195–232.

Hornstein, Norbert. 1996. Existentials, A-chains, and reconstruction. Ms., University of Maryland, College Park.

Hornstein, Norbert. 1999. Movement and Control. *Linguistic Inquiry* 30, 69–96.

Iatridou, Sabine. 1993. On nominative Case assignment and a few related things. In Colin Phillips, ed., *Papers on Case and agreement II*. (MIT Working Papers in Linguistics 19.) MITWPL, Department of Linguistics and Philosophy, MIT, Cambridge, Mass.

Jenkins, Lyle. 1997. Biolinguistics: Structure, development and evolution of language. *Web Journal of Formal, Computational and Cognitive Linguistics* 1(2). http://www.kcn.ru/tat_en/science/fccl/gp008.txt.

Jenkins, Lyle. 1999. *Biolinguistics*. Cambridge: Cambridge University Press.

Johnson, David, and Shalom Lappin. 1997. A critique of the Minimalist Program. *Linguistics and Philosophy* 20, 272–333.

Jonas, Dianne. 1996. Clause structure and verb syntax in Scandinavian and English. Doctoral dissertation, Harvard University, Cambridge, Mass.

Jusczyk, Peter. 1997. *The discovery of spoken language*. Cambridge, Mass.: MIT Press.

Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, Mass.: MIT Press.

Kidwai, Ayesha. 1996. Word order and focus positions in Universal Grammar. Ms., Jawaharlal Nehru University.

Kitahara, Hisatsugu. 1999. Eliminating * as a feature (of traces). In Samuel David Epstein and Norbert Hornstein, eds., *Working minimalism*. Cambridge, Mass.: MIT Press.

Lasnik, Howard. 1995a. Case and expletives revisited: On Greed and other human failings. *Linguistic Inquiry* 26, 615–633.

Lasnik, Howard. 1995b. Last Resort. In Shosuke Haraguchi and Michio Funaki, eds., *Minimalism and linguistic theory*. Tokyo: Hituzi Syobo.

Lasnik, Howard. 1995c. A note on pseudogapping. In Rob Pensalfini and Hiroyuki Ura, eds., *Papers on minimalist syntax*. (MIT Working Papers in Linguistics 27.) MITWPL, Department of Linguistics and Philosophy, MIT, Cambridge, Mass. [Reprinted, with minor revisions, in Lasnik, Howard. 1999. *Minimalist analysis*. Oxford: Blackwell.]

Lasnik, Howard. 1997. Levels of representation and the elements of anaphora. In Hans Bennis, Pierre Pica, and Johan Rooryck, eds., *Atomism and binding*. Dordrecht: Foris.

Lasnik, Howard. 1999. On the locality of movement: Formalist syntax position paper. In Michael

Darnell, Edith Moravcsik, Frederick Newmeyer, Michael

Noonan, and Kathleen Wheatley, eds., *Functionalism and formalism in linguistics*. Vol. 1, 33–54. Amsterdam: John Benjamins.

Lasnik, Howard, and Mamoru Saito. 1992. *Move α*. Cambridge, Mass.: MIT Press.

Lewontin, Richard. 1990. The evolution of cognition. In Daniel Osherson and Edward Smith, eds., *An invitation to cognitive science*, Vol. 3: *Thinking*. Cambridge, Mass.: MIT Press.

Lewontin, Richard. 1998. The evolution of cognition: Questions we will never answer. In Dan Scarborough and Saul Steinberg, eds., *An invitation to cognitive science*, Vol. 4: *Methods, models, and conceptual issues*. Cambridge, Mass.: MIT Press. [Update of Lewontin 1990.]

Locke, John. 1997. A theory of neurolinguistic development. *Brain and Language* 58, 265–326.

Manzini, Maria Rita, and Anna Roussou. To appear. A minimalist theory of A-movement and control. *Lingua*.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In Alexis Dimitriadis, Laura Siegel, Clarissa Surek-Clark, and Alexander Williams, eds., *Proceedings of the 21st Annual Penn Linguistics Colloquium*. (Pennsylvania Working Papers in Linguistics 4.2.) Penn Linguistics Club, University of Pennsylvania, Philadelphia.

Marcus, Gary. 1998. Can connectionism save constructivism? *Cognition* 66, 153–182.

Martin, Roger. 1996. A minimalist theory of PRO and control. Doctoral dissertation, University of Connecticut, Storrs.

McGinnis, Martha. 1998. Locality in A-movement. Doctoral dissertation, MIT, Cambridge, Mass.

Mehler, Jacques, and Emmanuel Dupoux. 1994. *What infants know*. Oxford: Blackwell.

Moro, Andrea. 1997. *The raising of predicates*. Cambridge: Cambridge University Press.

Nissenbaum, Jon. 1998. Movement and derived predicates: Evidence from parasitic gaps. In Uli Sauerland and Orin Percus, eds., *The interpretive tract*. (MIT Working Papers in Linguistics 25.) MITWPL, Department of Linguistics and Philosophy, MIT.

Orr, Allen. 1997. The softer side of sociobiology. *Boston Review*, Oct./Nov.

Pesetsky, David. 1982. Paths and categories. Doctoral dissertation, MIT, Cambridge, Mass.

Phillips, Colin. 1996. Order and structure. Doctoral dissertation, MIT, Cambridge, Mass.

Pollock, Jean-Yves. 1989. Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry* 20, 365–424.

Quine, Willard V.O. 1940. *Mathematical logic*. Cambridge, Mass.: Harvard University Press.

Raposo, Eduardo, and Juan Uriagereka. 1996. Indefinite *se. Natural Language & Linguistic Theory* 14, 749–810.

Reid, Wallace. 1991. *Verb and noun number in English*. London: Longman.

Reinhart, Tanya. 1993. *Wh*-in-situ in the framework of the Minimalist Program. Ms., TelAviv University. [Published in *Natural Language Semantics* 6, 3–28 (1997).]

Richards, Norvin. 1997. What moves where when in which language? Doctoral dissertation, MIT, Cambridge, Mass.

Rizzi, Luigi. 1986. Null objects in Italian and the theory of *pro. Linguistic Inquiry* 17, 501–557.

Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, Mass.: MIT Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman, ed., *Elements of grammar*. Dordrecht: Kluwer.

Sauerland, Uli. 1998a. Erasability and interpretation. Ms., MIT, Cambridge, Mass.

Sauerland, Uli. 1998b. On the making and meaning of chains. Doctoral dissertation, MIT, Cambridge, Mass.

Schütze, Carson. 1997. INFL in child and adult language: Agreement, Case and licensing. Doctoral dissertation, MIT, Cambridge, Mass.

Sigurðsson, Halldór Ármann. 1996. Icelandic finite verb agreement. *Working Papers in Scandinavian Syntax* 57, 1–46.

Sportiche, Dominique. 1995. Sketch of a reductionist approach to syntactic variation and dependencies. In Héctor Campos and Paula Kempchinsky, eds., *Evolution and revolution in linguistic theory.* Washington, D.C.: Georgetown University Press.

Ura, Hiroyuki. 1996. Multiple feature-checking: A theory of grammatical function splitting. Doctoral dissertation, MIT, Cambridge, Mass.

Uriagereka, Juan. 1988. On government. Doctoral dissertation, University of Connecticut, Storrs.

Uriagereka, Juan. 1996. Formal and substantive elegance in the Minimalist Program. In Chris Wilder, Hans-Martin Gärtner, and Manfred Bierwisch, eds., *The role of economy principles in linguistic theory.* Berlin: Akademie-Verlag.

Uriagereka, Juan. 1999 a. Minimal restrictions on Basque movement. *Natural Language & Linguistic Theory* 17, 403–444.

Uriagereka, Juan. Forthcoming b. Multiple Spell-Out. In Samuel David Epstein and Norbert Hornstein, eds., *Working minimalism*. Cambridge, Mass.: MIT Press.

Watanabe, Akira. 1992. Subjacency and S-Structure movement of *wh*-in-situ. *Journal of East Asian Linguistics* 1, 255–291.

Yang, Charles. 1997. Minimal computation: Derivation of syntactic structures. Master's thesis, Artificial Intelligence Laboratory, MIT, Cambridge, Mass.

Zwart, C. Jan-Wouter. 1996. *Morphosyntax of verb movement*. Dordrecht: Kluwer.

Zwart, C. Jan-Wouter. 1997. Transitive expletive constructions and the evidence supporting the multiple specifier hypothesis. In W. Abraham and E. van Gelderen, eds., *German-Syntax Problems. Problematic Syntax*. Tübingen: Narr.