

Derivation by Phase
Noam Chomsky

GLOSS: As I did before, I'll try to gloss Chomsky's recent paper, to the extent I understand it. Please do not circulate this commentary without permission, or Chomsky's paper in this format, since although he has given permission for the commentary, the paper is to appear in a festschrift for Ken Hale. In fact, the electronic manuscript I have is the one Chomsky sent, which is slightly less worked out than the version circulated in MITWPL (M-version). Also keep in mind that my commentary need not reflect Chomsky's own views, and I am responsible for mistakes in interpretation. Juan Uriagereka

Part One

[The original does not have parts, but I have divided it in two for presentation purposes.]

What follows extends and revises an earlier paper ("Minimalist Inquiries," MI), which outlines a framework for pursuit of the so-called "minimalist program," one of a number of alternatives that are currently being explored.¹ The shared goal is to formulate in a clear and useful way -- and to the extent possible to answer -- a fundamental question of the study of language, which until recently could hardly be considered seriously, and may still be premature: To what extent is the human faculty of language FL an optimal solution to minimal design specifications, conditions that must be satisfied for language to be usable at all?

This is a humble statement, to be kept in mind.

We may think of these specifications as "legibility conditions": For each language L (a state of FL), the expressions generated by L must be "legible" to systems that access these objects at the interface between FL and external systems -- external to FL, internal to the person.

The strongest minimalist thesis SMT would hold that language is an optimal solution to such conditions. SMT, or a weaker version, becomes an empirical thesis insofar as we are able to determine interface conditions and to clarify notions of "good design."

This should be considered a goal of this paper. Keep in mind also the very next sentence...

¹ Chomsky (1998). For background see references cited there and Lasnik (1999), among many others.

While SMT cannot be seriously entertained, there is by now reason to believe that in nontrivial respects some such thesis holds, a surprising conclusion insofar as it is true, with broad implications for the study of language, and well beyond.

Note the indefinite article: an optimal solution. "Good design" conditions are in part a matter of empirical discovery, though within general guidelines of an a prioristic character, a familiar feature of rational inquiry. In the early days of the modern scientific revolution, for example, there was much concern about the interplay of experiment and mathematical reasoning in determining the nature of the world. Even the most extreme proponents of deductive reasoning from first principles, Descartes for example, held that experiment was critically necessary to discover which of the reasonable options was instantiated in the actual world. Similar issues arise in the case at hand.²

In other words, you don't go to the math department to get your notion of 'good design', although you can ask them; but next you have to match that nice notion against your findings.

Tenable or not, SMT sets an appropriate standard for true explanation: anything that falls short is to that extent descriptive, introducing mechanisms that would not be found in a "more perfect" system satisfying just legibility conditions. If empirical evidence requires mechanisms that are "imperfections," they call for some independent account: perhaps path-dependent evolutionary history, properties of the brain, or some other source. It is worthwhile to keep this standard of explanation in mind whether or not some version of a minimalist thesis turns out to be valid.

This is a methodological point, of course.

These considerations bear directly on parametric variation, in this case yielding conclusions that are familiar features of linguistic inquiry. Any such variation is a prima facie imperfection: one seeks to restrict the variety for this reason alone.

Observe how this is an argument different from the familiar learnability one.

The same goal is grounded in independent concerns of explanatory adequacy/learnability, which require further that ineliminable parameters be easily detectable in data available for language acquisition. Both kinds of considerations (related, though distinct) indicate that study of language should be guided by the uniformity principle (1):

² See MI for discussion.

(1) In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances

One familiar application is the thesis that basic inflectional properties are universal though phonetically manifested in various ways (or not at all), stimulated by Jean-Roger Vergnaud's influential Case-theoretic proposals 20 years ago. Another is the thesis proposed by Hagit Borer and others that parametric variation is restricted to the lexicon, and insofar as syntactic computation is concerned, to a narrow category of morphological properties, primarily inflectional. These have been highly productive guidelines for research, extending earlier efforts with similar motivation (e.g., efforts to reduce the variety of phrase structure and transformational rules). What counts as "compelling" is, of course, a matter of judgment: there is no algorithm to determine when apparently disconfirming evidence is real or is the effect of unknown factors, hence to be held in abeyance.

I'd like to stress the continuity expressed in this paragraph, which could be extended to many other examples. In recent years a trend has emerged that sees minimalism as a break from the GB tradition. Although minimalism does use notions that were only incidental in GB (economy of derivations, economy of representations, last resort) the general concerns are old, and the technical solutions emerged in the course of limitations that previous versions of the model faced.

On such grounds, we try to eliminate levels apart from the interface levels,

Whatever those turn out to be, not an easy question.

and to maintain a bare phrase structure theory and the inclusiveness condition, which bars introduction of new elements (features) in the course of computation: indices, traces, etc. The indispensable operation of a recursive system is Merge (or some variant of it), which takes two syntactic objects $_$ and $_$ and forms the new object $_ = \{ _, _ \}$. We assume further that $_$ is of some determinate type: it has label $LB(_)$. In the best case, $LB(_) = LB(_)$ or $LB(_)$, determined by general algorithm.³This note is longer in the M-version, which speaks of how little use there is in this paper for labels, except for that of the new object formed (Collins's 'locus').

What follows is fast, and not without problems noted in my commentary to Minimalist Inquiries. In the M-version this paragraph is slightly elaborated on, but the technical issues persist. Let's set all of this to the side, though, assuming some appropriate, minimalist definition of 'command', e.g. in terms of active derivational workspaces.

³ On the possibility of dispensing with labels, see Collins (1999).

Merge yields two natural relations: Sister and Immediately-Contain (IC). Allowing ourselves the operation of transitive closure, we derive the relations Contain, Identity, and C-command.

While Merge "comes free," any other operation requires justification. Similarly, any features of lexical items that are not interpretable at the interface require justification. That includes most (maybe all) phonological features; these must be deleted or converted to interface-interpretable form by the phonological component. One might ask to what extent the phonological component is an optimal solution to the requirement of relating syntactic input to legible form, a hard question, not yet seriously addressed. We keep here to narrow syntax: computation of LF.

The empirical facts make it clear that there are (LF-)uninterpretable inflectional features that enter into agreement relations with interpretable inflectional features. Thus, the $_$ -features of T (Tense) are uninterpretable and agree with the interpretable $_$ -features of a nominal that may be local or remote, yielding the surface effect of noun-verb agreement.

The M-version has a footnote 4 associated to Tense, concerning the possibility of Agr nodes (because of Alternative II below). Some obvious difficulties with this alternative are mentioned there as well.

The obvious conclusion, which we adopt, is that the agreement relation removes the uninterpretable features from the narrow syntax, allowing derivations to converge at LF while remaining intact for the phonological component (with language-variant PF-manifestation).

We therefore have a relation Agree holding between $_$ and $_$, where $_$ has interpretable inflectional features and $_$ has uninterpretable ones, which delete under Agree.

An empirical question to ask here is whether this informal statement constitutes a necessary arrangement of interpretable and uninterpretable features, and if so what this follows from (good design?). In particular it is worth asking whether uninterpretable features can check uninterpretable ones, and if not why not (bad design?). See more on this below.

The relation Agree and uninterpretable features are prima facie imperfections. In M and earlier work it is suggested that both may be part of an optimal solution to minimal design specifications by virtue of their role in establishing the property of "displacement," which has (at least plausible) external motivation in terms of distinct kinds of semantic interpretation and perhaps processing.

The suggestion of an argument here is this: say there is some external motivation for displacement (unclear, but say that's true). Then the mechanism of uninterpretable features would be there in order to guarantee the displacement. There's a bit of a tautological flavor, though: since the system has displacement apparently associated to uninterpretable features, let's say that this is a consequence of good design, even if we don't understand how. An alternative, which as far as I can see has the same empirical coverage, would be to say: this is an imperfection, but one which the system uses to its advantage. Think of this as a 'viral' theory: you have to 'get rid' of uninterpretable features and you do, via movement. As a consequence the system finds itself in a new set of representations which can be used for new interpretive purposes. The second part has a reasoning like this.

If so, displacement is only an apparent imperfection of natural language, as are the devices that implement it.

If the alternative I mentioned is true, the imperfection is real. The facts remain unchanged, although not their motivation.

Displacement is implemented by selecting a target and a related category to be moved to a position determined by the target.⁴ The target also determines the kind of category that can be moved to this position. If uninterpretable inflectional features are the devices that implement displacement, we expect to find uninterpretable features of three kinds:

This is the technical statement. Note that the position here is this: there are three things you need in order to implement displacement via uninterpretable features (those in (2) below) and indeed you find three elements involved: the featural set, the EPP property, and Case.

(2) (i) to select a target _ [the probe]

(ii) to determine whether _ offers a position for movement and if so, what kind of category can move to that position

[in the M-version the 'and' clause is moved to (i); this shows, I think, that 'good design' is not always obvious...]

(iii) to select the category _ that is moved

That seems correct. For movement of a nominal to T, for example, the _-set and EPP-feature of _ = T serve the functions (i), (ii), respectively.

⁴ Terminology is often metaphoric here and below, adopted for expository convenience.

The M-version has a fn. 6 here: 'The EPP feature alone is not sufficient to identify a target; the $_$ -set (or comparable features, for other targets) is required to determine what kind of category K is sought.' And see my comment regarding (ii) above...

The category $_$ that is moved has uninterpretable structural Case, serving the function (iii). Agree is the relation between T and the moved category -- more precisely, their relevant subparts.⁵ This should be relevant, for instance, for adjectives. Let us say that the uninterpretable features of $_$ and $_$ render them active, so that matching leads to agreement. Locality conditions yield an intervention effect if probe $_$ matches inactive $_$ which is closer to $_$ than matching $_$, barring Agree($_$, $_$).

The picture seems to generalize over an interesting range. To the extent that this is true, uninterpretable features and the Agree relation are not true "imperfections," despite appearances.

Uninterpretability of features -- say, of phonological features, $_$ -features of T, or its EPP-feature -- is not "stipulated." The existence of these features is a question of fact: does L have these properties or not? If it does (as appears to be the case), we have to recognize the fact and seek to explain it: in the best case by showing that these are only apparent imperfections, part of an optimal solution to design specifications. Though motivated at the interface, interpretability of a feature is an inherent property that is accessible throughout the derivation. The phonological properties [\pm continuant], for example, are motivated only at the interface, but these "abstract" features are accessible throughout the derivation, which ultimately eliminates them in favor of narrow phonetic features interpretable at the interface.

This is not totally clear, and it is hard to distinguish from an alternative whereby interpretability is purely at the interface, but the interface is 'scattered'.

Similarly, interpretability of $_$ -features ([+] for N, [-] for T) is accessible throughout the derivation. For convergence, uninterpretable features must be deleted; in narrow syntax, we assume, by the operation Agree, establishing an agreement relation under appropriate conditions.

Suppose that L has generated the syntactic object K with label LB(K). On minimalist assumptions, LB(K) is the only element of K that is immediately accessible to L,

⁵ There is presumably a similar but distinct agreement relation, concord, involving Merge alone.

This is a natural stipulation, which tells you that the information under the 'top tier' of categorial representation is inaccessible to the system. I'm willing to grant that 'on minimalist assumptions', but an assumption does not follow, and as such it might be wrong.

so it must be the element that activates Agree, by virtue of its uninterpretable features: these constitute a probe that seeks a matching goal

In the M-version 'another collection of features' is added here.

within the domain of LB(K). What is the relation Match? The optimal candidate is Identity; we therefore take Match to be Identity.

There are two large paragraphs missing in the early version. The key idea in those paragraphs is presented later on in the present text: uninterpretable features (only them) enter the derivation without values. To avoid a confusing terminological issue, let's clearly separate a feature, a value, and a dimension. A feature is a valued dimension; for instance 'N' is a dimension '+' is a value and '+N' is a feature (basically, a concrete property). So what we're being told is that you can have mere dimensions in the lexicon without any value, if they are uninterpretable. These elements get their values determined by Agree, which results in their being deleted from narrow syntax, otherwise you couldn't tell them apart from interpretable features at LF, as those are valued to start with; nonetheless these guys remain in the phonological component. As a consequence of this implementation, it is not featural identity that is at stake in Agree, but actually dimensional identity. Chomsky's speaks in terms of 'non-distinctness', which seems unnecessary. In other words, all that a probe cares about is seeking an identical dimension lower down in its domain (not an identical feature). We can thus keep to the more minimalist notion of identity (which is obviously more elegant than 'non-distinctness'), and the important suggestion is this: grammar is sensitive to featural dimensions, not their specific values. That seems like a deep fact.

Another important technical paragraph states how Spell-out removes LF material which is uninterpretable and transfers the relevant object (WITH the uninterpretable stuff) to the phonological component. Fn. 8 of the M-version discusses the technical reason pointed out in MI of why this sort of system is necessary (overt syntax eliminates uninterpretable features, but they still have to have an effect on PF, thus the distinction in the Minimalist Program between 'deletion' and 'erasure'). Technically this is somewhat curious, I think, in that you need two representations of the relevant object K: one which is sent intact to

PF, and one which is sent to LF without uninterpretable stuff. Note also that Spell-out must determine what is uninterpretable, which is trivial prior to Agree (no values). This is used to motivate the cyclic application of Spell-out, or relevant information would be lost. Note also that if Spell-out applies without values having been assigned, Chomsky wants the derivation to crash at the interface. Here again you get that strange 'look ahead' relevant to not having the interface apply in a scattered fashion. You want to say that things crash at the interface even when, strictly, the interface as a level is something that the system doesn't reach until the end. Of course, Chomsky says that interpretability is a property that obtains throughout the derivation, even prior to reaching the interpretable components. I guess it's a bit like saying that you can violate an American law, being an American citizen, even when you're traveling abroad. You'll be tried upon your arrival in the US, but nonetheless you violated the law already in the place you committed the crime. Needless to say, as a consequence of this system there's no covert/overt distinction in the cycle. Chomsky also wants the phonological cycle to proceed in parallel, quite literally. He has to say that, as properties of relevant objects that don't make it to LF do make it to PF. One possibility that this suggests, of course, is that there are two entirely different systems at work: the syntactic one, which works with purely abstract categories, and then the phonological one, which access the syntactic one at the point of Spell-Out; that would eliminate part of the redundancy, but I'll return to this issue when Chomsky comments on Distributed Morphology.

I will keep here largely to Case-agreement and related systems: -features, structural Case, EPP, A-movement, and the core functional categories T, C, v (T = tense, C = complementizer, v a light verb that introduces verbal phrases).⁶This note is extended upon in the M-version, where the v* terminology is thus obliquely introduced. There is also a quick comment on Quirky Case, understood as inherent Case with a structural Case feature (I'm not sure I understand). Within these systems probe and goal match if features are valued for the goal and unvalued for the probe.

In the M-version there's an important added sentence: 'If -features were valued for the probe, it would be inactive and could drive no operation; if they were unvalued for the goal, they would receive no values from the (unvalued) matching features of the probe.' Note, however, that this still leaves the possibility (in principle) of mixed bags of features, with some valued and unvalued combinations.

If correct, the analysis should generalize to other core syntactic processes. Some

⁶ For expository purposes, we take the nominal with structural Case to be N and use T and C as cover terms for a richer array of functional categories, as in MI.

extensions to wh-movement are suggested in MI, but tentatively, for reasons indicated. This is the easiest case of A'-movement, since there are grounds to believe that features of probe and goal are involved. In other cases (e.g., topicalization, VP-fronting), postulation of features is much more stipulative; and throughout, questions arise about intermediate stages of successive-cyclic movement and island conditions.

Another humble admission. Keep it in mind if you work on A'-systems...

Matching of probe-goal induces Agree, eliminating uninterpretable features that activate them. A number of questions arise; specifically, with regard to the theses (3):

- (3) (i) Probe and goal must both be active for Agree to apply
(ii) must have a complete set of -features (it must be -complete) to delete uninterpretable features of the paired matching element

'Completeness' is a very interesting notion. Computationally, we will see that only complete categories are validly engaged in operations. I think representationally, also, there's something to pursue with regards to the notion 'completeness'. After all, there is an obvious correlation between complete categories and syntactic/semantic independency (e.g. compare complete T and partial T, for instance, or think of the difference between fully referential D's and indefinite or non-referential ones as complete D vs partial D; basically, the more features a category has the richer its semantic independence).

Let us tentatively adopt both theses, returning to the matter.⁷

For the Case-agreement systems, the uninterpretable features are -features of the probe and structural Case of the goal N. -features of N are interpretable; hence N is active only when it has structural Case. Once the Case value is determined, N no longer enters into agreement relations and is "frozen in place" (under (3i)). Structural Case is not a feature of the probes (T, v), but

The sentence:

it deletes under agreement if the probe is appropriate -- -complete, assuming (3ii).

is substituted in the M-version for: 'it is assigned a value under agreement, then removed by Spell-out from the narrow syntax. The value assigned depends on the probe: Nominative (NOM) for T, accusative

⁷ One question is whether (i) and (ii), if valid, follow from other properties of FL. That seems plausible, but there are many factors to be considered.

(ACC) for \underline{v} (alternatively ergative-absolutive, with different conditions).

The change of heart is understandable: mere deletion of Case would result in no Case realization, which might be fine for LF interpretation, but not for PF. So the idea is that the system somehow removes Case from the system, although it has not been valued in the same way as probing features are. The bottom line still is this:

Case itself is not matched, but deletes under matching of $\underline{\quad}$ -features.

So no matter how we cut it, Case is different from the other features. This poses a non-trivial conceptual question: could the system have worked optimally without this particular 'activation' system? Note, in particular, that it is not totally obvious why the system should involve *different* Case values; one could imagine a system whereby goals get activated on the basis of a single Case specification. Possibly, though, there are other reasons why the system requires different Case values, as opposed to a single one (which so far as I can see would have been enough for activation, since Case values carry absolutely no interpretive consequence, unlike all other values).

The following paragraph has been eliminated from the M-version:

We take uninterpretable features to be unvalued, receiving their values only under Agree. That is natural, given that the values are redundant. It is also empirically motivated by intervention effects (see MI). Accordingly, Match is not strictly speaking identity, but nondistinctness: same feature, independently of value.

Of course, this has now been incorporated before, as already noted.

In some cases, an active element E is unable to inactivate a matched element by deleting its unvalued features. E is defective, differing in some respect from otherwise identical active elements that induce deletion.

The M-version contains the clarification: 'The simplest way to express the distinction, requiring no new mechanisms or features, is in terms of (3ii): a nondefective probe is $\underline{\quad}$ -complete, a defective one is not.'

The M-version mentions participle-object constructions here, which can manifest (partial) $\underline{\quad}$ -feature agreement without Case assignment (the participle is defective). Then other cases:

The most familiar illustrations are raising constructions and their ECM counterparts, as schematically in (4i), where $\underline{\quad}$ is the matrix clause, $\underline{\quad}$ is an infinitival with YP a verbal phrase (the case most relevant here), and P is the probe: T with raising verb (case (ii)),

v with ECM transitive verb (case (iii))⁸:

(4) (i) [_{NP} [_{NP} [SUBJ [H YP]]]]

MATRIX INFINITIVAL

probe T

- (ii) (a) there are likely to be awarded several prizes
(b) several prizes are likely to be awarded

probe v

- (iii)(a) we expect there to be awarded several prizes
(b) we expect several prizes to be awarded

The Case-agreement properties of SUBJ [in (i)], and its overt location, are determined by properties of the matrix probe P, not internally to \bar{X} . \bar{X} is a TP with defective head T_{DEF} , which is unable to determine Case-agreement but has an EPP-feature, overtly manifested in (iii). Raising-ECM parallels give good reason to believe that the EPP-feature is manifested in (ii) as well, by trace of the matrix subject; preference of Merge over (more complex) Move gives a plausible reason for the surface distinction between SPEC- T_{DEF} in (ii) and in (iii) (see MI). In the (a) cases, the EPP-feature of T_{DEF} is satisfied by Merge of expletive; in (b) by raising of the direct object.

The following sentence has been displaced in the M-version, as noted:

The simplest way to distinguish T_{DEF} from a full probe, requiring no new mechanisms or features, is in terms of (3ii): a full probe is \bar{X} -complete, a defective one is not.

At this point there are several ways to proceed.

This is going to discuss the possibility that the EPP holds in infinitival clauses (the conventional approach) or not, the second alternative.

Alternative (I), the conventional approach adopted in MI, follows the path just outlined. SPEC- T_{DEF} of \bar{X} is filled either by Merge or Move, then associated to the higher probe. Examples (4ii, iiib) illustrate raising to subject, leaving reconstruction sites.⁹With

⁸In English, the expected form "awarded several prizes" surfaces more naturally as "several prizes awarded." We return to the matter.

⁹Reconstruction for trace of A-movement, as distinct from PRO, was discovered (to my knowledge) by Luigi Burzio, later published in Burzio (1986): e.g., such distinctions as "one interpreter each (was assigned t /*planned PRO to speak) to the visiting diplomats." For extensive review of these topics, see Fox (1999), Sauerland (1998). For a different approach to A-chains particularly, see Lasnik (forthcoming).

The first half of this note is interesting, since Chomsky does not usually speak of reconstruction in these terms.

P = T+raising verb, SPEC-T_{DEF} may apparently be manifested: e.g., in Icelandic defective infinitivals, possibly reflecting the TEC option (see MI).

This comment is interesting, because Chomsky does not usually recognize reconstruction possibilities for A-movement; we'll discuss the footnote later on.

T_{DEF} matches SUBJ in some of its features (to implement raising) but not all (to preclude inactivation). If P and T_{DEF} match SUBJ in the feature [person], then categories with this feature, and only these, can undergo raising (nominals but not adjectivals); on the simplest assumptions, T_{DEF} has no other --features.

The 'If P...' part is straightforward, but the '...and Tdef' part is less obvious. The 'then categories...' consequence would follow even if only P matches SUBJ in person. Then there is not much to say about the --features of Tdef.

Expletives too must have the feature [person], since they raise; and pure expletives of the there-type should have no other formal features, on the simplest assumptions.

One has to be careful with these 'shoulds', which can be misinterpreted. After all, we do know that expletives differ cross-linguistically, in terms of whether they do or do not induce agreement, or a definiteness effect, for instance; it would seem that the way to capture those differences is in terms of features other than [person] (in the more accurate terminology, a [person] dimension, or unvalued feature [person]).

In a framework that dispenses with categorial features, as is reasonable on minimalist grounds, [person] plays the role formerly assigned to [D] or [N] features.¹⁰

When --complete, T values and deletes structural Case for N. The --set of N (which is always --complete) both values and deletes the --features of T (with or without movement).

¹⁰ On eliminating categorial features in favor of root structures with functional heads, see Marantz (1997), MI. **This suggests an approach to syntactic computation along the lines explored by Murgia 1999 for ellipsis, which relates to the issue posed for the parallel PF and LF objects, mentioned in the text.**

On light verbs in particular, see among others Hale and Keyser (1993), Harley (1995). Functional categories lacking semantic features require complication of phrase structure theory (see MI), a departure from good design to be avoided unless forced. AGR elements, for example, should arouse skepticism.

right, but then see note 14. What follows is also plausible.

Similarly, D -- or at least one variant of D -- might be associated with referentiality in some sense, not just treated as an automatic marker of "nominal category"; nonreferential nominals (nonspecifics, quantified and predicate nominals, etc.) need not then be assigned automatic D (at least, this variant of D).

A slightly worrisome issue here is that, clearly, languages differ in whether their $_$ -set of N is overt. It may be that, for instance, the $_$ -set of English N is actually not complete, and thus for instance is incapable of licensing the null category 'one' that it licenses in Spanish or Portuguese, where it might be complete. If that straightforward point were correct, we'd need to slightly modify our assumptions about how the N set works. I'm raising this just as a matter of principle.

With defective probe, agreement is not manifested and Case of the matched goal is not assigned a value: raising T exhibits no agreement, and participles lack person;

But they do agree in number and often gender, which raises a warning flag (we return to participles).

neither determines the Case of matched N, which depends on a higher non-defective probe, T or v (see (18), below).

Similar properties hold in other constructions, e.g., attributive adjectival/participial constructions ("[old/smashed] car," "a car [old enough to buy/smashed into pieces]"). Whatever the correct analysis may be, these constructions involve a relation between N and the head of the predicate phrase; the complete $_$ -set of N values and deletes the matched uninterpretable features of the predicate, but the partial $_$ -set of the predicate does not value and delete Case in N, which still has to satisfy the Case Filter.

Right, so two possibilities emerge: a) these are special, probing preceding in the absence of Case activation; or b) there is also some form of Case activation in these instances, albeit not the usual one that shows up in argumental positions. For example, it might be possible that null Case is involved in these instances, and more generally whenever number (not person) features are checked. That of course presupposes a system of null Case assignment. Again, I'm raising these as possibilities that exist in principle, and do not seem particularly unreasonable.

In the MI framework C is one-one associated with $_$ -complete T (T_{COMP}):

(5) C selects T_{COMP} ; V selects T_{DEF}

That's surely a fact, but one whose cause we'd like to understand (e.g., why doesn't T_{COMP} select T_{DEF} ? What would be wrong with that?)

Control structures and finite clauses have the selectional relation $C-T_{COMP}$, while raising clauses have the relation $V-T_{DEF}$.

Of course, it is slightly odd to say that the T of Control structures is 'as

complete' as the T of regular complement sentences, if only because in the former you have obligatory *consecutio temporum*.

The reasons in MI were largely theory-internal, having to do with Case-agreement and related systems. The conclusions are consistent with those reached on other grounds. The earliest straightforward evidence that raising clauses fall together with finite TP, and control clauses with CP, was provided by Luigi Rizzi, who observed that control clauses, like CPs, are phonetically isolable in ways that raising clauses are not, nor of course finite TP (stranding its complementizer).¹¹ **[this footnote contains relevant data]**

These conclusions suggest a possible recasting of the account of defective elements, Alternative (II), with the invariant property (6):¹²

I will not comment on this alternative seriously because it is not pursued thoroughly here, and in any case it faces some non-trivial difficulties (e.g. the one pointed out in fn. 14, which forces the postulation of separate Agr nodes)

(6) C is $_$ -complete; T is $_$ -complete only when necessary

One case in which T must be $_$ -complete, it could be argued, is selection by $_$ -complete $_$ with uninterpretable features (specifically, $_ = C$). The selectional property could then be formulated in terms of Match/Agree: the $_$ -features of $_$ have to be deleted under Agree by T, which therefore must be $_$ -complete (crash with failure of match is detectable at once); we return to some problems. It is tempting to associate EPP with $_$ -completeness: C, and T selected by C, are $_$ -complete, and therefore allow an EPP-feature; T_{DEF} cannot have an EPP-feature. Accordingly, there is no internal raising to SPEC-T_{DEF}; raising is "in one fell swoop" in such constructions as (4ii), and there are no intermediate reconstruction sites.¹³ Case-agreement and EPP proceed as before, with T_{COMP}.

The symmetry of raising and ECM constructions suggests that the analysis should be extended to the latter as well.¹⁴ Second Merge of first-merged object of V makes little sense.

¹¹ **IMPORTANT DATA:** E.g., "it is to go home (every evening) that John prefers (*seems)"; Rizzi (1982). Other early evidence involved reconstruction effects (see note 9). Rizzi explained the distinction in terms of government, a notion not readily available in the framework here.

¹² Some related ideas are developed by Pesetsky and Torrego (forthcoming).

¹³ See Epstein and Seely (1999) for proposals to this effect. **And Castillo, Drury, and Grohmann (1999), who carefully argue for Alternative II.**

¹⁴ **This note is important.**

The extension is suggested by proposals of Koizumi (1995), Lasnik (1999, forthcoming), and Epstein and Seely (1999), differing from one another and from the account here, which is, furthermore, oversimplified.

The account should be restated, as in the sources cited, in terms of an AGR node selecting V; and by symmetry, selecting T (AGR selected by appropriate v and C). It is, then, AGR and not T/v that is the locus of $_$ -features, Case, and EPP, in the version presented here. Since I will not be pursuing this course, in part for reasons discussed later, I will keep to the simplified exposition. **Just as C_{comp} selects T_{COMP} , we might expect v_{COMP} to select V_{comp}**

The following...

-- that is, V with a complete set of $_$ -features; it is, then, the $_$ -features of V that enter into the Case-agreement system, parallel to T_{comp} selected by C. The light verb v is $_$ -complete in a construction with full argument structure: call it v^* , transitive \underline{v} or experiencer.¹⁵ The analogy, then, is C - T_{comp} , v^* - V_{comp} .

is partly deleted and partly put into fn. 9 (M) in the M-version.

Being $_$ -complete, C must select T_{comp} for its unvalued features to delete, and it allows an EPP-feature. For the same reasons, v^* (being $_$ -complete) must select V_{COMP} for its (unvalued) $_$ -features to delete, and it allows an EPP-feature. For both C and \underline{v} , the selectional property reduces to Match/Agree.

Unless selected by C or v^* , T and V are defective (raising T, passive/unaccusative V, respectively). They do not enter into Case-agreement, and have no EPP-feature. When selected by C or v^* , T and V are $_$ -complete, entering into Case-agreement structures (with raising of associate or not, depending on optionality of the permitted EPP-feature and availability of alternatives to satisfy it). In a transitive construction, the object agrees with V and is assigned Accusative Case (raising to SPEC-V if V has an EPP-feature). There is no internal raising to SPEC- T_{def} in raising or ECM constructions.¹⁶

Consider ECM constructions more closely [**under alternative II**]. The verbal phrase is: v -[V-TP]. If the light verb v is $_$ -incomplete (passive), then V is defective, as is T_v selected by V (in raising/ECM constructions). If $v = v^*$, then V must be $_$ -complete for convergence. But (6) does not require that T_v be $_$ -complete in this case, because the $_$ -set of V is valued and deleted independently by the embedded subject. Therefore T remains defective, and the raising/ECM parallelism remains intact.

The discussion suggests that T should be construed as a substantive rather than

especially this part:

¹⁵ Only \underline{v} was considered in MI, and I will largely keep to it for exposition below. Where assigned by V, not v, Case is inherent. Quirky Case largely falls under general Case-assignment principles if understood to be inherent Case with an additional structural Case feature associated with $_$ -complete v^* (as in MI). We return to examples, some problematic.

¹⁶ For T_{comp} , the EPP-feature is apparently obligatory; for V_{comp} , as well, for the sources cited in fn.14.

a functional category, falling together with N and V, perhaps others, a possibility that is neutral between Alternatives (I) and (II). We can regard T as the locus of tense/event structure (see note 6). The C-T relation is therefore analogous to the v^* -V relation.

This effectively ends discussion of alternative II (only further reference in notes). Observe how the comment about T being substantive is highlighted, since Chomsky wants to use that later on within the confines of alternative I.

Alternative (I) takes the locus of Case-agreement/EPP to be T, v^* ; Alternative (II) takes it to be T, V. Let us refer to the two alternatives as $\text{LOCUS}_{\text{TV}^*}$ and LOCUS_{TV} , focusing on their basic conceptual difference: the choice of category relevant to Case-agreement/EPP.

The agreement issue noted in footnote 14 is also a serious conceptual difference.

Much of what follows is neutral among these alternatives or variants that merit consideration as well.¹⁷ I will continue with the conventional choice $\text{LOCUS}_{\text{TV}^*}$, returning to others where appropriate, along with lingering problems.

Suppose that the label LB(K) of K has an uninterpretable selectional feature (by definition, an EPP-feature), which requires Merge in SPEC of LB(K). That can be satisfied by Merge of an expletive, in which case long-distance agreement may hold between LB(K) and the goal. Alternatively, an active goal G determines a category PP(G) (Pied-Piping), which is merged in SPEC-LB(K), yielding the displacement property.

The ‘uninterpretable selectional feature’ has gone by many names in the last forty years. To call it ‘selectional’ suggests a connection with theories of selection that does not seem obvious. Moreover, it is peculiar that this property should be satisfiable either by expletive Merge or by Move; a disjunction always hides a generalization being missed. (You may think that Move includes Merge, but that’s given up by the end of the paper.) Also, by allowing this sort of situation in, the elegance of the Probe/goal situation is considerably weakened. So it is possible that this is a genuine imperfection, and an interesting one (Chomsky doesn’t even try to say that this one is not an imperfection, at least not until now). You should perhaps keep in mind how similar expletive-associate relations are to case-NP relations in languages where the case-marker is overtly associated to a full phrase. A question to ask is whether expletive-associate dependencies might not be (one of) the origins of case systems

¹⁷ One variant might supplement (6) by taking T (and substantive categories generally, including V) to be always defective, so that the locus of Nominative Case and subject-verb agreement is C, not T (see reference of note 12 for related ideas). In the notation just used, this alternative falls under $\text{LOCUS}_{\text{CV}^*}$.

(but see fn. 18).

The combination of Agree/Pied-Pipe/Merge is the composite operation Move, preempted where possible by the simpler operations Merge and Agree.¹⁸

FL specifies the features **F** that are available to fix each particular language L. The Ml framework takes L to be a derivational procedure that maps **F** to {EXP}, where an expression EXP is a set of interface representations. As a first approximation, take EXP to be {PF, LF}, these being symbolic objects at the sensorimotor and conceptual-intentional interfaces, respectively.

Note: this is just a ‘first approximation’, probably a wrong one.

We adopt the conventional (usually tacit) assumption that L makes a one-time selection [F_L] from **F**. These are the features that enter into L; others can be disregarded in use of L.

There is a serious issue lurking here. Chomsky appears to be taking the view that the features of the syntax are much like those of phonology, a subset of those allowed by universal phonetics. However, we saw before that phonology probably does not have to obey the uniformity principle in (1), in part because its properties are readily available to the learner. The issue for syntactic properties is much less clear, though. Two possibilities arise. One is that if a language presents no evidence to the learner of the presence of a given syntactic category, then that category simply does not exist. For example, take noun classifiers in Japanese and in English, overt in the first instance. Does English have them? According to this view, no. However, as a consequence, Japanese and English LFs are going to be radically different, for example in the quantificational system. That is not necessarily a problem for learnability if the English system is the default one and you need positive evidence for the Japanese one. There is still an issue, learnability aside, as to whether this situation would be an ‘imperfection’, which would create two rather different systems of thought (for this view, see e.g. Gil (1987)). The other alternative is this: both Japanese and English have classifiers (Muromatsu (1998)), albeit they are covert in the latter. In this general view the problem is descriptive: under what circumstances does a language has an overt or a null classifier. The learnability issue is

¹⁸ It has occasionally been suggested that the EPP-feature of the probe P is a Case-assigning feature F_C. If correct, that would scarcely change the picture: as before, F_C is an uninterpretable selectional feature inducing Merge (sometimes Move), distinct from the features of P that enter into Agree. But it does not seem to be correct. There is good reason to believe that structural Case correlates with agreement, hence also long-distance agreement (without raising to SPEC), and accordingly that EPP is satisfied without Case-assignment (see (18) and discussion). If either is correct (both seem to be), Case-assignment and EPP are independent phenomena.

unchanged and the elegance question is now resolved. But note that in the second instance the set of actual and possible syntactic features is really identical, and the task of the grammarian is essentially to find out the set of universal syntactic categories.

Assume further that L assembles [F_i] to lexical items LI of a lexicon LEX, the LIs then entering into computations as units. In the simplest case, LEX is a single collection, but empirical phenomena might call for "distribution" of LEX, with late insertion in the manner of Distributed Morphology (DM).¹⁹ In any case, we can think of LEX as in principle "Bloomfieldian," a "list of exceptions" that provides just the information required to yield the interface outputs, and does so in the best way, with least redundancy and complication.

This aside is in slight antagonism with the Bloomfieldian character. After all, why should the lexicon be exceptional or accidental? Presumably because it is historical, to a large extent. If so, though, why should it be yielding any optimality? What kind of optimality? More or less than in phonology more generally?

In the simplest case, the entry LI is a once-and-for-all collection (perhaps structured) of (A) phonological, (B) semantic, and (C) formal features.

The parenthetical comment 'perhaps structured' is also a bit perplexing. If these features are indeed structured, what does that structure follow from? It must be either accidental (in which case it is possibly irrelevant) or else follow from some principled mechanism. But what is the nature of that pre-derivational mechanism? This is something to keep in mind, also, for those theories of theta-roles which take them to be features of lexical items. If they are, they should be a random collection, with no hierarchical or other structural properties. If on the other hand roles are ordered, somehow, we want to understand what is the principle that predicts that ordering, and what component of grammar is responsible for it.

The features of (A) are accessed in the phonological component, ultimately yielding a PF-interface representation; those of (B) are interpreted at LF; and those of (C) are accessible in the course of the narrow-syntactic derivation. Language design is such that (B) and (C) intersect, and are disjoint from (A),

In principle, either language design or an accident of evolution resulted in this property, assuming it is right. It is perhaps worth emphasizing that if we just blame whatever property we find on language design, without any

¹⁹ On the current state of DM, see Harley and Noyer (1999).

justification, the theory becomes circular.

though there is some evidence, to which we return, that presence or absence of features of (A) might have an effect on narrow syntactic computation.

These are a kind of stylistic rules discussed at the end of the paper, which pose a non-trivial complication of the model, although one that has very interesting potential consequences.

It also seems that FL may retain something like (B)-(C) and the narrow syntax in which they enter while the phonological component is replaced by other means of sensori-motor access to narrow-syntactic derivations, as in sign language.

There's a presupposition here that I'm not sure one can directly grant; that in signed languages you literally 'replace' whatever sensori-motor mechanism you had in spoken languages. That of course presupposes that the latter is prior, e.g. in evolutionary terms. So far as I know, this is a controversial claim. Bottom line: signed languages function just like spoken languages do, with a different kind of sensorimotor support. So 'sound' PF cannot be essential to the language faculty, it has to be something more abstract.

Of particular interest is the subset of (C) that is not in (B): uninterpretable formal features that appear, *prima facie*, to violate conditions of optimal design.

For open classes, the optimal account is typically the simplest, for obvious learnability reasons: LI is a unitary collection, including the phonological matrix. LEX is distributed when departure from the simplest account is warranted in favor of late insertion, typically for inflectional elements and suppletion.

So there is a distinction being drawn here between substantive and grammatical items. Note that the fact that the first type is a 'unitary collection' doesn't entail that it has to be inserted early. It is perfectly possible (hard to determine empirically) that the unitary collection in question is inserted immediately before Spell-Out.

Throughout, answers depend on predictability of phonetic outcome by general phonological principles that satisfy UG conditions, and in all cases, the simplest choice should win. For roots and highly predictable inflectional elements (say, English progressive), the distinctions between single-LI and several independent contributions to LI (as in DM systems postulating universal late insertion ULI) seem to have little empirical content,

Again, you don't care whether the rules determining the outcome take

place at a single point, in a separate phonological cycle, or they are distributed throughout the syntactic cycle. However, in less predictable instances...

but they might, for example, when an idiosyncratic feature F of a root has syntactic effects. ULI then requires postulation of a redundant syntactic feature F' as a "place holder" in narrow syntax for F, with a stipulation that F' must be replaced under late insertion by a root with F (i.e., F' is effectively identical with F).²⁰ A unitary LEX avoids the redundancy and stipulation (and is preferable on conceptual grounds in any event). The substantive results of DM remain unchanged.

If that comment seemed incomprehensible, don't worry; it is. It pertains to how you want to do your morphology in those instances where it matters, and whether you need to code it in terms of an abstract feature in the syntax that you replace upon matching (a stipulative matching in this instance). Since I know very little about how morphology should be done, I will not dwell on this now.

Narrow syntax maps a selection of choices from LEX to LF; the phonological component, in contrast, has further access to [F_i].

I suppose that, in some instances, PF has to manipulate the items that the syntax has provided (e.g. when you get a verb and a clitic together in the syntax, and then you delete some consonant or change some vowel). That's fine, however...

Like the extraction of [F_i] from F, these assumptions, largely conventional, reduce the computational burden for the procedure L while adding new conceptual apparatus.

I'm not sure in what sense what I said reduces computational complexity, to be honest.

More controversially, MI extends the same reasoning to individual derivations: L makes a one-time selection of a lexical array LA, a collection of LIs (a "numeration" if some are selected more than once), and maps LA to EXP. Again, there is a reduction of computational burden, in this case a vast reduction, since LEX, which virtually exhausts L, need no longer be accessed in the derivation once LA is selected.

That's certainly true, but then I'm perplexed: why did the system take the step of reducing F to FI, in turn assembled into LEX, if after all this is all nothing compared to the reduction from LEX to LA? Why not go from F to LA? Two answers come to mind. a) The intermediate steps (FI, LEX) happen to be historical accidents or evolutionary ones; yes, it would have

²⁰ For an illustration in the case of Latin deponent verbs, see Embick (1999).

been nice to go from F to LA, but the super-engineer was blind, after all. b) The super-engineer wasn't blind at all, we are; if you didn't have intermediate steps FI, LEX, watch what would go terribly wrong: ... (Unfortunately I'm no super-engineer, so don't know how to fill in the blanks.) Incidentally, F to LA, or more precisely F to (PF, LF) is generative semantics. Option a) above would say that generative semantics is the optimal path to take, but it so happens that things went otherwise; option b) above has to take the view that lexicalism is more optimal than its alternative. It's hard to see how that can be argued on the basis of the arguments in the late sixties, which were all based on facts that stressed the peculiarity of the lexicon: unlike the syntax, the lexicon is non-transparent, idiosyncratic, unsystematic -not precisely an argument of elegance.

The new concept LA (numeration) is added,

Strictly, 'numeration' is a KIND of LA: one where lexical TOKENS are coded.

while another concept is eliminated: chains are determined by identity, with no need for indices or some similar device to distinguish chains from repetitions, also violating the inclusiveness condition.

That is, chains are determined by identity with regards to the lexical tokens where they originate (the derivation manipulates those tokens, in essence, by turning them into categorial operators which take sets of configurational contexts as arguments -see Part Two). A non-trivial issue is how to notate tokens. If we do it by indices (the common practice) we replace one notation for another. On the other hand, it is interesting that the grammar has devices such as Case which would seem to serve no purpose other than 'notating' (internal to the derivation) different argument chains. If that's the right view, different Case *values* (as opposed to the Case dimension) and similar formatives for other kinds of chains might well be tokenizers.

As in the other cases, the tests are ultimately empirical; on purely conceptual grounds, one could argue either way. As noted, the nature of optimal design that is instantiated in FL (if any) is a matter of discovery, within certain guidelines.

Proceeding further, MI proposes another reduction of computational burden: the derivation of EXP proceeds by phase, where each phase is determined by a subarray LA_{ij} of LA, placed in "active memory."

The problem noted above for the generative semantics vs. lexicalism

conceptions comes to the fore again now. Why shouldn't the system map directly from thought to phases? Don't think that this would give you nothing. In fact, it would give you a structure very similar to that in early Neo-davidsonian systems, or to dynamic binding. True, each phase would have to be a separate thought, but you could connect each thought paratactically to others, in much the same way you connect conjunctions or successive texts (this very paragraph). Of course, we know factually that syntax is largely hypotactic, but why should that be? It would if phases must for some reason happen within the confines of a numeration, not directly the lexicon, or thought, or whatever. I think that's how phases work, but the question is why should they, and computational complexity alone does not give you that, unless I'm missing something. As a point of comparison, consider the logic I gave in Multiple Spell-out (MSO) for what I called cascades there. Cascades are of course similar to phases, but the system goes into cascades not because of any need to reduce complexity (although it gets that as a *result*), but because if the system had not gone into the relevant sub-chunk, the derivation would crash. The reasons for this are irrelevant now (in fact, in different papers I showed that there could be different reasons, but this is immaterial to the argument). The important point is that we must decide whether a phase should be something that emerges in the derivational dynamics because, otherwise, convergence would for some reason not be possible (alternatively, because convergence would be then possible in a more optimal way), or else whether phases are stipulated as computational reductions. If the latter (Chomsky's view in this paper) I think we have to address why the system didn't take a radically 'generative semantics' route.

When the computation exhausts LA_{α} , forming the syntactic object K , L returns to LA , either extending K to K' or forming an independent structure M to be assimilated later to K or to some extension of K .

The first of these two options is entirely straightforward: you finish K , so you get stuff from LA to make K larger still, and so you build K' . The other option should be worked out in detail. What is this business of forming a structure that you assimilate later? When? In the same phase? What happens with arbitrarily complex phrases that could 'suck up' lexical material from higher phases. Take for instance [*A thing that I said*] *impressed him* vs. *I said [that a thing impressed him]*. How do we form the latter? Taking material from the higher phase? Putting the 'extra' material already in the lower phase? If the latter, how many phases are there in the 'lower' phase? I'm sure all these questions have answers, but you want to think about what they are. See also fn. 21.

Derivation is assumed to be strictly cyclic, but with the phase level of the cycle playing a special role.²¹ Suppose that $LA_{\underline{it}}$ is eliminated and phases are constructed from LA (or the entire LEX). The cost is greater computational burden, in a certain sense. Again, the conceptual arguments are not decisive.

Keep in mind the question whether the cycle is deduced or stipulated.

A subarray $LA_{\underline{it}}$ must be easily identifiable; optimally, it should contain exactly one lexical item that will label the resulting phase.

I don't see this. Why would that be optimal and why should the item be lexical (or is 'lexical' not used here as opposed to 'functional')?

Assume that the substantive categories nominal and verbal (perhaps T as well) are headed by functional categories: for verbal phrases, a light verb (see note 10).

Headed? Is that the right term? Does a light \underline{v} literally head VP? In the M-version that sentence has been replaced for: 'Assume that substantive categories are selected by functional categories: V by a light verb, T by C.' That appears to be more accurate. Unfortunately, that sentence is placed after the connective 'if so...' below:

The evidence reviewed in MI suggested that the phases are "propositional": verbal phrases with full argument structure and CP with force indicators, but not TP alone or "weak" verbal configurations lacking external arguments (passive, unaccusative). If so, phases are CP and v^*P , and a subarray contains exactly one C or v^* .

Although I don't understand how C and \underline{v} related phrases are 'propositional', I do understand that phases are perhaps related to verbs

²¹ $LA_{\underline{it}}$ is a subset of LA, drawn from LEX,

Note: 'drawn' from the lexicon (a set of types) not a sub-set of the lexicon.

but the objects it makes available are those labeled by elements of $LA_{\underline{it}}$, perhaps complex objects already constructed in the course of the derivation, which proceeds in parallel.

I don't follow this well. What does 'it' refer to? LAi? if so, what does this mean: 'the objects LAi makes available are those labeled by elements of LAi'? Does 'it' refer to LA? If so: 'the objects LA makes available are those labeled by elements of LAi'. That is at least coherent, but what does it mean? In fact, what is to 'make available'? A technical term?

Suppose that LA is eliminated in favor of iterated selection of $LA_{\underline{it}}$ from LEX. The cost is that reduction of computational burden is far less and that search of arbitrary depth is needed to determine whether an item selected from LEX is new.

Is this meant seriously or metaphorically? Do we really know what the 'cost' of the 'search'?

and Comp, and 'if so', phases are CP and v*P. But after the phrase above, 'assume that substantive categories are selected by functional categories...' is placed before 'if so...', I don't see how then it follows that phases are CP and v*P. Or is this supposed to mean that phases are determined by functional categories selecting lexical categories? (That is VP selected by v*P, and TP -which Chomsky suggests is substantive-selected by CP, but not v*P selected by TP). If that's what is meant, it is somewhat interesting, as then phases would be defined as substantive structure and their corresponding functional layer (perhaps a 'tokenizer'), not really a proposition or a 'proposition', which is good news, as the new notion is clearer.

The choice of phases has independent support: these are reconstruction sites, and have a degree of phonetic independence (as already noted for CP vs. TP).

It is clear that CP is a reconstruction site, but it is much less obvious that v*P is, especially if we also want to say that vP or TP are not. In fact, I don't know of any argument to that effect (although see the references in fn. 22, which I haven't read).

The same is true of vP constructions generally.²² If these too are phases, then PF and LF integrity correlate more generally.

Suppose, then, we take CP and vP to be phases. Nonetheless, there remains an important distinction between CP/v*P phases and others; call the former strong phases and the latter weak.

I find the presentation confusing. What 'latter' phases are we talking about? Haven't we explicit said that CP and v*P are the phases? Or is the oblique comment above about 'vP constructions more generally' meant to extend the notions from v*P to other vPs? Other categories too? DP? But not TP... Why not? Reconstruction to TP has been argued, at least, and TP certainly is the target of movement, obligatorily in some accounts (standard EPP). So I'm not sure which are the other phases, although we'll be using the term *strong* phase to refer to CP and v*P. The more distinctions of this sort that we make, the more notions we'll have to justify in less obvious grounds.

The strong phases are potential targets for movement; C and v* may have an EPP-feature, which provides a position for XP-movement, and the observation can be generalized to head-movement of the kind relevant here.²³

²² See Legate (1998), and for the broader picture with regard to reconstruction in such cases, Fox (1999).

²³ Namely, head-movement involving inflectional categories: to C, or to T (hence to a position between the vP and

The special role of strong phases becomes significant in the light of another suggestion of MI that I will adopt and extend here: Spell-Out is cyclic, at the phase level.

In all fairness, that suggestion is considerably older than MI. It goes back at least to Bresnan's work in the early seventies. Note how the M-version changes the phrase 'at the phase level' for 'at the strong phase level'. To say that spell-out happens at the phases is stipulative (cf. the MSO system, where if that hadn't happened, the derivation would not converge); to say that spell-out happens at the *strong* phases only is of course an even narrower stipulation.

These sentences:

In contrast to EST-based systems, there is no overt-covert distinction with two independent cycles; rather, a single narrow-syntactic cycle. Furthermore, the phonological cycle is not a third independent cycle, but proceeds essentially in parallel.

have been displaced in the M-version (we've already seen them). The reasoning above -a justification for why only strong phases induce spell-out- continues as follows:

The intuitive idea, to be sharpened, is that features deleted within the cyclic computation remain until the phase level, at which point the whole phase is "handed over" to the phonological component. The deleted features then disappear from the narrow syntax, allowing convergence at LF, but they may have phonetic effects.

All of that would be true, also, if it were regular, not strong phases, the ones that go to early spell-out. So we still have to understand why it is that things happen in a class of phase and not in others.

The material that follows is now in fn. 8 of the M-version:

In these terms we can overcome a paradox of EST-based systems with a single position for Spell-Out: the "overt" part of the narrow-syntactic computation eliminates uninterpretable features, but they have to remain until the stage of Spell-Out of the full syntactic object, because of their phonetic reflexes.

The following replaces that material: 'Spell-out seeks formal features that are uninterpretable but have been assigned values (checked); these are

CP phases; see note 6). We return to the status of head-movement.

removed from the narrow syntax as the syntactic object is transferred to the phonology. The valued uninterpretable features can be detected with only limited inspection of the derivation if earlier stages of the cycle can be “forgotten”; in phase terms, if earlier phases need not be inspected.’

The computational burden is further reduced if the phonological component [**added in M-version: “too”**] can “forget” earlier stages of derivation. That follows from the Phase Impenetrability Condition (PIC) (MI (21)), for strong phase HP with head H:

- (7) The domain of H is not accessible to operations outside HP, but only H and its edge [**are accessible to operations outside HP**],

the edge being the residue outside of H-bar, either SPECs or elements adjoined to HP.

Once again, we’re being given a computational, complexity argument. The idea is to have mechanisms in the system (in this instance the PIC) that prevent any sort of global ‘search’ to determine the *structural description* of a transformation; that is added to the assumption that the *structural change* of a transformation is cyclically constrained (the Extension Condition). The computational window at any given point H is very narrow: You have access to what is in the (in the older versions, complement) domain of H -and even that, only up to the next phase- and what is in H’s immediate edge (old ‘checking domain’). Again one could ask why the system doesn’t seek this complexity *directly*. It seems as if something forces the system into a familiar, embedding structure, and once that’s in place for some reason, the optimizations we’re talking about kick in. But it is far from obvious why there is a sort of ‘optimization hierarchy’, which seeks the best possible arrangement (in complexity terms) among those permitted by this sort of structure. This sort of situation arises in successive steps of evolution, but we’re not supposed to be looking at that...

Accessibility of H and its edge is only up to the next strong phase, under PIC: in (8), elements of HP are accessible to operations within the smallest strong ZP phase but not beyond.

- (8) [_{ZP} Z... [_{HP} _ [H YP]]]

Local head movement and successive-cyclic A- and A'-movement are allowed, and the phonological component can proceed without checking back to earlier stages. The simplest assumption is that the phonological component spells out elements that undergo no further displacement, with no need for further specification.²⁴

²⁴ The idea that the phonological component can choose which element of a chain to spell out has been investigated

I'm not sure why that's the simplest assumption. I would have thought that the simplest assumption is to spell-out only once per (strong) phase, as that involves a single operation. But if we go with spelling-out elements that have stopped moving, or whatever, then we're going to be spelling out as many times as elements of that sort exist in the structure. Isn't it computationally simpler to execute the operations once and for all within the relevant sub-domains?

In effect, H and its edge $_$ in (8) belong to ZP for the purposes of Spell-Out, under PIC. YP is spelled out at the level HP. H and $_$ are spelled out if they remain in situ. Otherwise their status is determined in the same way at the next strong phase ZP. The question arises only for the edge $_$, assuming that excorporation is disallowed.

It is interesting to ask why this state of affairs should arise:

 | [_{ZP} Z... [_{HP} $_$ | [H YP]]]

For those familiar with Drury's top-down system, this does not come as a surprise, since the relevant chunk of structure is a unit (as in Phillips's thesis) and the unit goes to Spell-out upon merging to the rest of the structure. The question is whether there is any natural way of predicting the fact that there is a mismatch between phrasal objects for LF and for PF in non top-down systems; the way Chomsky puts it: 'In effect, H and its edge $_$ in (8) belong to ZP for the purposes of Spell-Out, under PIC.' PIC does give you that, but by stipulation, unlike in Drury's system, where things followed from the architecture.

The picture improves further if interpretation/evaluation is uniformly at the next higher phase, with Spell-Out just a special case. Assuming so, we adopt the guiding principle (9) for phases PH_{ij} :

(9) Interpretation/evaluation for PH_1 is at the next relevant phase PH_2

This is meant as a mere 'guiding principle', since 'relevant phase' is not determined yet.

What are the relevant phases? As noted, because of the availability of EPP, the effects of Spell-Out are determined at the next higher strong phase: CP or v*P. For the

since Groat and O'Neil (1996). Any such approach requires either new UG principles or language-specific rules to determine how the choice is made.

same reason, a strong phase HP allows extraction to its outer edge, so the domain of H can be assumed to be inaccessible to extraction under PIC: an element to be extracted can be raised to the edge, and the phonological component can spell out the domain at once, not waiting for the next phase.

This is in the tradition of *Barriers* and the Subjacency condition, and it relates to a technical problem that arises in 'chapter 4' and MI: *who did you see* -how to get 'object shift' in a language where the process is not overt (English is SVO). Suppose you moved directly from object position to the Wh-site. Then at LF (in 'chapter 4' terms) you'd have to do something with the features of \bar{v} , but in this system you can't just move a trace, which is part of a non-configurational object. The other alternative is to have moved through the \bar{v} spec overtly, as in OS languages. Although normally you would not do that in English. In order not to violate Last Resort, then, you'd have to postulate some feature for the intermediate movement. The solution provided in MI, extended here, is to treat \bar{v} and C as you treat T, as EPP targets. Then all languages satisfy EPP requirements in these categories, although some do it more systematically than others (see Part Two for discussion of these and related matters). We can then use these syntactic areas where edges seem relevant as a heuristic to determine 'next relevant phase' in (9).

Keeping to the optimal assumption that all operations are subject to the same conditions, we restate (9) as (10), where PH_1 is strong and PH_2 is the next highest strong phase:

(10) Interpretation/evaluation for PH_1 is at PH_2

At this point I have a real hard time distinguishing a system along the lines of Epstein and Seely, or what I called the 'radical' version of MSO, from what is being said here. There might be differences of detail in what (doesn't) go to interpretation, but the bottom line is that interpretation is happening cyclically. Of course, there is a hedge in (10), this business of 'evaluation', a notion I'm not familiar with. I don't know if this is meant to distinguish this system from others where LF is, in effect, not a unitary level of representation. For what it's worth, I think a system that does not have to stipulate unitary levels is indeed the simplest. Although to be fair an even simpler one would be a system that stipulates nothing about levels *or* about cyclic interpretation: one that allows you to do interpretation cyclically or non-cyclically, as you please. If that is the simplest option, we'd have to understand why the system has gone with the cyclic alternative -assuming that's the right one. In the MSO model this was because otherwise you could not converge. Here, so far as I can see, it is a desire to unify everything, this idea that 'all operations are

subject to the same conditions'. The latter is a purely methodological point.

On similar grounds, PIC should fall under (10). We therefore restate PIC as (11), for (8) with ZP the least strong phase:

Let me repeat the PIC and (8), or this won't make any sense:

PIC: The domain of H is not accessible to operations outside HP, but only H and its edge [are accessible to operations outside HP],

(8) [_{ZP} Z... [_{HP} _ [H YP]]]

And here's the restatement of PIC, where I italicize the new parts:

(11) The domain of H is not accessible to operations at ZP [*the least strong phase in (8)*], but only H and its edge

We can henceforth restrict attention to phases that are relevant under (10), that is, the strong phases.

That's what you want to take as your main conclusion. The presentation is not felicitous, but the idea is simple enough. This is *less* natural than saying that you just restrict attention to phases period, since obviously a 'strong' phase is more stipulative than a phase.

For the same reason, we restrict attention to v* rather than light verb v generally, unless otherwise indicated.

Considerations of semantic-phonetic integrity, and the systematic consequences of phase-identification, suggest that the general typology should include among phases nominal categories, perhaps other substantive categories.

I'm not sure what 'semantic integrity' refers to, but I assume we're talking about the fact that not just CP or v*P are units for interpretation. Surely an extension beyond seems warranted, but then a serious empirical task remains ahead; for instance, would-be nominal phases have to be special in not allowing A-movement across, and limiting Wh-movement to the spec DP position (Torrego, Longobardi/Giorgi). Same thing with PP's some of which might be phases, while not others (Raposo 1999).

If categorial features are eliminated from roots, then a plausible typology might be that phases are configurations of the form F-XP, where XP is a substantive root projection, its category determined by the functional element F that selects it.

Once we go this far we might assume it is F, and not XP, that determines tokens for the numeration. Hence the numeration would not need indices for lexical items, various token entries of *man* or whatever. If *man* is to be used twice in a derivation we need two D elements. The issue about indices can be raised also about those, but now it is restricted to functional elements. Again, it would seem as if notions such as ‘Case value’ are part of the system in order to distinguish token D’s within a given domain.

CP falls into place as well if T is taken to be a substantive root, as discussed earlier. Phases are then (close to) functionally-headed XPs.

Funny use of ‘headed’, not corrected in the M-version. The intention is clear, the term confusing. I’m not sure what the ‘(close to)’ hedge is supposed to be for. Is the worry that certain functionally-headed XPs are not phases? Are there phases which are not functionally headed?

Like TP, NP cannot be extracted stranding its functional head. The same should be true of other non-phases.²⁵ It seems as if ‘fast’ can adverbially modify into the content item only if you do not separate the light verb *have* from this content item. [see fn. 25] Some phases are strong and others weak; with or without the EPP option respectively, hence relevant or not for Spell-Out and the general principle (10).²⁶

The extension to weak/strong differences is much less natural.

Let us return to (8), repeated as (12) (HP and ZP strong) [let me repeat that HP and ZP strong]:

²⁵ For VP, testable only with a nonaffixal light verb.

In other words, you want to find a language with an auxiliary-like \bar{v} (perhaps serial constructions of the sort studied by Dechaine in Haiti) and see whether you can detach the VP; the prediction is you can’t, just as you can’t do it for TP with respect to C or for NP with respect to D (although in the latter instance you have to say something about ‘peanut’s, I’ve eaten some’; note, though, that you can’t do that with generalized quantifiers: ‘*peanuts I’ve eaten most’, suggesting that weak quantifiers like many might be substantive, perhaps a number category of the Ritter sort, analogous to T. The following suggestion is of course very interesting and worth pursuing:

One might consider treating preposition-stranding within the same context.

Note also: a. He may have a fast dinner [2 readings]

b. What kind of fast dinner may he have? [1 reading: Hotdogs.

*With chopstiks]

²⁶ Head movement aside; see note 23. For conclusions about next-higher-phase evaluation drawn from the theory of control, see Landau (1999).

stage _
 ←-----
 (12) [_{ZP} Z... [_{HP} _ [H YP]]]

Suppose that the computation L, operating cyclically, has completed HP and moves on to a stage _ beyond HP. L can access the edge _ and the head H of HP. But PIC now introduces an important distinction between _ = ZP and _ within ZP, for example _ = TP **[that is, an element within ZP]**. The probe T can access an element of the domain YP of HP; PIC imposes no restriction on this. But with _ = ZP (so that Z = C), the probe Z cannot access the domain YP.²⁷ **Of course the cost is two types of phases now. [this note is crucial]**

A head can probe into the next strong phase down, but not any further.

If Z is C in (12), then its complement TP is immune to extraction to a strong phase beyond CP, and only the edge or head of HP (a strong phase CP or v*P) is accessible for extraction to Z. **[Subjacency]** The same holds for Z = v*, and the observations extend to Agree **[Barriers]**. But T in the domain of Z can agree with an element within its complement, for example, with the in-situ quirky NOM object of its v*P complement.²⁸ **The latter are interesting, suggesting that somehow strong phases can be voided in these kinds of constructions. [This note has interesting empirical consequences.]**

If such ideas prove correct, we have a further sharpening of the choices made by FL within the range of design optimization: the selected conditions reduce computational burden for narrow syntax and phonology, and eliminate distinct LF and phonological cycles. Spell-Out (deletion) is determined quickly, under PIC. The computation is "almost efficient," in something like the sense of Frampton and

²⁷ **The key (which shouldn't be buried in a footnote):**

It is accessibility of the domain of weak phases at _, under (10), that permits the extension of phases to weak phases, yielding the preferred result of relative phonetic-semantic integrity of phases, an extension barred in MI.

²⁸ The consequences include some barrier/relativized minimality-type phenomena (cases of ECP, subjacency, and the Head-Movement Constraint). They extend partially to Huang's CED if phases include DPs, as just suggested. For A'-movement, (3i) is not obviated; thus, PIC prevents access to an inactive subject of CP from the next strong phase, but not from higher T with no intervening strong phase (as in *"T be believed [that John is intelligent]," with Agree(T, John) barred by (3i) but not PIC.

Importantly, these have been shown to be possible in some languages, e.g. the claim exists in the literature for Greek (Ingria, Rivero) and Chinese and other East Asian languages (Xu, Ura), at least.

For A'-movement, questions arise of the kind mentioned earlier, along with others, among them Slavic-style multiple @u<wh>-movement violating subjacency.

Gutmann (1999), with bounded memory load, up to the next phase.

Unification of cycles has consequences for semantic interpretation: no level constructed by the phonological component can yield more than very limited semantic interpretation.

[You may be wondering: Any semantic interpretation? See Part Two.]

Phonological rules typically render semantically significant units unrecognizable: they eliminate "trace" so that chains cannot be reconstructed, blur or remove boundaries between units and change their phonetic content, delete semantic features (which would cause PF crash), etc. Hence displacement rules interspersed in the phonological component should have little semantic effect. We expect (13), if the steps reviewed are on the right track:

(13) Surface semantic effects are restricted to narrow syntax

If for no other reason, this is interesting because there is a class of phenomena with 'semantic flavor' that must involve overt, hence narrow syntax; for instance, topicalization and scrambling. An intriguing issue is why those sorts of elements do not have 'in situ' counterparts, the way quantifiers or Wh-elements do. Possibly, they are 'surface semantic effects', and hence for some reason along the lines just discussed they are restricted to narrow syntax.

As discussed in MI and sources cited, there is mounting evidence that the design of FL reduces computational complexity. That is no a priori requirement, but (if true) an empirical discovery, interesting and unexpected. One indication that it may be true is that principles that introduce computational complexity have repeatedly been shown to be empirically false. One such principle, Procrastinate, is not even formulable if the overt-covert distinction collapses: purely "covert" Agree is just part of the single narrow-syntactic cycle.

As above, Chomsky occasionally starts by saying, 'IF P, then q'. Then he proceeds to introduce further premises based on q or not q, and by the time the presentation is done, readers occasionally forget that this was a conditional. Observe the following sentence:

With the motivation for Procrastinate gone, considerations of efficient computation would lead us to expect something like the opposite: perform computations as quickly as possible, the "earliness principle" of Pesetsky (1989).

The rhetoric here is too strong: we were in the conditional mode. One should not lose perspective of the fact that the existence of QR is still

being debated, and there are various processes that look like purely LF processes (or they seem to involve mechanisms that do not have obvious overt correlates), for instance ‘split’ constructions -NPI, ‘was fur’ split, etc. Second, I am not so sure, either, that I see precisely how ‘considerations of efficient computation’ lead us to expect earliness, any more than they lead us to expect ‘procrastination’. Is the concern here one about computational memory?

Thus, if local (P, G) match and are active, their uninterpretable features must be eliminated at once, as fully as possible; there is no option of partial elimination of features under Match followed by elimination of the residue under more remote Match. In particular, if probe P requires Move (i.e., has an EPP-feature), then the operation must be carried out as quickly as possible. We return to some illustrations. A natural principle, which has been suggested in various forms, is (14)²⁹It seems a bit peculiar to cite a preliminary version of McGinnis (I don’t know if that preliminary version has circulated) when, for instance, a final version of Nunes (1995) had the idea fully developed. [this note is interesting]:

(14) Maximize matching effects

I must confess that I am perplexed by the presentation here. Surely (14) is natural, I have no qualms with that. My problem is relating (14) to the previous sentence. Is this meant as an explanation of the idea that uninterpretable features must be eliminated as soon as possible? But wasn’t that what Earliness was supposed to instantiate? Or is (14) perhaps the current formulation of Earliness? I’ll assume the latter, since that’s the only way I can make sense of what’s being said, although I

²⁹ It follows that the computation cannot ignore matching of (P, G), ultimately crashing because of cyclicity and intervention effects.

In other words, if you can make a match, you must make a match, regardless of the consequences. Importantly:

Not to be confused with such reduction of complexity (very limited, because of PIC) is backtracking of the kind associated with the principle Greed, which allows recovery of a derivation along a more complex path.

To speak of a ‘more complex path’ here can be misleading. I’m not sure I see what is a more complex path in a greedy situation, since the other logically possible paths are not even derivational alternatives, precisely given Greed.

Other false starts are still not barred. E.g., suppose that $_$, T, V are available and the first choice selects T as the complement of $_$. Cyclicity prevents selection of V by T, and the derivation crashes.

So within a phase, and so long as you satisfy (14) and other relevant principles, you can restart as many times as you need to until you obtain a convergent output.

RE:

On maximization principles of the kind (14), see among others a preliminary version of McGinnis (1998).

might be wrong in this assumption.

One concern of MI is to show that central properties of inflectional morphology, displacement and related matters can be accommodated within a framework of the sort just outlined. I will assume that account (modified in accord with revisions here) to be essentially accurate as far as it goes, and turn to some extensions and further modifications.

Consider raising constructions with unaccusatives, abstracting for the moment from English-specific idiosyncracies so that (i) converges as (ii):

- (15) (i) [C [T be likely [EXPL to-arrive a man]]]
(ii) there is likely to arrive a man

The expletive EXPL has the uninterpretable feature [person]. Under local Match, EXPL agrees with T and raises to SPEC-T. The operation deletes the EPP-feature of T and the [person] feature of EXPL, but the $_$ -set of T remains intact because EXPL is incomplete. Therefore Agree holds between the probe T and the more remote goal man, deleting the $_$ -set of T and the structural Case feature of man (assuming, as throughout, the George-Kornfilt (1981) thesis that structural Case is a reflex of agreement)

There's an issue here. It seems as if some forms of agreement (e.g. participial one) do not assign Case, at least not standard Case. So it seems as if it is only personal agreement that is implicated in standard, structural Case. Nonetheless, Chomsky is not going to take this view, see below.

The values assigned under Agree are transmitted to the phonological component: the values of man for the $_$ -set, Nominative for structural Case. Uninterpretable features delete, and the derivation converges as (ii).

Suppose the smallest phase is v^*P , not CP:

- (16) (i) [C [we [@-< v^*P > v^* -expect [EXPL to-arrive a man]]]]
(ii) we expect there to arrive a man

If the derivation is parallel to (15), then Agree holds of (v^* , EXPL), deleting the [person] feature of EXPL but leaving v^* intact so that Agree holds of (v^* , man). The $_$ -set of v^* deletes; structural Case of man is assigned the value Accusative, and deletes.

So far, both derivations are alike. But note one difference:

If v^* lacks an EPP-feature here, then (15) and (16) differ in that there is no raising

to SPEC-v*.³⁰

In (15) and (16) no intervention effect is induced by EXPL. That follows for (15) under the principle (17), conceptually plausible and empirically supported³¹:

(17) Only the head of an A-chain (equivalently, the whole chain) blocks matching under the Minimal Link Condition (MLC)

If you take chains seriously, this makes sense: a chain is a set of phrase-markers, thus a relation. A chain is an n-place operator of categories, each link introducing a new argument. The arguments of a chain are the configurational contexts that define it. For example, in *John was arrested t* the chain is the set {arrested, T'}. The different occurrences of *John* are purely notational, they just signal the various contexts involved in the non-configurational object {arrested, T'}. In that sense, they are like a book-keeping device, like the T-shirts you put on the members of a soccer team to indicate that they play together, as a unit. It makes no sense to say that the center forward won, or the goalie lost; it's teams that win or lose. Similarly, it makes no sense to say, in this view, that a part of chain creates an intervention effect. It's the whole chain that does. A given player may have played lousily, and the team still wins, or the other way around, the team loses with a player playing great; it's a macroscopic event, where the behavior of the whole is what counts.

For (16) the same principle would suffice if raising takes place in ECM constructions.³²

The same results hold for both (15) and (16) if matching observes the maximization principle (14), which entails that the intervention effect is nullified unless intervention blocks remote matching of all features. That is not the case in these examples: the probe (T or v*) matches EXPL in one _-feature, but other _-features of the probe do not match EXPL and are therefore free to seek a goal G, establishing (T, DO) or (v*, DO) match, which value and delete all features of the probe and goal under (14). That makes sense, and the principle is independently plausible. We return to evidence that matching should be construed in this way.

³⁰ A separate question is whether EXPL raises to a position within the v*-complement. See note 14 [in Part 1]. The raising-of-object proposed does not have the syntactic and semantic properties of raising to SPEC-v* (Object Shift), and therefore presumably is raising internal to v*-complement, distinct from the kind of raising in (15). Lasnik and Epstein-Seely take Case of a man to be determined in situ, along the lines of Belletti (1988): an inherent Case in our terms, perhaps partitive, but in any event independent of agreement and distinct from NOM/ACC. This seems questionable, in part for reasons that appear directly.

³¹ MI, (51) and discussion. See further (36), below.

³²32 The observation is irrelevant if Case is assumed to be assigned to the object in situ, as in the proposals cited (see notes 14, 30).

What follows is the key. In this instance (14) and (17) redundantly determine the absence of an intervention effect.

Summarizing, an intervention effect is barred in (15) by principles (14) and (17), and in (16) by (14) and possibly (17), depending on the correct analysis of ECM-constructions.³³

Consider the slightly more complex case of participial passives:

- (18) (a) [C [_ T seem
 [EXPL to have been [_ caught several fish]]]]
 (b) [_ v expect

Again we have double agreement: the probes (T or v) agree with EXPL and fish.

As before, the key:

T deletes the uninterpretable feature of EXPL (and induces raising), and assigns Nominative (NOM) to fish, as in (15); v deletes the uninterpretable feature of EXPL (but without raising to SPEC-v; see note 30), and assigns Accusative (ACC), as in (16).

Now the new element, not present in the examples with *arrive*:

But in (18) there is another possibility beyond (15) and (16): the participle (PRT) could agree with the direct object (DO) fish. Languages differ in this regard: Agreement is manifested in Icelandic but not Mainland Scandinavian (MSc) or Romance.

This is a mistake, removed from the M-version. In both Romance and some mainland Scandinavian languages there is indeed agreement here.

In Icelandic, furthermore, there is also Case agreement: PRT is NOM with probe T, and ACC with probe v.

This is, so far as I can tell, the main reason Chomsky doesn't restrict Case to the personal system: in some languages it manifests itself, indirectly, in participial systems which are not personal. However, we might be clouding things unnecessarily, here. Surely, in many languages adjectival elements manifest Case morphology; you don't have to go to Icelandic participials to see that. Think of extreme instances of Case

³³ Another possibility is that probe-goal match is evaluated first for the most local pair, subjecting deleted features to Spell-Out, then the most local remaining pair, etc. Thus, the first step deletes the single --feature of EXPL rendering it invisible for the next step relating (probe, DO). The proposal would not, however, yield other consequences of (14) and (17).

realization in purely adjectival forms in Latin, for instance: *rosa alba* 'rose-nom white-nom', *rosam albam* 'rose-acc white-acc', *rosae albae* 'rose-gen/dat white-gen/dat'. Are we going to say that we need to check those two separately? That would complicate the system (an unlimited number of adjectives can go with any given noun) and besides, it would miss the generalization that the adjective always gets whatever Case the noun does. So it would seem as if this is an instance of mere concord (as opposed to agreement), where the checking properties of the head noun somehow resonate on the adjective. The issue is whether Case agreement in participials is like that in adjectives or not. So in *Gallia est divisa* 'Gaul-nom is divided-nom' do we have to do some checking with regards to *divisa* or only with regards to *Galia*, the Case of *divisa* being determined by mere concord?

Note that Case-assignment is divorced from movement and reflects standard properties of the probes, indicating that it is a reflex of Agree holding of (probe, goal); the EPP-raising complex is a separate matter. The constructions illustrate both Case assignment without raising to probe and also EPP without Case assignment (namely EXPL in (b)).³⁴

The second conclusion is slightly more controversial; in any case, this is a central topic of this paper, and also of MI.

Under the uniformity principle (1), we conclude that Case is assigned in the same way even where not overtly manifested: structural Case of DO in unaccusative/participial constructions is NOM or ACC depending on the probe (T or \bar{v} , respectively) in Romance, English, MSc, etc., a conclusion with ramified consequences, as recent literature illustrates.

Note by the way that the idea that Case is assigned 'long distance', nominative or accusative depending on the assigner (in the old days a governor, now a probe) was one of the central topics of *Long Distance Case Assignment*, an old Raposo & Uriagereka paper in LI that had many minimalist elements in the late eighties.

Consider more closely the mechanics of (18). The first stage of the cycle that concerns us is \bar{v} articulated further as (19)³⁵:

³⁴ Consider Alternative (II), with EPP a property of the matrix transitive verb (expect in (b)), however the idea is implemented (see notes 14, 30; also 18). The EPP property should be independent of the embedded clause: whether it is transitive (with the subject raising into the expect-phrase) or non-transitive (as in (18)); whether the language always exhibits EXPL overtly (as in English) or only in certain contexts (as in Icelandic). If so, the matrix v^*P in (b) uniformly has an EPP-position, but no Case.

³⁵ To simplify, take PRT to be a light verb distinct from v^* . Here T is \bar{v} -complete, selected by C. If C is missing and T defective, then the same analysis just transfers to the first \bar{v} -complete

(19) [PRT [catch [o several fish]]]

PRT is adjectival: its $_$ -set may therefore consist of (unvalued) number, gender, and Case, but not person. The $_$ -sets of PRT and DO match, inducing Agree. DO is $_$ -complete. Hence for PRT, number and gender receive the values of DO and delete (maximally, under (14)). But Case is unvalued for both PRT and DO, so neither can assign a Case value to the other.³⁶ [important fn.]

To schematize, the probe is not complete, it only has number and gender dimensions. As a result of this, it cannot value features (see footnote). But it can still probe the goal, which is complete, and thus be maximally matched by identical dimensions in the goal, appropriately valued. As a consequence its uninterpretable, matched, number and gender features delete. Nothing happens with the Case dimension, and recall that I suggested this particular dimension might be unnecessary in the personless, participial probe.

PROBE:PRT	...	DO:GOAL
0<--[number].....[Xnumber]		
0<--[gender].....[Ygender]		
[Case]		[Case]

Next we turn to stage $_$ of the cycle. Here again there is double agreement: (probe, EXPL) and (probe, DO). Case of DO is NOM with probe T, and ACC with probe $_$. Probe and DO goal lose uninterpretable features.

That's familiar business; however:

What about PRT? Its $_$ -features are deleted at stage $_$ and should therefore be invisible to Match by the probe. Case of PRT cannot be valued and the derivation crashes, contrary to fact.

To be fair, this is a problem that might arise even if the Case of PRT can be valued and erased by mere concord with whatever Case is on the nominal. The thing is that, at the lower phase, even *the nominal's* Case dimension has not been valued, and thus cannot determine (Case) concord on anything. You have to wait until the next phase, and hope that at that (strong) phase the participial is still accessible to the system, even if its features are not. One possibility is that concord, whatever it is, does not care about $_$ -feature accessibility in the sense

probe. If the proposal of note 17 is tenable, then C is within $_$, extending the parallelism of (a) and (b) of (18).

³⁶ Furthermore, PRT is $_$ -incomplete (hence unable to value features), and both PRT and DO lack the (structural) Case-assigning property of T, $_$.

needed for Agree, perhaps because, unlike Agree, Concord is short distance, a relation among elements in a projection. The other possibility is of course that Concord does require the same sort of accessibility as Agree, in which case the following conclusion would be right regardless of whether the participial has Case determined by concord:

The problem is overcome if Spell-Out takes place at the strong phase level. Then the $\bar{_}$ -features of PRT are still visible at stage $\bar{_}$ of the cycle, though deleted; they disappear at the strong phase level CP or $\bar{_}$ P, as the phase is transmitted to the phonological component.

This is a confusing presentation -the idea that the features of the participial are visible though deleted, and they disappear later. I suppose it means that the deletion operation is really marking something for elimination at the PF interface. In any case, the idea is that from the vantage point of the strong phase, with $\bar{_}$ -features not deleted, there is no issue about visibility, and the relevant Case in the participial can be then valued and consequently deleted.

At stage $\bar{_}$ of the cycle, the $\bar{_}$ -features of PRT are valued by PRT-DO matching, as just discussed. At the next stage, the probe T/ $\bar{_}$ matches the (still visible) goal PRT, valuing its Case feature; and the probe matches the goal DO, valuing the Case feature of DO as well as its own features (since DO is $\bar{_}$ -complete). At the phase level CP- $\bar{_}$ P, the (now valued) uninterpretable features are eliminated from the narrow syntax as the syntactic object is handed over to the phonological component.

The result, then, is that we have triple matching/agreement: (probe, EXPL), (probe, DO), and (probe, PRT). PRT and DO agree with one another: directly for number/gender, indirectly for structural Case (since each agrees with the probe).

So far as I can see this is indistinguishable, empirically, from saying that PRT and DO 'agree' with one another directly for all features: by Agree (in the technical sense) and by Concord, something that will have to be said, anyway, for PRT in adjectival contexts in relevant languages. There is, however, a technical issue that arises: what exactly are the conditions under which PRT and DO establish their putative concord in Case? We don't want to say that these are identical to the conditions of Agree, since we want to separate that relation from mere Concord. Intuitively, it should be the case that Concord is more local, but how do we get the DO to be in a local configuration (adjectival like) with PRT? This, in the languages where it obtains, might relate to another issue which Chomsky's system leaves admittedly unsolved: why, in some languages, does participial agreement correlate with movement of DO? (See fn. 41 of the M-version, fn. 38 below) If concord is implicated, the

answer is trivial: in those languages (where as in Latin there is an extra uninterpretable feature that needs to be checked in PRT) an EPP feature is invoked in PRT *in order to eliminate the relevant uninterpretable feature by Concord*, that is, bringing the element that can enter into the relation to a ‘sister-like’ configuration (see Holmberg 2000 for related ideas concerning the parameterization of PRT phases, to which we return).

Once again there is no intervention effect induced by EXPL, or in this case by PRT either. In the case of (16) there were two possible reasons: principle (17) (assuming raising of EXPL to within matrix VP) or principle (14), which requires maximal (probe, goal) effects. In (18), there is no raising of PRT, so we must resort to principle (14), which is therefore available for (16) as well. The [number/gender] features of the probe bypass EXPL, and its [person] feature bypasses PRT, allowing probe-DO match.

Note that this is the case regardless of whether the Case feature of PRT is valued by Agree or by Concord.

The same reasoning extends to more complex cases, e.g. (20), with probe T or v in a raising or ECM construction:

(20) (i) probe...EXPL...PRT₁ believe t_{EXPL} to have been PRT₂ caught [DO
several fish]

(ii) there were believed (we expected there to have been believed)
to have been caught several fish

Here there is full agreement of PRT₁, PRT₂ and DO for number, gender, and Case; indirectly for Case via agreement with probe. How the features of PRT are manifested phonetically is a language-particular matter.³⁷

True, but this may be setting aside too many interesting issues. So for example, not all languages are alike with regards to agreement with direct objects. Kayne’s examples involve overt movement (languages differ on whether of the clitic or Wh- variety); in some languages (e.g. Catalan, see Cormack (1997)) you have agreement *without movement*. So it seems as if a complete treatment has to be sensitive to various factors. First, it appears that participials are outside of the personal

³⁷ In these terms, we can partially overcome a problem that arises within the LOCUS_{TV} framework (Alternative (II), above), if selection is reduced to Match/Agree. The $\bar{_}$ -features of T_{COMP} and V_{COMP} are unvalued until they function as probes, then deleting. But they must be available to delete the $\bar{_}$ -features of C and v*, which select them. That is unproblematic if Spell-Out takes place at the strong phase level, though another problem remains: the final operation, capturing selection, violates (3i), which may be untenable anyway in this form for reasons to which we return.

system that we see in T-v, which is why normally they do not determine Case properties. Second, nonetheless in some languages participials overtly agree with subjects and, more surprisingly, with direct objects, in the latter instance via movement -though not in all relevant languages, cf. Catalan. Third, it seems, then, that concord has to be implicated, which might allow you to introduce an EPP feature at the participial level, a costly operation (as it has to involve the Move system), but perhaps the only way for some languages to salvage relevant derivations where 'extra' uninterpretable features are involved (e.g. Case in Latin, etc.). At the same time, fourth, you have to allow for the possibility to eliminate certain uninterpretable features without concord, clearly the case in Catalan for number/gender features in participials. That must mean that the inflectional system, although personless in participials, may still be more or less incomplete. Concretely, it would seem that in Catalan it still is capable of probing in terms of number/gender specifications, whereas perhaps in French it is only capable of entering into concord relations, instead of simpler (though morphologically more 'demanding') relations of the Agree sort. That surely must correlate with the decay of the number/gender system of French (gender reduced almost to orthography), in contrast with the extreme vitality of this system in Catalan (where it manifests itself even in Wh-heads *cuales/cualas* 'which-masc.-pl./fem.-pl.'). If we just say that phonetic manifestation is 'a language particular' matter we may be missing all this.

Looking back at the assumptions that enter into this account, we find only one that goes beyond what seems fairly uncontroversial on minimalist grounds: the special role of the strong phase level in narrow-syntactic computation and Spell-Out. We therefore have good additional evidence that phases exist within the framework of cyclic derivation, and that evaluation/interpretation takes place only at the strong phase level.³⁸ I think those problems arise for a solution in terms of Agree or any other transformation, as they have to do with the notion of checking domain and the fact that participial agreement does not go outside its clause. It's not clear to me that these problems would emerge for the local notion of concord.

³⁸ To sharpen the issues one would like to find a language with overt non-clitic expletives, ECM, free use of unaccusatives, and a rich enough manifested morphology to tell whether verb-object agreement holds uniformly between the probe and DO of the lower verb, as in Icelandic.

This relates to much of the discussion I raised in the text, and the following is the commentary that I alluded to.

Among unresolved questions are the reasons for Romance-style participial agreement contingent on movement, as discussed by Kayne (1989) and subsequent work. It is simple enough to state the parameter (spell out $_$ -features of PRT-goal only if probe induces Move), but there should be a more principled account. In any approach in the neighborhood of those surveyed here, movement to SPEC-PRT (specifically, in non-transitive constructions) would be an additional stipulation, hence no improvement over the straightforward description in terms of probe-goal match; the idea faces other problems, noted by Philip Branigan and Dominique Sportiche (see Chomsky 1995, chap. 4, (132)).

[interesting note]

Masterful rhetoric! One could have said: given that the notion, and use we make, of strong phases 'goes beyond what seems fairly uncontroversial on minimalist grounds', this is a stipulation we need to remove from the system, or at least we have to accept that the system presents imperfections. However, we are invited to contemplate a different alternative: since our assumptions are right, it must be that we have found yet 'good additional evidence that phases exist within the framework of cyclic derivation, and that evaluation/interpretation takes place only at the strong phase level'. In other words, outside conditions justify these crazy strong phases; how? Well, somehow. Mind you, the point is consistent with everything else said in the paper, and has a good lesson in presentation: when you are against the ropes, come out hitting as hard as you can.

END OF PART ONE