The Promise of Derivations: Atomic Merge & Multiple Spell-Out[φ]

John Edward Drury
University of Maryland, College Park
jdrury@wam.umd.edu

Introduction
This paper pursues three central theoretical themes in service of suggesting an answer to the following question: what counts as motivation for a derivational theory of syntax? The three central themes are these: (i) the *direction* of derivation matters, (ii) derivations are partitioned into sub-derivations that only interact with each other as *units*, and (iii) precedence relations follow from the notion of a serial derivation and command can and should be understood as a derivational subcase of precedence. As these themes unfold, I argue that (i) and (ii) lead us to a sensible perspective on the main question and point towards heading the direction of the third theme (iii).

In what follows, a novel combination and elaboration of existing proposals for derivational systems is introduced based on a pair of primitive operations—Atomic Merge and Spell-Out—and its properties are investigated in (re)analyses of parasitic gap constructions and control into adjuncts. Although the purpose of this empirical component is aimed more towards demonstrating how the proposed system works than anything else, the choice to expend the analytical energy of this paper on these phenomena is not arbitrary. Recent analyses of these constructions has argued that they involve so-called Sidewards Movement (Nunes 1995; Hornstein 1996, 1998).[1] These are movement operations occurring at some intermediate derivational stage which relate positions contained in separate, unconnected subtrees which are later associated to form a single-rooted object. The effect of such operations is movement-established relations between positions that are not in a c-command configuration in the output. The technical innovation is to allow movement to occur *before* these positions are dominated by the same root (contained in the same phrase-marker).

This paper suggests that these proposals are correct in spirit but not in detail. Roughly, it is counter-proposed here that while the kind of relations that Sidewards Movement has been suggested to mediate *are* in fact movement relations, there is never movement between complex, unconnected subtrees. The reason in this paper's system is simple—there are *never* intermediate steps in a derivation that contain unassociated complex subtrees. That is, the 'structural description' for Sidewards Movement is never met. The analysis offered here is argued to avoid the postulation of additional constraints that are demonstrated to be necessary to constrain the Sidewards Movement analyses.

However, all of this will aim to speak to the larger issue of what sorts of considerations motivate a derivational system. After all, we *might* consider the representational view to the default—at the end of the day, we're all worried about generating well- versus ill-formed output structures. And, if there is an intuition that has had a grip on generative syntax since its inception, it is the centrality of *sentences* (or rather, the structural descriptions corresponding to them); the *outputs*; the point at which a derivation is "finished". Viewing a derivation as simply a *sequence of representations*, where each is related to the next/last in some law-like way, underscores the obscurity of the derivational/representational distinction (henceforth: D/R-distinction). To the extent that each of the step-wise, locally established structures are presein the last member of such sequences, we should ask whether or not we

---

[φ] This essay is the theoretical component of work in progress. The core proposals were presented at the Fall 1997 UMCP Linguistics Student Conference and again in my generals paper (Drury 1998a). I would like to warmly thank and acknowledge the following for extensive discussion and criticism of this paper and the big project behind it: Mark Arnold, Juan Carlos Castillo, Kleanthes K. Grohmann, Norbert Hornstein, Howard Lasnik, Philip Resnik, Juan Uriagereka, and Amy Weinberg. My committee (Amy, Juan, and Norbert) for the generals defense deserves special mention for their encouragement and involvement. I have also benefited from conversations about this material with Aitziber Atutxa, Mathew Kaiser, Jeff Lilly, Colin Phillips, and Acrisio Pires. Special thanks to both Kleanthes K. Grohmann and Werner Abraham for their interest in this manuscript. Comments and criticism are more than welcome. JED.

[1] Additionally, these operations have been argued to render head movement (and adjunction generally) consistent with Chomsky's Extension Condition (Bobaljik & Brown 1997, Uriagereka 1996, 1998). I will only very briefly touch on these issues in what follows. The label "interarboreal operation" comes from Bobaljik & Brown.

need to appeal to these sequences as opposed to directly stating our principles over the outputs.[2] However, the D/R-distinction has an inherent, slippery kind of obscurity that has (judging from the literature) allowed it to alternate between a formal triviality and an issue of crucial theoretical import. The fact is that it simply isn't clear whether or not this distinction even *could* be important. As Chomsky (1998:p8) puts it:

> it is unclear whether or not these are real questions; on the surface they look like the question whether $25 = 5^2$ or $5 = \sqrt{25}$. If the questions are real, they are subtle. [...] The apparent alternatives seem to be mostly intertranslateable, and it is not easy to tease out empirical differences, if there really are any.

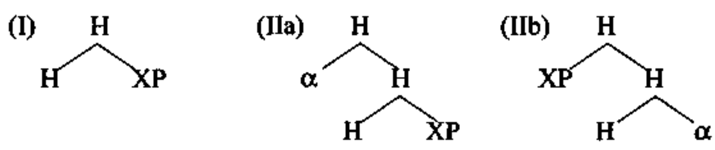
Imagine that someone were to tell you that it matters to the enterprise of, say, correctly characterizing all and only well-formed triangles, whether this is accomplished by listing properties/direct definition (e.g. sum of all angels must equal 180°) versus some sequence of (e.g. arc-drawing) instructions guaranteed to generate such objects. This, one must admit, is not an immediately inspiring distinction. But since this is, nonetheless, what this project is about, let me say a word or two about why this is worth doing. The interest of this issue will improve as we proceed.

Our starting point is a methodological observation—that the *potential* importance of the D/R-distinction is likely to remain obscure if theoreticians freely draw from both kinds of systems. The point is simple: hybrid approaches can never tell us what we lose/gain by going one way or the other. Further, in practice, these mixed systems have presented undesirable redundancies.[3] For instance, consider a recent example from the literature which illustrates both the sort of redundancies I'm referring to *and* what this paper takes to be the *right* research strategy. If one believes that grammars which deploy both movement operations and chain conditions are redundant (as do both Brody 1995, 1998 and Hornstein 1996, 1998) then one should attempt to formulate a theory that makes use of just one or the other set of devices. Brody (1995:p8) writes:

> we have an argument from conceptual economy against a theory that makes use of both concepts. If Move α is independently motivated, then the theory that uses chains to capture antecedent-trace relationships is wrong; if a principle like Form Chain is independently necessary, then a theory incorporating Move α must be mistaken.

He goes on to take the representationalist route and dispenses with movement operations; Hornstein, on the other hand, argues for the elimination of chains in favor of a derivational Copy + Merge formulation. Although we will see below that the D/R-distinction isn't as cleanly represented by this debate as it might seem at first blush, for now it is enough to point out that this is the correct route into the discussion—with evidence and argument that crucially hinge on *one or the other* of the extreme, endpoint positions. This paper (initially) takes a more indirect route, beginning with a technical question internal to the derivational perspective.

The technical question that sets this investigation in motion is raised in a recent manuscript by Chomsky (1998:p27). He considers the following object (I) and two logically possible ways an item α could Merge to it (in IIa and IIb):



About these two options he has this to say:

---

[2] This is not the same as the issue of how many levels of representation there should be. There can obviously be multiple levels of representation without any crucial ordering. The question here is whether there is any necessary directionality to the mapping.

[3] Although as Brody (1998:367, fn.1) points out:

> The choice between a purely representational and a purely derivational theory is of course a more difficult matter. Note that we cannot in principle exclude mixed theories either, when these succeed in avoiding the unwanted redundancies and duplications.

The point of this paper is not that hybrid approaches should be excluded in principle, but that given the mission of attempting to uncover places—either conceptually or empirically—where the D/R-distinction makes a difference, hybrid approaches are likely to be unhelpful.

Plainly, the desired outcome is [IIa], not [IIb]; that has always been assumed without discussion. But the reasons are not immediately obvious. Each outcome satisfies a natural condition: [IIa] satisfies the familiar ***Extension Condition***; [IIb] satisfies the condition of ***Local Merge***, which requires the merged element α to be as close to the attractor as possible. A reasonable basis for choosing [IIa] over [IIb] is that [IIb] tampers with the fundamental relations involving the label of the target category β, while [IIa] leaves them intact; in fact, in [IIb] the category $\beta = \{XP, H\}$ disappears altogether. Suppose, then, that we make explicit this "least tampering" condition:

> [*LEAST TAMPERING*]
> An operation cannot change relations involving labels.

There are two such relations: sisterhood and c-command...Operation [IIa] preserves sisterhood and c-command for {H, XP}; operation [IIb] modifies both (Chomsky 1998:p 27) [emphasis mine-JED].

As indicated in the passage, it is the Extension Condition which has, in one or another formulation, been assumed in almost all work under the umbrella of "minimalist" investigations.[4] There is (to my knowledge) only one recent place where the mechanics illustrated by [IIb] have been seriously explored in an investigation of syntactic competence—Phillips (1995, 1996).[5] This paper traces the paths that lead from these two starting assumptions, examines the kinds of systems that follow from each, and adopts a formulation which takes the mechanics proposed by Phillips as a point of departure. The purpose of this comparative process is to suggest a positive answer to a technical question which bears suggestively on the D/R-distinction. The technical question is this: ***does the direction of derivations matter***? I argue that the decision about whether to exploit a derivational versus a representational theory turns on *exactly the sort of distinction* that distinguishes between systems governed by Extension versus those governed Local Merge.

For example, Chomsky suggests a motivation for embracing Extension—the absence of 'tampering' properties which follow from Local Merge.[6] I argue the opposite—that it is *exactly these kinds of properties* that allow us to see content in the D/R-distinction, and that allow us to exploit the formal dimension (a sequence of representations) made available by the derivational view in ways that don't have obvious translation into representational accounts. I will suggest that derivational systems that *don't* exhibit such tampering properties really are quite indistinguishable from corresponding representational approaches.

Here is the plan. The §1 introduces and investigates some of the properties of the mechanics of this paper's proposed system—the Atomic Merge/Multiple Spell-Out system

---

[4] Other sorts of conditions have been introduced with essentially equivalent effects—e.g. the Strong Feature Condition of Chomsky (1995:chap4), or Watanabe's (1995) Avoid Redefinition (see also Kitahara 1997). Various issues have arisen regarding the status of this condition—for example, whether head movement (and adjunction generally) obey it; what the status of covert/LF movement is with respect to this condition, etc.. This harkens back to early discussions about cyclicity (see e.g. Freidin 1978 among others). These issues are reviewed and discussed in ongoing joint work (Castillo, Drury, & Grohmann, to appear).

[5] Phillips identifies his derivational system with the (syntactic sub-component of) on-line processing mechanisms (the "Parser is the Grammar" thesis). I mention his work exclusively here since he argues in a purely *competence* mode for his proposed derivations, and then offeres independent *performance* arguments that the parser has roughly the same 'shape'. Grammaticality, under this view is simply "parsability in the limit.". Also, Weinberg (1997) is an effort to formulate what a 'minimalist' parser would look like, but without the commitments to the parser-grammar identity argued for by Phillips. This work has been extended in Weinberg (forthcoming) to include some of the specific mechanism introduced here and in earlier versions of this material (Drury 1997) in order to account for various garden-path phenomena.

[6] Curiously, other sorts of tampering properties are alluded to by Chomsky (1995:223) as conceptual motivation for derivations:

> Viewed derivationally, computation typically involves simple steps expressible in terms of natural relations and properties, with the context that makes them natural 'wiped-out' by later operations, hence not visible in the representations to which the derivation converges. Thus, in syntax, crucial relations are typically local, but a sequence of operations may yield a representation in which the locality is obscured.

Chomsky goes on to use head movement as an example—a case where superficially 'long-distance-looking' relations are derivationally factored into a series of local steps. But what is crucial about this story is the *locality*, not the notion of a *sequence of operations*. Lasnik (1997:15-18) quotes this same passage from Chomsky and points out that once a Copy Theory of movement is adopted (as Chomsky does) then its simply not obvious that the intermediate steps of head movement would be 'wiped-out' thereby obscuring the locality. This appears to be a case where there is a relatively straightforward translation into a representational system. To be fair, Chomsky considers the possibility of a representational re-coding a page later, but argues that in cases of movement with multiple steps, only the head and the tail of the resulting object are relevant. Although as Lasnik points out (p18) this depends on the status of intermediate traces/copies left by movement operations. If these are motivated for interpretation (or for some other reason) then the objection stands (i.e. that the derivational and representational views on local movement of heads are intertranslateable) otherwise not. I refer the reader to Lasnik's discussion of these matters. We will revisit the issue of derivational 'tampering' throughout this paper.

(henceforth: AM/MSO)—in the context of discussions of *order* in phrase structure and of the treatment of movement/chain relations. It is argued that command relations can and should be derived as a subcase of precedence relations whereas the latter are associated with the very notion of a serial derivation. 'Movement' is proposed to be understood as simply the (re)Merge of an item to a different position (creating 'loops' in the structure). The issue of where a displaced item is 'pronounced' is pinned on the operation Spell-Out in a way that dispenses with typically assumed operations like Copy, Delete, and Form Chain. In addition, all relations are argued to require licensing in sisterhood configurations, and some basic assumptions about clause structure are introduced. Also of interest will be a difference between the AM/MSO system and 'bottom-up', Extension-regulated systems—the latter but not the former *requires* intermediate derivational steps which contain complex, unconnected phrase-markers. As mentioned already, recent analyses of parasitic gaps and adjunct control capitalize on such intermediate steps and propose them as the structural description for a kind of 'tree-to-tree' movement. These analyses are reviewed in §2 and some conceptual and empirical problems are raised; then §3 deploys the AM/MSO system in a reanalysis, including an account of some recalcitrant cases introduced as problematic for the Sidewards Movement analyses in §2. The final section summarizes the results, coming back to the 'big' questions regarding the D/R-distinction. Some possible extensions of the system are then offered, as well as some suggested directions for future research.

1 Atomic Merge and Multiple Spell-Out

The structure building operation of the grammar that is posited here is a version of the Merge operation of Chomsky's (1995) Bare Phrase Structure (BPS). The goal of BPS is to eliminate X-bar theory in favor of a sparse, set-theoretic, 'virtually conceptually necessary' underpinning. Chomsky assumes a lexicon and a computational system. The lexicon is 'accessed' by the computational system via a *numeration*—formally a multi-set (i.e. a set with members indexed with integers indicating multiple tokens) of lexical items—which is essentially a convenient reference to the subset of the lexicon implicated in particular derivations.[7] There are two "free" operations: Select and Merge. Select removes a lexical item from the numeration (reducing the item's index by 1). Merge associates two such lexicon objects to create a third, derived object. Consider the simplest instance of the Merge operation—the association of two lexicon items α and β, to create a third derived object γ (the label is de-emphasized):

(1) Select(α), Select(β), Merge (α, β)

$$\alpha \Leftarrow\underline{\text{MERGE}}\Rightarrow \beta$$
$$\Downarrow$$
$$\{\gamma, \{\alpha, \beta\}\}$$

In a system that obeys the Extension Condition, further items to be added to the structure necessarily target the entire object that has been generated up to that point (i.e. target the root of the phrase-marker) as in (2), where δ is added creating a new object labeled η (again, the labels are de-emphasized):

(2) Select(δ), Merge(δ,{γ, {α, β}})

$$\delta \Leftarrow\underline{\text{MERGE}}\Rightarrow \{\gamma, \{\alpha, \beta\}\}$$
$$\Downarrow$$
$$\{\eta,\{\delta,\{\gamma, \{\alpha, \beta\}\}\}\}$$

Given this notation, which represents the labels of the derived objects (above: γ, η) as members of the generated sets, Chomsky introduces a definition for the notion of syntactic object or *term*. These are the objects that the syntax manipulates:

TERM: For any structure K:
(i) K is a term of K (the entire set is a term), and
(ii) If L is a term of K, then the members of the members of L are terms of K.

[7] Although it was in fact introduced to paly a substantive role in the theory of economy; i.e. to delimit comparison sets over which transderivational economy metrics are stated. This is not important for the moment, but I will return to this in §2 in the context of criticism of Sidewards/Interarboreal movement analyses in a discussion of some technical problems that arise for Chomsky's version of the theory of economy.

Thus, for the object generated by the merge illustrated in (2), the terms are:

(3) (a)	$\{\eta, \{\delta, \{\gamma, \{\alpha, \beta\}\}\}\}$	the entire set
(b)	δ	member of a member of (a)
(c)	$\{\gamma, \{\alpha, \beta\}\}$	member of a member of (a)
(d)	α	member of a member of (c)
(e)	β	member of a member of (c)

I propose to limit the Merge operation to its simplest instance, illustrated in (1) above. This will drive the mechanics of what Chomsky refers to as Local Merge (along with its characteristic 'tampering' properties). Call this operation Atomic Merge:

ATOMIC MERGE: Merge combines only *simple* (memberless) terms.

To illustrate its application, consider the derivation for the same object generated in (1) & (2):

(4) (a) Select(δ), Select(α), Merge(δ, α)

$$\delta \Leftarrow\underline{\text{MERGE}}\Rightarrow \alpha$$
$$\Downarrow$$
$$\{\eta, \{\delta, \alpha\}\}$$

(b) Select(β), Merge(α,β)

$$\{\eta, \{\delta, \alpha \Leftarrow\underline{\text{MERGE}}\Rightarrow \beta\}\}$$
$$\Downarrow$$
$$\{\eta, \{\delta, \{\gamma, \{\alpha, \beta\}\}\}\}$$

There are two things to note about (4). First, the arguments of the Merge operation are always *individuals* (i.e. they are *memberless/simple/atomic*). Second, the Merge operation, in virtue of being required to target material that is already associated, operates 'within' the object already formed, *changing* the relations established by the previous step. In (4a) δ and α are sisters/form a unit, while in (4b) they do not so form a unit. Thus Atomic Merge, as in Chomsky's 'Local Merge', is allowed to 'tamper' to basic relations. We turn now to the question of *order* in such objects.

1.1 Phrase Structure and Order Atomic Merge introduces only two basic relations: dominance (or containment) and sisterhood. This subsection discusses two other relations: precedence and command. The view that I adopt is a somewhat radical ***derivational*** interpretation of what I take to be the intuition behind Kayne's Linear Correspondence Axiom (LCA)—that hierarchical structure "inherits" its asymmetries from the "more basic" linear ordering of terminal items. The derivational interpretation of this idea here is that the ordering of terminal items is pinned squarely on the asymmetry inherent in the very notion of a serial derivation, and, abstractly, is seen to be "ultimately related to the asymmetry of time".[8] This intuition that I will exploit—that precedence is 'prior'—comes through in Kayne's *The Antisymmetry of Syntax* (henceforth: AS) in several places. Consider:

> The [LCA]...explains certain basic properties of phrase structure that standard X-bar theory has not...it does so by in essence attributing certain properties of linear order to hierarchical structure, in effect taking linear order to be of more fundamental importance to the human language faculty than is generally assumed (AS:p xv).

> X-bar theory is not a primitive component of UG...[it] expresses a set of anti-symmetric properties of phrase structure...[which]...will be seen to be inherited, in effect, from the ***more basic*** anti-symmetry of linear order (AS:p3 [emphasis mine—JED]).

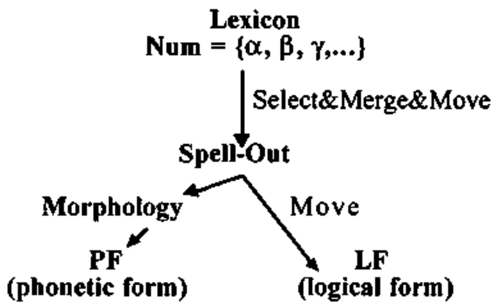
> This S-H-C property of UG, as well as the fact that UG does not make both orders available, is thus seen to be ***ultimately related to the asymmetry of time*** (AS:p38 [emphasis mine—JED]).

> From this perspective, linear order turns out to be more fundamental to syntax than is normally thought. As a result of the LCA, the property of antisymmetry that a linear ordering has ***is inherited by the hierarchical structure*** (AS: 131 [emphasis mine—JED]).

[8] Note that this is the opposite of Chomsky's interpretation of Kayne's LCA (i.e. to be a PF condition/bare output condition).

The LCA insists on a total/linear c-command ordering of terminals. Kayne proposes in addition (AS:p33-38) to interpret this order as mapping directly to precedence relations. This is thus a *formal* requirement on phrase-structure, mandating that structures have a certain 'shape'. In fact, Kayne attempts to derive commonly stipulated properties of X-bar theory from this axiom. I mention this because current work within the Minimalist Program has imported the LCA with a much different interpretation. Chomsky, for example, views the properties of phrase structure as following from the nature of the 'virtually conceptually necessary' properties of the basic Merge operation (plus some assumptions about the nature of syntactic categories—see Chomsky 1995:241-249). The resulting BPS theory doesn't have many of the properties that Kayne views phrase structure to have in virtue of the LCA (e.g. vacuous projections; category/terminal distinction, etc.). Chomsky thus argues that the LCA exists in the grammar as a 'bare output condition'. Consider in (5) the architecture of the system under current minimalist assumptions:

(5)



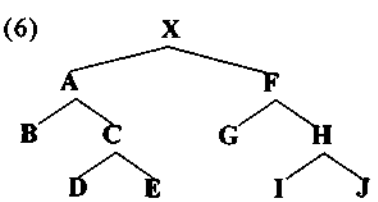
There are two places that conditions/principles can come from under current 'minimalist' assumptions: from either 'virtual conceptual necessity' or from 'bare output/interface conditions'. Thus, things happen in the system either to meet some requirement of one of the two levels of representation, or else they must follow from the indispensable bare mechanics. With respect to the question of command-precedence relations, Chomsky proposes to view the need to establish a linear ordering of terminals as a PF-condition. The way that this output requirement is met, he proposes, is by recruiting the command relations that independently exist in phrase markers. This 'linearization' process is taken to be "a principle of the phonological component that applies to the output of Morphology" (Chomsky 1995:p340).

The intuition here was (to my knowledge) first expressed by Higginbotham (1983), and is discussed by Chametzky (1996). Chametzky writes that, "its hard to see what, other than the facts of speech, would lead to the postulation of a precedence relation" (p11).[9] He quotes a passage from Higginbotham (1983:p150-151) who argues that

> the notion of precedence is to reflect the ordering of formatives in speech...[and]...that they will be so ordered is a consequence of the application of the laws of physics to the human mouth...reflecting the physics of speech (quoted in Chametzky 1996:p14).

This is essentially the intuition that Chomsky is pursuing—that precedence relations are a PF-interface requirement. Note that in order for this to work the command relation must provide a linear ordering of the terminal elements. But this hierarchical ordering doesn't (and in fact, *can't*) follow from the LCA within Chomsky's assumptions. Consider the case of a complex specifier and the structure it is associated to:

(6)

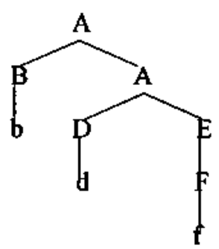


[9] Chametzky argues that there are no basic precedence relations. I will return to this below.

Here, nodes A and F are in a symmetric c-command configuration, so how can the nodes that A and F dominate be linearized with respect to each other? Chomsky blames the required asymmetry on a property he attributes to intermediate (non-X^0, non-X^{max}) projections—according to Chomsky's definitions they simply cannot enter into command relations (Chomsky 1995:p339-340). Thus, if the subtree dominated by A (above) is a specifier, A will asymmetrically c-command into the subtree dominated by F, but F (being an intermediate projection) will not c-command into the subtree dominated by A. Strangely this means that F must be relevant to calculating command relations *for the nodes it dominates*—if it were simply 'invisible' to all aspects of calculating command relations then X would be the first branching category for the nodes F dominates, and the asymmetry Chomsky is attempting to build-in dissolves immediately. So intermediate projections, presumably in virtue of their 'intermediate' relational phrase structure status, also enjoy 'intermediate' status with respect to command—they cannot themselves be command*ers*, but they must crucially be involved in establishing command relations between *other* elements.

These issues simply do not arise for Kayne. Since for him the LCA mandates that phrase structure must have certain properties, he takes his axiom to insist that all specifiers are in fact adjuncts—that is, dominated by a segment of the category that they adjoin to and not by the category itself. His assumed definition of c-command, which makes use of the category/segment distinction, is crucial. For example, consider (7):

(7)



Here B is a singleton specifier, argued by Kayne to be an adjunct (thus the two segment category A). He defines c-command as follows:

C-Command (AS:p 16): X c-commands Y iff:
- X and Y *are categories;*
- *and* X excludes Y;
- *and* every category that dominates X dominates Y

Thus in (7), B will c-command what A dominates, *but not vice-versa*—nothing that the (two segment) category A dominates can c-command B since A does not dominate B in virtue of the definition of exclusion assumed (from May 1985; Chomsky 1986):

Exclusion: X excludes Y if no segment of X dominates Y

The conclusion is that no matter how complex the specifier B, everything that it dominates will asymmetrically c-command what A dominates.

It is worth noting that both Kayne and Chomsky are forced, in order to exploit the intuition of the LCA, to build-in asymmetries into the hierarchical structures. For Kayne, this is principled since his formulation takes the LCA to be a condition on phrase-markers. For Chomsky, it is less clear—one might question the strange role that intermediate projections are being asked to play. And, for both, their definitions of c-command are absolutely crucial. The fair question to ask at this point is, "why c-command?" Suppose we were interested in engaging in the task of answering this question. For both Kayne and Chomsky, this explanatory enterprise of attempting to understand what c-command is and why its an important structural relation in natural language is, in a way, preempted. The answer to the question "why c-command" (or better, "why c-command in terms of exclusion?" or "why c-command that ignores intermediate projections?") has to be "because it doesn't work otherwise".

There are however, attempts in the literature to understand what command relations are and why they should be important. I know of four.[10] There is the earlier effort by Kayne (1981, 1984), and three recent ones: Chametzky (1996), Epstein (1996), and Chomsky (1998). The latter two are within the assumptions of BPS. I will omit here discussion of Kayne's earlier work on "unambiguous paths" and focus on the other three.

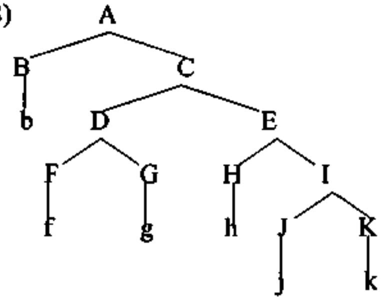
1.1.1 Generalized Sisterhood and Minimal Factorization Chomsky (1998) notes that the basic Merge operation makes available two central notions: containment (or dominance) and sisterhood. Given these, c-command is available under "very weak assumptions" if, for example, we are permitted to compose relations. This yields two: *contain(sister)* and *sister(contain).* The former is the transitive closure of the containment relation[11] and the latter is the familiar c-command relation.

Chametzky (1996:chap 2) makes a related point, observing that command is really a "generalization of the sisterhood relation" and advances a perspective on c-command (from Richardson & Chametzky 1985) inverting the standard understanding of the relation which takes the 'point-of-view' of the command*er.*[12] Instead, they define the relation from the perspective of the command*ee*. In addition, the relation is not viewed as one holding between two nodes, but rather *between a node and a set of nodes*, as follows (p27):

> **C-Command:**
> For any node α, the c-commanders of α are all the sisters of every node which dominates α.

Chametzky's formulation of phrase-structure axioms—part of what he dubs 'Minimal Phrase Structure Theory' (MPST)—includes only dominance as the sole ordering relation of phrase-marker nodes. Chametzky notes since there (empirically) seems to be the need for some non-dominance and non-sister relations, something else is required. The most natural 'something else', he argues, is c-command. The reasoning is that from the point-of-view of the command*ee* we can see why c-command in particular is the 'most natural'. For any node X, the set of nodes that c-command X provides *the minimal factorization* of the phrase-marker ("an exhaustive, non-overlapping constituent analysis of a PM"—MPST: p175, n.2). For a quick illustration, I reproduce Chametzky's example below (MPST:p30) in (8):

(8)



The minimal factorization "of the PM...with respect to node G" is the set {B, F, E}. This is so because:

> there is no other set of nodes that, when unioned with the set {G}, is both smaller than the set in question (has fewer members/a smaller cardinality) and also offers a complete, nonredundant constituent analysis of the structure (MPST:p. 29).

---

[10] There is also Barker & Pullum (1990). Their paper is discussed by Chamtezky (1996) who points out that their enterprise is more towards the question of "what is c-command", as opposed to the explanatory issue of "why c-command". B&P formalize a family of command relations in a survey of various formulations that have been proposed in the literature. See Chametzky (1996) for discussion.

[11] Perhaps a way to make a version of 'government' available "under very weak assumptions"?

[12] This exploits what was perhaps the origninal intution behind Klima's (1964) notion of "in construction with", which, as noted by Barker & Pullum (1990:p13-14), is essentially the inverse of Reinhart's (1974) notion of "superiority" (i.e. "c-command"). Klima's notion is reproduced by B&P as follows:

> A constituent is 'in construction with' another constituent if the former is dominated by the first branching node that dominates the latter.

Although it turns out that 'in construction with' is actually not a command relation according to B&P's criteria. See B&P for discussion.

For another example, the minimal factorization with respect to node J would then be {B, D, H, K}—the c-commanders of J. In §1.4 below, we combine this insight with both the (somewhat radical) derivational interpretation of Kayne's LCA and the strategy Epstein (1996) takes, to which I turn directly.

1.1.2 Piggy-Backing on the Asymmetry of "Serial Derivation" Questioning the status of c-command leads Epstein (1996) to a cluster of important sub-questions:

- Why does command exist at all? Why doesn't a enter relations with all (or any) constituents in the tree?
- Why is the *first* branching nodes relevant? Why not: "the first or second or third (*n*th?) node dominating a must dominate β?"
- Why is *branching* relevant?
- Why doesn't α c-command the *first branching* node dominating α, but instead c-commands only categories *dominated* by the first branching node?
- Why must α not dominate β?
- Why must α not equal β?

Epstein suggests that this structural relation can be deduced from the derivational mechanics Chomsky's Bare Phrase Structure in a way that provides principled answers to (I-VI). The idea exploits a redundancy that Epstein observes between the (representational) definition of c-command and the (derivational) MERGE. C-command, he argues, can be understood as follows:

> **Derivational C-Command:**
> α c-commands all and only the terms of the category β with which α was paired/concatenated by Merge or by Move in the course of the derivation.

The idea is that command relations are derivationally established—items command one another just in case they were merged with one another. This is the answer to Epstein's first question (why c-command?). The second question (as does the answers to the rest) is thus claimed to follow from the nature of Merge and Move—what these operations do is pair two items and project the features of one of them. Since these operations are binary, the object that their concatenation produces can be viewed tree-theoretically as a binary branching node dominating the two items that merged to produce it. Thus *it could only be* the '1st node up' that is relevant for the command relation. Further, branching is relevant (third question) by assumption—BPS stipulates that there is no vacuous projection—structure beyond the lexicon is only created by merging items. Regarding the fourth and fifth questions, nodes don't command what dominates them (or what they dominate) because they were not paired (by Merge) with what dominates them (or what they dominate). The last question follows from the assumption that Merge is not reflexive—items don't merge with themselves—so α can't equal β.[13] Thus Epstein's investigation has the effect of re-locating the (representational) properties of c-command by blaming them on some 'bare' assumptions about the nature of an operation no theory can do without—Merge.

But are the crucial properties actually all *derived*? Consider Epstein's story for why, when α c-commands β, α also c-commands what β dominates—*but not vice-versa*. The crucial part of his definition is that *α c-commands all and only **the terms** of the category β*. So here is the question: Why?

Here is the logic of Epstein's argument. He is relying on the assumptions of Chomsky's BPS—specifically, the idea that "terms correspond to nodes of the informal representations, where each node is understood to stand for the subtree of which it is the root." (Chomsky 1994, 1995). Consider the Merge of α and β where β is an already assembled complex subtree. The idea, according to the definition, is that when α and β are merged, α not only c-commands β (the root) but everything contained in β (all the terms of β). The intuition is that because β is complex, and because the root 'stands for' the entire subtree, when α and β are associated by Merge, α is, at that derivational 'moment', also so associated with each and every subpart of β. But α and the root of the complex subtree β stand in a *symmetrical* c-command relation. If the terms of β are c-commanded by α in virtue of the Merge operation,

[13] But see Guimarães (1997) on the notion of 'Self-Merge' and vacuous projections in BPS.

why don't the terms of β also c-command α as the root of β does? Put another way, ***where does asymmetric c-command come from***? There is nothing in Epstein's system that *derives* the fact that the symmetry of c-command is restricted *to just the root* of a complex subtree that is merged with another item. That c-command can be an asymmetric relation remains a stipulated property.

A quick note on Epstein's strategy—there is a presupposition here that we should subject to some scrutiny. There is a tendency in Epstein's presentation to freely exchange "stipulative" and "definitional" for "representational" while "derivational" seems to frequently be going for "explanatory". Given the asymmetry which is defined-in to this derivational definition, nothing in this discussion would change if we exchanged the derivational *operation* Merge for the (representational) sisterhood *relation*. Under such a view, which are the command relations? Answer: α asymmetrically c-commands its sister's descendants. Why don't α's sister's descendants command α? Answer: It's an axiom. Command is asymmetrical.

This is a good time to make a general point: there is nothing *inherently* more explanatory about having a derivational coding of a representational relation. By itself, saying that a relation 'has been reduced to' an instance of an operation isn't saying much if the latter simply duplicates or 'generates' the former. A 'logical process' doesn't provide us any new information if the (local) outputs of the 'process' are straightforwardly maintained in the eventual (global) output. The *potential* interest in derivations is the possibility of creating and *then changing* relations from one member of a sequence of representations to the next. Otherwise derivationally understood primitives have an obvious representational coding and this makes it hard to see why we shouldn't just appeal to conditions on outputs (i.e. on the final members of a sequence of representations/outputs of derivations). Remember what is at stake. If relations that are established by operation are 'set in stone' to never be tampered with again, then these will exist in the output representation and thus be characterizable over *those* 'endpoint' structural descriptions, and the need for *sequences* of representations dissolves.

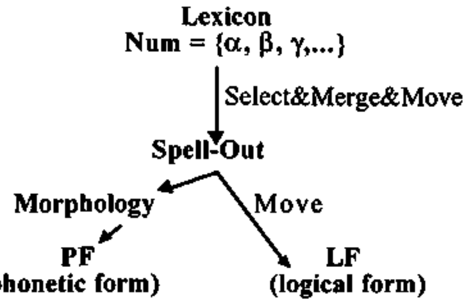
Nonetheless, Epstein's strategy is *extremely* attractive—pinning an ordering relation in the grammar on the ordering that comes 'for free' given the adoption of serial derivations. We have already seen that asymmetries have been 'built-in' in various ways in various sets of proposals. I take the fundamental and important insight of Epstein's paper to be that structural asymmetries *might be* pinned on the asymmetry inherent in the 'logical time' of a derivational approach. This is exactly the insight that I wish to pursue. The answer to both the question of why both precedence and command have the property of asymmetry is then: because derivations have that property. A crucial component here will be Uriagereka's (1996) Multiple Spell-Out proposal, to which we turn next.

1.2 Locally Total is Sufficient (removing the last residue of S-Structure) Uriagereka's (1996, 1997, 1998) project is essentially to maintain a coherent combination of Kayne, Chomsky, and Epstein's views on hierarchy and order. Recall from above that for Kayne and Chomsky it was important to build asymmetry into phrase-markers given their respective definitions of c-command—either via the notion of specifiers as adjuncts or via the stipulation that intermediate bar-levels cannot command. Now juxtapose this with the agenda of Epstein who, contra Kayne, sees command as holding in phrase-markers in virtue of the nature of the process that generates them—*not* in virtue of the axiomatic need to establish linear order (i.e. not in virtue of the LCA). So here is where we are:

- **Command is a reflex of Merge (Epstein)**
- **Linearization of terminals (establishment of precedence relations) is a grammar-external/bare output requirement (Chomsky)**
- **The *way* linearization is established is via asymmetric c-command (Kayne)**

Uriagereka suggests a minimalist way to retain the central insight of Kayne within BPS—by removing the stipulation that Spell-Out applies only once in a derivation. Consider again the standard minimalist view of the grammar:

(9)



Although not discussed thus far, the assumption of the system as presented in Chomsky (1994, 1995) is that the system selects and merges items from the numeration until all the indices of the items are reduced to zero, then,

> at some point in the (uniform) computation to LF, there is an operation Spell-Out that applies to the structure Σ already formed. Spell-Out strips away from Σ those elements relevant only to π, leaving the residue Σ_L, which is mapped to λ by operations of the kind used to form Σ. Σ itself is then mapped to π by operations unlike those of the $N \rightarrow \lambda$ computation. We call the subsystem of C_{HL} that maps Σ to π the *phonological component*, and the subsystem that continues the computation from Σ_L to LF the *covert component*. The pre-Spell-Out computation we call *overt*. Let us assume further that Spell-Out delivers Σ to the module Morphology, which constructs word-like units that are then subjected to further phonological processes that map it finally to π (p 229)

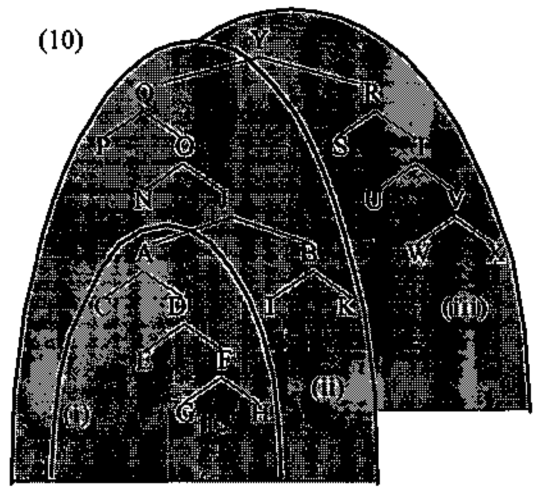
PF and LF interfaces traffic in different kinds of information—each uninterpretable by the other. What unifies a particular bit of 'sound' (PF-related) with a particular bit of 'meaning' (LF-related) is a derivational history from the some delimited subpart of the lexicon (NUM). But since the levels really are taken to traffic in different sorts of information, the computation is taken to split, sending the relevant information to the relevant levels. This splitting of the derivation is called Spell-Out. Chomsky assumes this operation applies freely at any point in a derivation but assumes that it applies *only once*, stripping away the material relevant to PF and leaving the rest to undergo further 'covert' operations.

Uriagereka (1996, 1998) observes the problems that arise (mentioned above) from trying to fit the LCA into the context of Chomsky's BPS. He suggests that the *locally linear* orders made available by (asymmetric) c-command are sufficient for linearization in the BPS system if the rule of Spell-Out applies more than once—the idea is that all (complex) non-complements are forced to Spell-Out when they are merged. In effect, he proposes that *just the base step* of the LCA is required to map phrase-markers to linear orders:

> LCA (MULTIPLE SPELLOUT VERSION):
> If α asymmetrically c-commands β then α precedes β

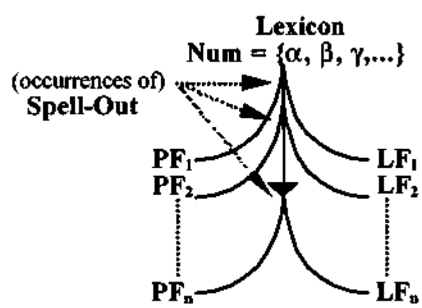
Uriagereka offers the helpful (visual) metaphor of a Calder-mobile—phrase-markers in BPS don't code any linear order—one or another linear order is forced on 'the mobile' if, for example, it is laid on the floor (or photographed). That is, Merge doesn't *concatenate* lexicon items, it merely *associates* them. What imposes order is the LCA—locally construed so as to linearize sub-derivations which, from that point on, are essentially treated at terminal items. For an illustration, consider the following structure in (10) which contains three 'command units':

(10)



Uriagereka's suggestion is that phrase-markers are Spelled-Out in 'derivational cascades'—structures are linearized one command unit/cascade at a time. In the above illustration, (i)-(iii) represent occurrences of the Spell-Out operation. This 'radically derivational' system, as Uriagereka points out, calls into question the status of the very notion of "level of representation"—that is, if it is command units that are cyclically shipped to the interfaces, then in what sense is there ever a unified object that is subjected 'all at once' to representational conditions? The resulting model Uriagereka dubs "dynamically bifurcated" and is illustrated in (11):[14]

(11)



Under this view, Spelled-Out constituents are literally shipped to the interface—they are no longer part of the derivation. Thus there is no *unitary* level of representation at which phrase-markers are evaluated. This interesting and extreme view, Uriagereka suggests, provides a rationale for the existence of *agreement* phenomena in natural language.[15] The central idea that will be recruited is the notion that structural descriptions are composed of separate sub-derivations, each of which 'lives its own life' so to speak, and only interact with each other ***as units***. The way to understand Uriagereka's proposal is to say that what Spell-Out does is convert complex (molecular) objects into simple (atomic) ones. Below, we will view this as following from a 'minimalist' restriction on Merge—that it may only take individuals (and never sets) as input.

This kind of proposal is another instance of a system that has the tampering properties of interest—that offers a salient difference between derivational and representational systems. Under Uriagereka's view, the well-formedness of a structural description is factored into sub-structures that, other than as units, ***shouldn't interact***. That is, after a command unit has been spelled-out, its internal structure should simply be inaccessible to future operations because it literally ***isn't there***. To underscore the potential utility of this view—Uriagereka (1998)

[14] This model, as Uriagereka notes, is a conceptual cousin to that proposed by Jackendoff (1972) in which interpretation involves cyclic access to the syntax.

[15] I will not pursue this here, but the basic idea is that agreement markings are 'addresses' that the interface component(s) use to figure out what goes with what—since phrase-markers are sent to the interfaces in parts, the proposal is that agreement functions to unify the parts post-syntactically.

suggests that this might be a fruitful way of viewing islands. Recall that under an Extension-regulated version of Chomsky's BPS, there has to be multiply-rooted intermediate stages of derivation. According to Uriagereka, a Merge operation that involves any complex non-complement (i.e. subjects/adjuncts) will necessarily involve an application of Spell-Out. If subjects and adjuncts must Spell-Out before they are Merged, then the system has a straightforward way of understanding strong island phenomena.

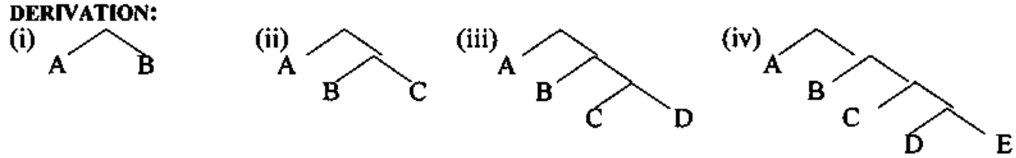
This is exactly the sort of 'tampering' operation that yields a derivational system that won't have any straightforward recoding in a representational system. The reason is that this kind of system alters the structural information as it goes—there is never a unified object that is subjected to conditions. Before turning examine the fit of this idea with the proposed understanding of Merge introduced above (Atomic Merge), we turn briefly to Phillips (1996), who is the only other author I know of to explore the mechanics of the sort that are adopted here.

1.3 Merge Right Phillips (1996) proposes viewing derivational structure building as working incrementally left-to-right on a given linear order of words. His agenda is to identify this kind of derivation with the on-line, real-time syntactic component of sentence processing—the Parser is the Grammar thesis. But the arguments for the kind of derivations he proposes are not solely motivated by the desire to make the grammar look like a plausible parser—his aim is to provide independent evidence from two directions. From both the competence perspective *and* from the performance perspective, he argues, there is evidence that suggests the same basic structure-building operations. What have traditionally been taken to be two investigations of two rather different sorts of things (a competence theory of grammars and a performance theory of parsers) are argued to really be two different studies of the *same thing*. This paper does not have a position to take on this very interesting program, nor will it have anything to say about the evidence and argumentation that Phillips marshals in support of this thesis from the sentence processing perspective. What is of interest here is his basic architecture—the core of which is summed up by the following conditions:

> MERGE RIGHT:
> New items must be introduced at the right edge of the structure
> BRANCH RIGHT:
> *Metric*: select the most right-branching available attachment of an incoming item
> *Reference set*: all attachments of a new item that are compatible with a given interpretation

The guiding idea is that derivations, given a sequence of items, attempt to structure them in a way that creates the 'best match' between precedence and c-command.

(12) INPUT: ⟨A, B, C, D, E⟩
DERIVATION:



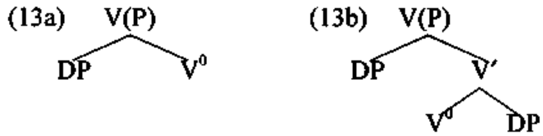
This system roughly takes the intuition of Kayne's (1994) Linear Correspondence Axiom as a (violable/best-match) condition on left-to-right structure building.[16] In addition to the two conditions above, Phillips (1995) also states a condition which enforces monotonicity:

> PRESERVE DOMINANCE:
> (a) If X is immediately dominated by Y in the input structure, then X is immediately dominated by Y in the output structure
> (b) If X immediately dominates Y in the input structure, then X immediately dominates Y in the output structure.

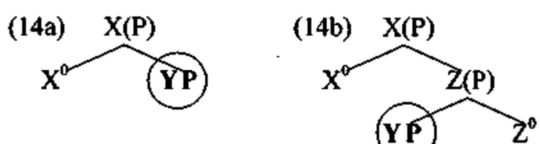
[16] And thus from the processing perspective he attempts to assimilate structural parsing preferences to something akin to Right Association (Kimball 1973). Other essentially equivalent strategies have included Late Closure (Frazier 1978), Attach Low (Abney 1989), Recency (Gibson 1991). See Phillips (1996:110) for discussion.

There are four things to note. First, the system makes reference to the "right" edge of the structure. This belies Phillips' parsing concerns—he assumes that the input to structure-building processes is the order of the input for the comprehension task.[17] Second, BRANCH RIGHT is a violable constraint in the following sense—it does not mandate that *all* structural additions are right-branching, just that out of the possible ways of adding to the structure, the chosen addition must the *most* right branching. In his system, *most* right branching is cashed out in terms of the shortest path of arcs between the item to be added and the one previous. Third, alternative possible additions to structure form a class (the reference set) which is defined by compatibility with a "given interpretation". Note that the reference to *interpretations* in defining the set of comparable attachments has conceptual cousins in several proposed instantiations of economy conditions in the Minimalist Program ("Have an Effect on the Output", see Chomsky 1995:294; Johnson & Lappin 1997).[18]

The fourth thing to note about the Phillips-style derivations is *that they change constituency relations as they proceed.* This feature, Phillips argues, allows for an account of contradictions between various constituency diagnostics. A glance at the derivation illustrated above in (12) demonstrates the point—an item that begins its 'derivational life' in a sisterhood relation with the *previously* attached item is 'pushed-down' into the structure as the *next* item is attached (e.g. in the first step (i) of the derivation in (12), A and B are sisters and form a constituent; in the second step (ii) they do not form a constituent exclusive of C). For this to work in a way that won't make 'wild' changes to already established relations the PRESERVE DOMINANCE constraint is important. Note that it necessarily treats X^0, X', and X^{max} as identical (in the spirit of Bare Phrase Structure—Chomsky 1995:chap4) so that in the following derivation, for example, this step-to-step condition on immediate dominance relations is not violated.



Here (13a) is the input structure to which the object DP is to be added, yielding the output in (13b). Since V^0, V', and V^{max} are viewed as essentially the 'same item' PRESERVE DOMINANCE is respected. Phillips does not explicitly discuss—and it is not totally clear from his discussion that he would want to allow—cases in which the material that is added to the structure becomes the specifier of the *following* head.



In (14), the YP *is to become* the specifier of the next head (Z). Here the condition on dominance is clearly *not* respected—it is the incoming head that projects, and not the specifier, and therefore *immediate* dominance is not respected from the input (14a) to the output (14b). Below I suggest that this sort of process should be admitted in the AM/MSO system.

This understanding of constituency as a *dynamic* notion gives him an explanation of constituency test conflicts. To illustrate consider Phillips' discussion of VP-Fronting.[19]

[17] Or the order of the input to the production systems (thought?).

[18] To anticipate the discussion of economy that will become important below, there have been proposals in which derivations can be members of a reference set (which establishes comparability for economy purposes) only if they correspond either to the same LF (in some proposals) or to the same interpretation (in other proposals). Chomsky (1995:294) states the following condition on 'forming' numerations: α enters the numeration only if it has an effect on the output. Lasnik (1997:32) and Johnson & Lappin (1997) point out the complexity entailed by such a condition. Phillips (1996:29, fn.16) notes that his condition's reference to interpretations entails that interpretation happens *incrementally*, but doesn't offer an approach to how this would work, citing the work of Steedman (1996) as an example of one sort of proposal that offers such incremental interpretive mechanisms.

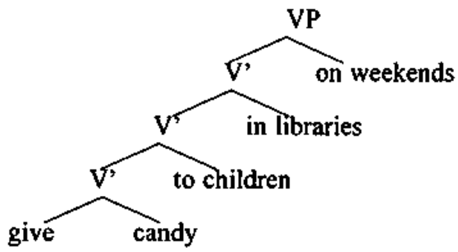
[19] For full discussion see Phillips (1996:chapter 1).

These cases are of interest because they suggest a left-branching structure for VPs since collections of items starting from the left can be fronted leaving the rest behind. Consider (15):

(15) a. ...and [give candy to children in libraries on weekends] he did.
b. ...and [give candy to children in libraries] he did on weekends.
c. ...and [give candy to children] he did in libraries on weekends.
d. ...and [give candy] he did to children in libraries on weekends.
e. *...and [to children in libraries] he did give candy on weekends.
f. *...and [in libraries on weekends] he did give candy to children.

This indicates a structure like that in (16):

(16)

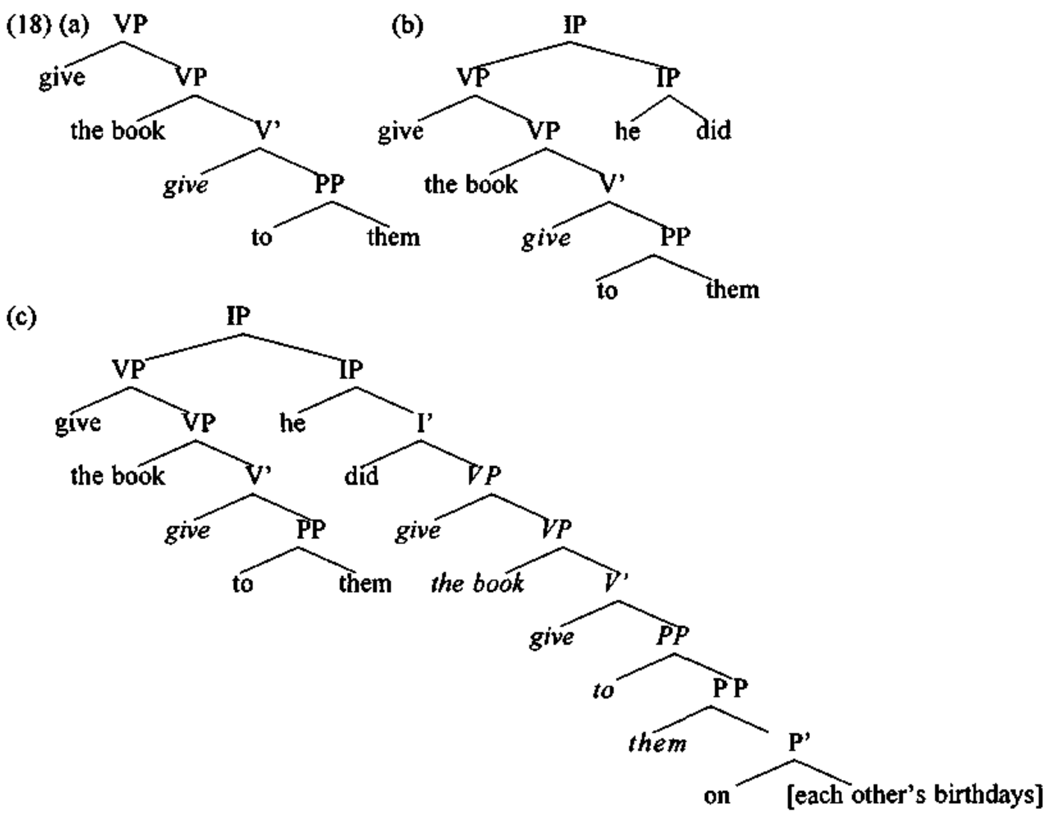


Now consider the cases in (17) (from Pesetsky 1995):

(17) a. ...and [give the book to **them**$_i$ in the garden] he did on **each other's**$_i$ birthdays.
b. ...and [give the book to **them**$_i$] he did in the garden on **each other's**$_i$ birthdays.

Assuming the binding of reciprocals requires c-command by an antecedent, (17) conflicts with the structure in (16). Contained in the fronted position, *them* does not command *each other's*. And given (16), it would not command the reciprocal even if it was 'reconstructed' into the gap site. So, the facts about which parts of the VP are viable targets for fronting (i.e. are constituents) and the command requirement for the binding facts create a conflict, suggesting two different structures.

Phillips' left-to-right derivation for (17) goes as in (18):

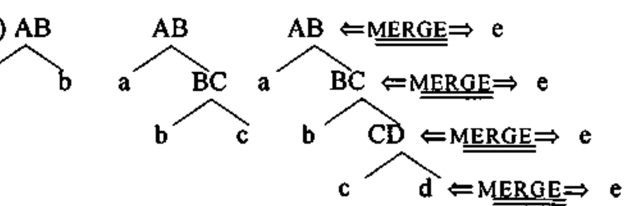


(18a) is the left-to-right derivation of the fronted material. The de-emphasized *give* is a null 'lowered' copy. Phillips takes this sort of derivation of a Larsonian VP (Larson 1988) to follow from his Merge/Branch Right. That is, the Larsonian VP-shell is the 'most right-branching' possibility. The point, however, is that the fronted material is assembled as a right- and not left-branching structure. The IP is constructed (18b) and the 'fronted' structure is 'lowered' (thus the de-emphasized 'copy' in (18c)) and the rest of the VP is added. The trick is to have whatever processes target the portion of the VP to be 'fronted' (here, 'lowered') has to happen before right edge elements are added to the structure thereby changing the existing constituency relations.

This is *exactly* the sort of benefit from a 'tampering' with relations alluded to above. To underscore once again this investigation's underlying theme—benefits that come *in virtue of* assuming that what can be established at one point in a derivation can be *undone* at some subsequent point are to be prized since they pose a challenge for strictly representational systems. Indeed, such benefits should be even more valued if they are only reached by assuming ***a particular directionality of derivation***. These basic mechanics will form the point of departure for the AM/MSO system.

1.4 Putting the Pieces Together: the AM/MSO system Now that the components and their recent antecedents in the literature have been reviewed, we can illuminate the entire picture. Recall from above the introduction of Atomic Merge, which restricts its inputs to simple terms. Lexicon items under this view have (as Williams 1994 puts it) "top-level" properties with respect to the syntax (i.e. with respect to Merge). That is, even though there is/are some level(s) at which these items have internal structure—that is to say, these items "have parts"—the syntax treats them as individuals, not sets.[20,21]

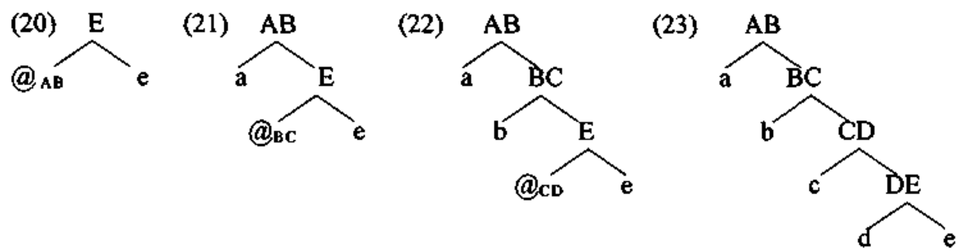
Given the introduction of Uriagereka's Multiple Spell-Out proposal above, the next move should be obvious. Merge requires atomic terms. And, if we feel pretty confident that complex syntactic structures relate to one another (which its safe to assume that we do), and if Spell-Out is viewed as 'converting' complex (molecular) structure into individuals (atoms), then we can view Spell-Out as occurring when it is "forced to". If an input item is required to target some already formed complex structure to satisfy some property (either of the item or of the target) then Spell-Out must apply to render the complex structure atomic—that is, to make the complex structure a licit input to the Merge operation. To illustrate the system in the abstract (we'll turn to some more concrete examples below), consider the following. The derivation in (19) has assembled the lexicon items (a, b, c, and d), and a fifth item e is about to be added. The possible attachments and required Spell-Outs are depicted below in (20)-(23):

(19)


output in (20)
output in (21)
output in (22)
output in (23)

[20] Recall that Chomsky (1995) discusses the Merge operation in set-theoretic terms. He worries about whether the derived item resulting from an application of Merge should be understood as having their properties determined by the union vs. intersection vs. identity of the sets of items from which it is derived. But this means that the input items as sets. That is, if the operation itself is some variety of standard set-theoretic relation, then, for example, $\{\alpha, \beta\}$ must be the result of Merging $\{\alpha\}$ and $\{\beta\}$ and not just α and β. Or, if α and β are themselves sets (say, a triple of LF, FF, and PF, each in itself a matrix or set of features) then the Merge of α and β is really the Merge of $\{LF\alpha, FF\alpha, PF\alpha\}$ and $\{LF\beta, FF\beta, PF\beta\}$, which would output the set $\{LF\alpha, FF\alpha, PF\alpha, LF\beta, FF\beta, PF\beta\}$. We avoid the potential confusions that arise here by stipulating that the inputs to Merge are memberless (individuals). The point regarding whether or not the input items to Merge are sets is discussed in fn.27 of Chomsky's (1995:chapter 4) where he attributes the question to Chris Collins. I thank Norbert Hornstein for bringing this note to my attention.

[21] This probably doesn't require mention here, but on the difference between 'having parts' and 'having members' see Lewis (1991). It turns out that the difference isn't quite so clear as I'm making it here. The singleton-individual distinction—which, if you think about it, I'm making much of here—is a pretty obscure conceptual division that Lewis argues that we really don't understand.



The @ is a notation which will represent the label of the Spelled-Out structure. Following Uriagereka, these addresses/labels representing Spelled-Out material are not special formatives—they are simply the top-label of the Spelled-Out structure (the rest having been 'stipped away'). Crucially, this will behave like a lexicon item (a sort of 'derived' terminal) in that it will be considered a simple term. In (20) the root of the structure in (19) has been targeted, causing the entire object to Spell-Out—this leaves a label ($@_{AB}$) which merges with the item *e*. In (21), the node BC is targeted, causing everything below it to Spell-Out, leaving the address/label $@_{BC}$, which merges with *e*, etc.. These derived objects are viewed as having 'parts'/internal complexity, and may have certain properties in virtue of this internal complexity, but the internal structure is viewed here as *inaccessible* to further syntactic operations. I will subscript the @'s with their type information as well as other relevant properties (e.g. if this unit becomes Case or θ-marked, etc.).

The system works like this:

- **Merge requires simple terms;**
- **Merge to any *non-terminal* forces Spell-Out to apply so that both of the items that are to be merged are simple terms.**

Returning to the status of precedence/command relations, note first that here Spell-Out is *not* being viewed as applying when forced by linearization requirements. Instead, Spell-Out is forced by the nature of the Merge operation itself—*Merge requires simple terms as input.* Now, following this paper's interpretation of Kayne's insight (which sees hierarchy and linear order as inextricably linked) suppose that we consider what happens if we take precedence to be 'more basic' and command to be derived. The line of reasoning might go like this:

- **Precedence is mapped from command (Kayne);**
  AND,
- **Command = "derivationally combine" (Epstein)**

Given this, it seems worthwhile exploring the consequences of giving up the 'middleman' here, since, given the above pair of claims, we might try out the idea that:

- **Precedence follows from Derivational Combination**

Suppose this is the case. Precedence is then identified with the order of Select/Merge. Command, under these assumptions, can be viewed as "precedence at time **t** of a derivation"—that is, command can be understood as precedence up to Spell-Out:

- **Command = Sub-derivational Precedence**

Precedence relations are a global property of structural descriptions. This property is here associated with the very nature of serial derivation. Command is then a local sub-case of precedence—it is the order of merge operations within 'derivational cascades'. Put another way, the asymmetries we worried about earlier are now related to the asymmetry of the 'logical time' of the derivation. Note that what has happened here is that command relations have essentially been *abandoned*—in their place is the notion "sub-derivational precedence".[22]

[22] As such, this understanding of 'command' doesn't count as a command relations according to the criteria of Barker & Pullum, who include the property "ambidextrousness" (i.e. that command relations are insensitive to precedence). This just underscores the point that is being made here—that the relations that have been taken to be mediated by command, can now be understood in terms of a derivationally local sub-case of precedence. Importantly, if an item β is

Notice also that this captures (a version of) the 'minimal factorization' insight of Chametzky mentioned above. For any item in the structure at any point in the derivation, a look 'back up the tree' reveals nothing but the items that we would ordinarily say c-command that item.

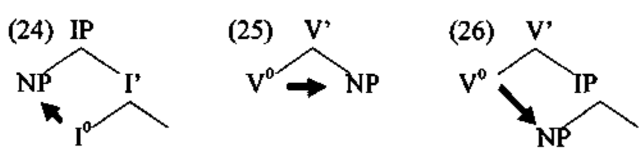
Before moving on to consider some of the properties and potential utility of this system, a word is in order regarding the identification of derivational combination and precedence. Chametsky (1996:16-17) argues that while dominance is a well-founded formal relation, precedence is not. Consider the following remarks:

> the...reason to maintain dominance is that branching part-whole structures are...not specific to linguistic theory. Simon (1962) demonstrates that such hierarchical structuring is the "architecture of complexity" with respect to natural systems...There might appear to be an unwarranted asymmetry in the use of "the architecture of complexity" to support dominance and the use of the "the physics of speech" above to reject precedence...However,..."the physics of speech"...is parochial in more than one way: (1) it is limited to human language and (2) it is not even general there.

The last part of these remarks is referring to the existence of signed languages, where presumably the "physics of sign" may allow for simultaneity in ways that the "physics of speech" might not. I think that the argument that Chametzky is anticipating (and trying to preempt) can be made with respect to an 'asymmetry' between his conceptual arguments for maintaining dominance and rejecting precedence. Note that the same *kind* of argument can be made for precedence, based on the directionality of time. Certainly this is as general a feature or constraint on complex systems as hiearchical part-whole relations are. Clearly this is a boundary condition on the system under investigation—one that holds equally well of spoken and signed languages, where the latter contrast could be viewed simply as a problem of grain or scale. Surely the simultaneity that exists in sign exists in speech, but the 'grain' or 'scale' is such that this is far less obvious. And surely sign, like speech, is subject to ordering of formatives. The 'arrow' of time insists on it. That the simultaneity that exists in the production of sign is, well, so ***visible***, is just an accident of the size of the articulators.

Regardless of the status of these conceptual arguments, what I wish to investigate below is the what sort of understanding we gain by going this route (by having precedence as basic in virtue of assuming a derivational system). By exploiting the notion of a sequence of representation, or a logical ordering of construction procedures, what sort of understanding of the system do we arrive at? I've suggested above that the asymmetries theoretician's 'build-in' to their systems in an attempt to derive the properties like precedence are in a sense basic—that current views have things essentially backwards.

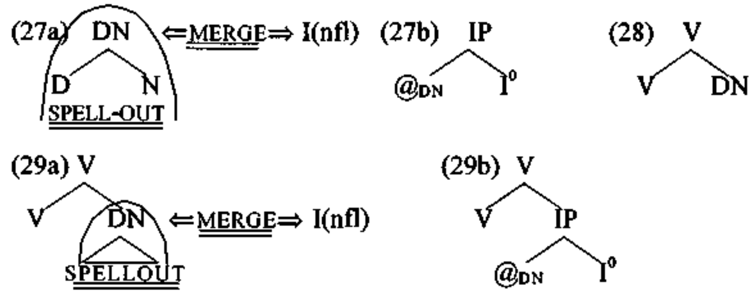
1.5 Some Properties of the AM/MSO System One immediate property of this architecture is that it suggests that all syntactic relations can and perhaps should be captured/licensed by what is probably the simplest syntactic relation that there is: sisterhood. Relations that, following Chomsky (1993), have been recently subsumed under the Spec-Head relation, can now be recast in terms of sisterhood. To see this, consider Case relations and their various associated structural configurations in GB:



The first was taken to be the configuration under which Nominative Case was licensed/checked/assigned. The second was taken in GB to the canonical government configuration in which Accusative, Dative and various 'inherent' Cases were licensed. The third configuration is that of Exceptional Case Marking, requiring S-bar deletion or its more modern equivalent. Chomsky (1993) proposed (adopting the split-IP of Pollock 1989) that all three of these relations can and should be uniformly licensed in Spec-Head relations. Thus, in English for example, both direct objects and ECM subjects were taken to covertly raise (post-

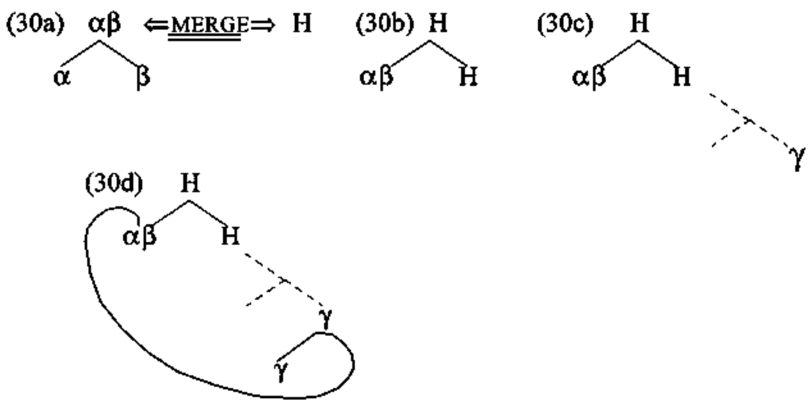
introduced in a derivation 'after' an item α, then β cannot command nor precede α. There is one exception—if α is 'displaced' in a latter derivational stage so as to be introduced, as it were, both before and after β. To the extent that we require 'backwards' command relations, they will have to recast given some sort of movement/displacement operation.

Spell-Out movement) to the specifier position of a functional head to 'check'/license Case. Given the view proposed here, we can assimilate all of these relations to the one that GB took to be the canonical one. To illustrate:



In (27), the introduction of a head targeting already constructed complex structure forces Spell-Out (predicting the inaccessibility of material in this "left branch"). Following this the head can be targeted by further merge operations (say, to form a complement). The second example (28) is just the familiar head-complement relation, but interestingly, if another head was to be introduced as in the third case (29), the complement would be forced to Spell-Out (if complex), essentially becoming the specifier of the following head. This suggests a natural mechanical way to treat ECM constructions. The ECM subject can 'begin' as the complement of the ECM verb, and then be Spelled-Out to become the specifier of the next head.

Consider movement. The items α and β have been merged (30a), and a following head (H) has merged (30a-b) to the complex $\alpha\beta$, triggering Spell-Out. The derivation has proceeded to the point depicted in (30c), where γ is introduced and merged. Suppose that γ has some property that $\alpha\beta$ can satisfy, or vice-versa (e.g. $\alpha\beta$ is a *wh*-phrase, $H = C^0$ and γ is a θ- or Case-assignor). Suppose that $\alpha\beta$ simply ***Re-Merges***, associating with γ as in (30d)



This is how the AM/MSO system will instantiate displacement operations in general. Notice that there is no need for an operation Delete—Spell-Out, by hypothesis, has "stripped away from [the derivation] those elements relevant only to π [PF]" (Chomsky 1995:p.229) at the point where the head (H) was introduced. No issue arises as to whether the item $\alpha\beta$ is pronounced in the upper or lower position. Given the status of precedence relations in this system (order of merge = order of formatives) and the requirement on merge, $\alpha\beta$ must be 'pronounced' in the upper position—so long as the object $\alpha\beta$ is ***complex***. If it is not then the expectation is that it should be pronounced in multiple positions. This means that in cases of, for example, *wh*-movement in English, elements like *what* must be *syntactically* complex—or consider names or pronouns—where such items only appear 'pronounced' in one position, this system is saying that this is so because these items involve complex syntactic structure. This seems plausible, and I will assume that this is the case in what follows.[23]

[23] A number of analytical issues arise here with respect the possibility of capturing *wh*-copying in German, or languages in which there is copying of clitics, etc.. I do not pursue these matters here, but note that there is a natural route to take in treating such phenomena in this system hinging on whether or not the items involve syntactic vs. morphological complexity. On complexity of DPs, see Longobardi (1995) among others. This is also a promising route

But what sort of object is created? Thinking of things graph-theoretically, 'movement' (re-merge) in this system creates 'loops' in the structure—it turns a rooted, acyclic graph into a cyclic one. There is no need for an operation Copy in this system either. There are just two operations—operations that every minimalist theory has—Merge and Spell-Out. Essentially the result of this (re)merge is an object that coresides in more than one position in the sequence. The arcs connecting remote nodes in a tree will be viewed as *fundamentally no different* than the arcs that standardly indicate the *is a* relations. It will emerge that this is not simply a "notational variant" of one or another formulation of movement—the item that will be viewed as occupying two positions standardly understood as being related by movement will literally be the same item—not a copy, not a copy associated with a trace, not a chain (unless we simply want to 'call' these objects chains, this I have no problem with)—the *same* instance of the *same* item. Call this the *Coresidence* view of movement. This view of movement seems to me to be quite close to at least one reading of Chomsky's (1995:ch.4) assumptions:

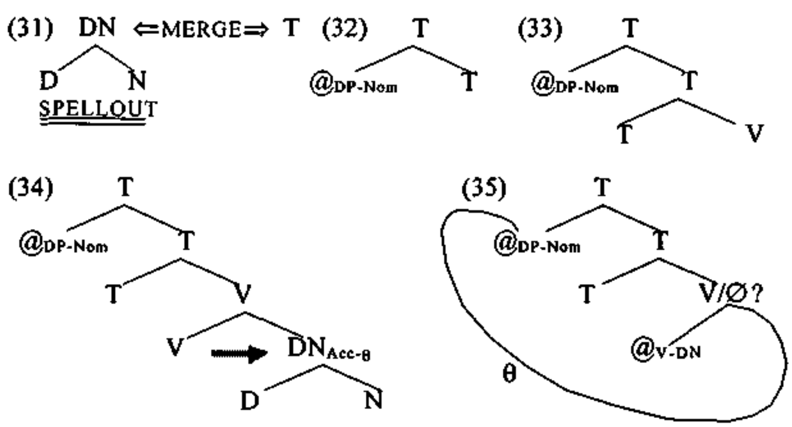
> Under the copy theory of movement...a two element chain is a pair ⟨α, β⟩, where α = β...we want to distinguish the two elements of the chain CH formed by this operation. The natural way to do so is by inspection of the context in which the term appears. Given the way syntactic objects are formed, it suffices to consider the co-constituent (sister) of a term, always distinct for α and its trace (p 251-52)

The idea is to identify the 'links' of a chain with the positions in which the links reside. For example, if an item α that is a sister to an item β is copied and merged ('raised') and the copy becomes the sister of an item K, the resulting chain can be viewed as a pair of pairs {⟨α, K⟩, ⟨α, β⟩}. Further,

> though α **and its trace are identical,** the two positions are distinct. We can take the chain CH that is the object interpreted at LF to be the pair of positions (p252 [emphasis mine-JED).

That is, we can identify the chain CH as {K, β}. This is essentially the idea that this paper is advocating—the difference is going to turn on whether there is *any* sense in which the 'copies' will be treated as distinct—as having their own independent 'derivational reality'. Under the so-called Copy Theory of movement, a second operation of Deletion is required to ensure that an item is only 'pronounced' (and/or 'interpreted') in one position. The view here will be that the Copy and Delete view of movement introduces operations that are superfluous—we can view 'movement' as just the (Re)Merge of an item following its Spell-Out.

Consider now a basic case to illustrate the mechanics. Below is the derivation of a simple transitive clause.

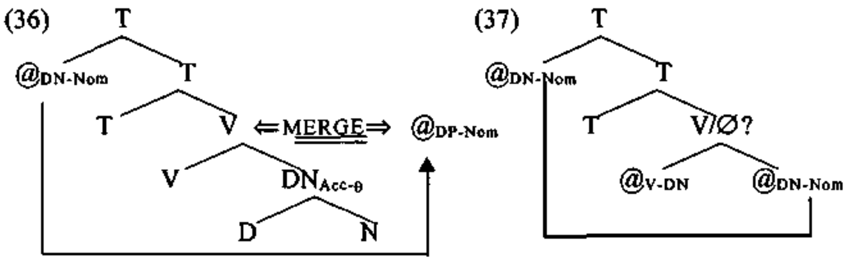


In (31), D and N Merge to form a DP.[24] The introduction of T forces DN to Spell-Out. I introduce here a notation (@) to represent Spelled-Out units, with subscripts to identify them. T assigns Nominative case to the Spelled-Out DP, yielding (32). Next, the verb and

into understanding the *wh*-copying that occurs in L1 acquisition of English (e.g. *what do you think **what** Cookie monster eats*?). See Crain & Thornton (1998).

[24] As above, I occassionally adopt a notation neutral to the status of projection. Thus DN = combination of D and N with no commitment here to the nature of projection. Were I do not do this is unimportant here (e.g. assuming T projects when combined with V, etc.)

object are added to the structure. I assume here following Bobaljik & Thrainsson (1997) that in English, both Accusative Case that the internal/direct object θ-role are assigned in the same position.[25] Step (35) is non-obvious, but has attractive properties. What has happened in step (35) is more perspicuously represented in a pair of steps below in (36) & (37):



Assuming (following Dowty 1982 and many others) that verb-object units compositionally θ-mark subjects, I assume that the subject targets the entire unit. In general I will also assume that Spell-Out units do not project—but being unsure whether or not it matters for this Verb-Complement unit, I include the Ø here for the nonce. This representation captures the familiar 'complement-dependent' nature of the subject θ-role, discussed for example in Marantz (1984). Consider the following:

(38) a. throw a baseball
b. throw support behind a candidate
c. throw a boxing match (i.e. take a dive)
d. throw a party
e. throw a fit
f. take a book from the shelf
g. take a bus to New York
h. take a nap
i. take a letter in shorthand
j. kill a cockroach
k. kill a conversation
l. kill an evening watching T.V.
m. kill a bottle (i.e. empty it)
n. kill an audience (i.e. wow them)

That is, the interpretation of the D/NP that is to be the subject of (38a-n) appears to be heavily reliant on the nature of the complement. Consider why its worth mentioning these kind of facts. In a bottom-up derivational system, it is trivial to capture the compositional nature that subject θ-roles might have given (38)—but it is not obvious how to do this in a system that works roughly in the opposite direction. In the AM/MSO system, what this means is that the subject has to essentially 'wait' until the verb-object unit has been formed before the subject can associate with it to receive a θ-role.[26]

This representation has another advantageous feature in that it creates a transpositional object—the subject related to two positions (Case, θ)—which 'brackets' the object position. This kind of 'interleaving' has been argued by Hornstein (1995) to provide an account of relative quantifier scope. There it was argued that the operation of Q(uantifier) R(aising) is suspect given minimalist assumptions and should be abandoned. In place of QR, Hornstein

[25] This is one of the three conceivable options with respect to object Case assignment. The other would be to argue for an analogue of the proposals by Johnson (1990), Koizumi (1990), and Lasnik (1995) that both objects and verbs move 'overtly' in English. Here, the translation would require that objects are spelled-out in their Case position, and their label is re-Merged to the θ-position. The option standardly assumed (that object Case-checking/licensing) involves covert movement is difficult to instantiate here. I will not explore the various alternatives here for reasons of space, but note that either the overt movement of objects (i.e. they're pronounced in their upper/Case positions) or the view that objects receive Case and θ in the same position are the most natural of the three logically possible stories. I assume here without argument that, following Bobaljik & Thrainsson, objects stay *in situ*.
[26] What this also does is structurally assimilate subjects and adjuncts in a way that I will put to work in extensions of the analysis of parasitic gaps in Drury (to appear).

provides an A-movement account of the familiar ambiguity in (39), which has the two possible interpretations paraphrased in (39a) and (39b).

(39) Everybody loves somebody
(39a) "For everybody x, there is a person y, such that x loves y"
(39b) "There is a person y is such that for every person x, x loves y"

His analysis adopts the idea that the licensing of structural CASE is regularized to the Spec-Head configuration (Chomsky 1993) and that CASE features are uninterpretable at the interface and must thus be eliminated via checking prior to LF to allow convergence. Hornstein argues that given this view of CASE theory, there is no need to have a separate operation of QR to handle the ambiguity of (39)—if both the subject and the object must raise to check CASE, then we have the following LF structure for (39):

(40) $[_{AGRsP}\ \textbf{everybody}_i\ [\ AGR_S^0\ [_{TP}\ [\ T^0[_{AGRoP}\ \textbf{somebody}_j\ [\ AGR_O^0\ [_{VP}\ \text{everybody}_i\ [_{V'}\ V^0\ \text{somebody}_j]]]]]]]]$

Here the subject raises to Spec-Agr_S and the object raises (covertly) to Spec-Agr_O in order to check Nominative and Accusative CASE respectively. What is crucial here is that the resulting CHAINS are interleaved—the head of the object CHAIN commands the foot of the subject CHAIN. If the semantic notion of scope is understood as being read-off of the syntactic command relations the ambiguity of (39) can be seen to result from a structure like (40) depending on which copy of the chain is interpreted at LF—the upper or the lower one. Hornstein assumes that, (i) all but one copy of a CHAIN must delete and, (ii) the grammar can freely delete either the upper or lower copy. Thus, the two readings of (39) result from whether the head or the tail of the relevant CHAINS are retained at LF, as in (41).

(41a) $[_{AGRsP}\ \textbf{everybody}\ [\ AGR_S^0\ [_{TP}\ [\ T^0[_{AGRoP}\ \textbf{somebody}\ [\ AGR_O^0\ [_{VP}\ \cancel{\textbf{everybody}}\ [_{V'}\ V^0\ \cancel{\textbf{somebody}}]]]]]]]]$
(41b) $[_{AGRsP}\ \cancel{\textbf{everybody}}\ [\ AGR_S^0\ [_{TP}\ [\ T^0[_{AGRoP}\ \textbf{somebody}\ [\ AGR_O^0\ [_{VP}\ \textbf{everybody}[_{V'}\ V^0\ \cancel{\textbf{somebody}}]]]]]]]]$

In (41a), the top copy of the subject is retained so that at LF, *everybody* c-commands *somebody* and thus scopes over it, yielding the reading in (39a). In (41b), the bottom copy of *everybody* is retained and the top copy of *somebody* is retained so that the object c-commands the subject at LF and this yields the reading in (39b). Thus the operation of QR is argued to be eliminable in favor of the independently motivated tenets of Chomsky's (1993) version of CASE theory (which separates the domains of θ-assignment (VP) from CASE assignment (the split IP complex—$Agr_{S,O}$P, TP)) coupled with the mechanism of free deletion of all but one copy of a CHAIN.

However, this analysis of relative quantifier scope requires a crucial assumption—what Hornstein calls the ALL FOR ONE PRINCIPLE (AFOP). Consider again the structures in (41) given the Copy Theory of Movement. Recall that the subject and object move to check CASE features that must be eliminated or else the derivation will crash. If what underlies movement is COPY and MERGE, then the upper copies that enter into CASE-checking configurations will have their CASE features checked, but the bottom copies will not. The analysis offered for the quantifier scope ambiguity in (39) crucially relies on the possibility that *any* member of a CHAIN can be the one that is retained at LF. But if only the top copy has checked its ([-interpretable]) CASE features, a derivation that keeps only the bottom copy (as is necessary for the reading of (39b) under Hornstein's assumptions) should crash since the copy that is retained is exactly the one that has not checked any features. Thus the AFOP is introduced.

AFOP: Every link in a chain meets the morphological conditions satisfied by any link in a chain.

This principle allows for the tail of a CHAIN to have its CASE features checked if the head of that CHAIN has entered into a checking configuration. Put another way, CHAINS check features under this view, not individual copies/links of CHAINS—it is in virtue of having a copy/link in a certain (checking) configuration that a CHAIN has its features checked. Given the AFOP, *any* copy in a CHAIN can safely delete and Hornstein's A-movement analysis of quantifier scope ambiguities can be maintained.[27]

[27] Hornstein has since abandoned this view, as mentioned in the introduction. As we will see in §2 in the discussion of parasitic gap constuctions and adjunct control.

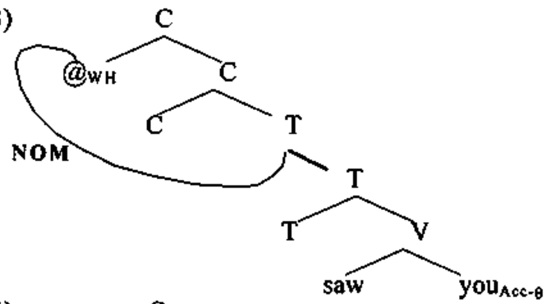
Given the interleaving of subject and object in the AM/MSO-system introduced above, these effects can be captured under the umbrella of A-movement. Notice however that the assumption (imported from Bobaljik & Thrainsson 1997) that objects stay in situ (check/are assigned Case and θ in one position) yields an interesting difference between this system and the basics of the story that Hornstein tells. Notice that under Hornstein's selective deletion view, there are actually *four* possible outputs:

(42) (a) [**everyone**...[...**someone**...[...~~everyone~~...[...~~someone~~...]]]]
(b) [**everyone**...[...~~someone~~...[...~~everyone~~...[...**someone**...]]]]
(c) [~~everyone~~...[...~~someone~~...[...**everyone**...[...**someone**...]]]]
(d) [~~everyone~~...[...**someone**...[...**everyone**...[...~~someone~~...]]]]

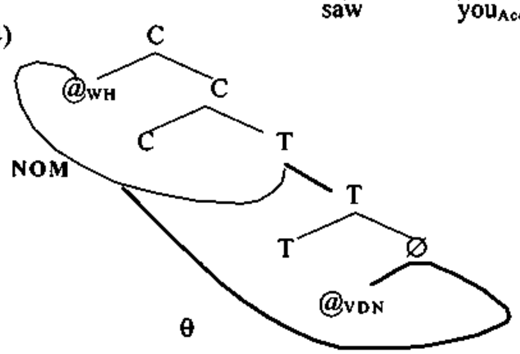
But of these, only the last yields an output that has the object scope over the subject.[28] This 'spurious ambiguity' isn't damaging in any way, but it is simply not present under the assumption that the object doesn't move (where there are then only two scope possibilities).

Let's examine the status of movement in the AM/MSO system in a little more detail. Take a case of simple *wh*-movement: *which man saw you?*

(43)



(44)



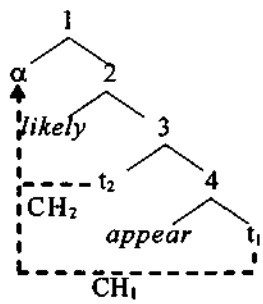
What has occurred here is that the *wh*-element has merged to C and been forced to Spell-Out (again, assuming these items to be syntactically complex). It Re-Merges to be assigned Case by T in (43). The derivation continues in (44), where the V is added, the complement constructed and assigned both Case and θ, and the *wh*-element (now Case-marked) re-Merges to the verb-complement unit to be compositionally assigned the subject θ-role. There is an interesting property that these sort of derivations have that suggests a formal correspondence between Merge and Move—here argued to be identity (i.e. there is just Merge). Uriagereka (1998:p.399) observes in a discussion of Chomsky's view on chains that there is an interesting symmetry between the formation of chains and the building of phrases by Merge. Consider a case of raising that he uses to make the point:

(45) John is likely t_2 to appear t_1 to be late.

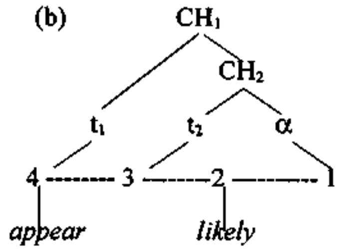
He speculates about the situation in the abstract, rotating the standard graph representing phrases to examine the nature of the object formed by "movement". I reproduce his figures here in (46) (his 78a,b):

[28] Juan Uriagereka has suggested (p.c.) that this property of Hornstein's system might be made to account for scope *preferences*. The idea would be that the reading that is overdetermined by multiple derivational routes compared to one that has less or just one derivation might be the 'preferred' reading. While this is possible, and if there is some preference for the subject with wide scope, this could also be conceivably accounted for in terms of precedence (i.e. prefer to have 1st scope over 2nd).

(46) (a)



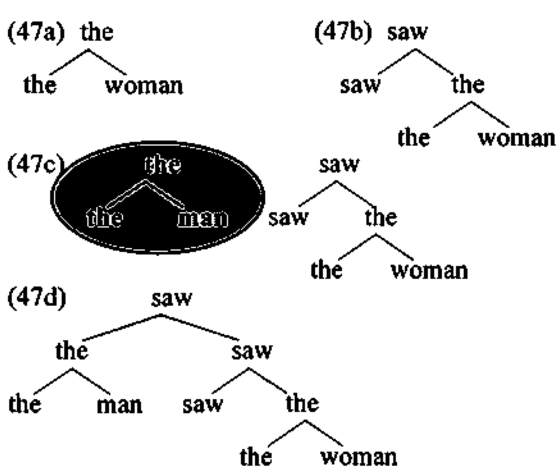
(b)



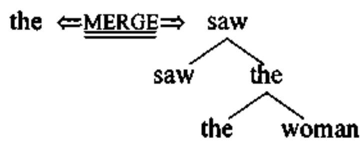
Above (46b) is simply a 'rotated' version of (46a), illustrating the properties of chains that parallel those of phrases. The (speculative) intuition is that "chains are merely instances of a higher-order Supermerge, where the constituent items aren't selected from a numeration, but instead are taken from a phrase marker that has already been assembled (using Merge at a lower dimension)." (p397). The idea one that fits with the guiding idea here about movement, except that instead of Supermerge, the relevant operation is simply (Atomic) Merge. Note that in the above "rotated graph", intermediate traces of movement/intermediate chain links are viewed as corresponding to the intermediate projections of phrases built by Merge. And the full object (CH_1) is taken to correspond to the notion of 'maximal projection' in phrase-building.

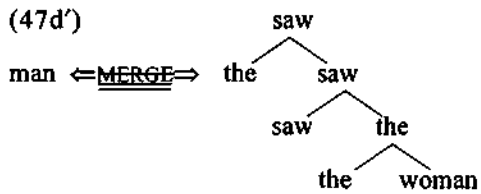
The AM/MSO system proposed here embraces this sort of intuition. The idea again is simply this—when the construction procedure has reached a point in which there are non-adjacent (in the sequence) elements which could satisfy properties of one another, they are simply (re)Merged. The object that is created is essentially *transpositional* (and as we'll see later, potentially *transderivational*)—it is an item that *coresides* in two (or more) positions. Since any such object that (re)Merges, if complex, will have had to Spell-Out, it will be pronounced in its initial Merge site, and share properties with other positions to which the label/address (now standing in for the Spelled-Out structure) associates.

To lead into the next section, consider a curious difference between the AM/MSO system and Extension-regulated, bottom-up systems like Chomsky's BPS. The latter, as mentioned in the introduction, actually requires that there be *intermediate* derivational steps in which there are ***multiple*** roots (unconnected complex subtrees). To illustrate, consider the derivation for the VP *the man saw the woman*:



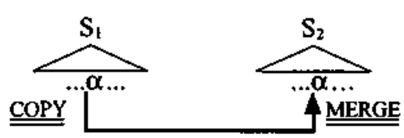
Step (47c) is the important one. In order to form a non-singleton specifier[29], either *man* or *the* must be selected and **not** merged. Select must apply a second time before the subject [$_{DP}$ the man] can be separately formed. The two subtrees shown in (47c) are then merged to construct the VP as in (47d). The point is that in order to build structures that contain complex non-complements, the system must allow for the possibility of separately assembling subtrees and then later combining them to create a single-rooted structure.[30] This is intermediate multiple-rootedness is thus mandated by the Extension Condition Imagine an alternative derivational step (47c′) where either *the* or *man* is added to the root instead of first separately combining to form a DP.

(47c′) 

(47d′) 

Now the determiner constitutes a singleton specifier, and the only way to add *man* into the structure is to violate the Extension Condition and merge below the root as in (47d′).[31] The DP must be separately constructed and then merged to satisfy Extension. Such intermediate derivational stages in a way exist in the AM/MSO system as well in virtue of (Multiple) Spell-Out. That is, what the Extension-regulated system brings out in an obvious way is the need to factor derivations into subderivations which interact as units. This is intuition that the AM/MSO system exploits. The difference, which at first blush seems inconsequential, is that the AM/MSO system starts with a connected object which is then necessarily 'factored' into subderivations in order to meet the requirements of simplicity of operations—i.e. in order to meet the requirement that the simplest instance of Merge (atomic merge) is the only operation that is permitted.

This is of interest given the possibility that there might be movement relations between such unconnected subtrees in systems that require them. Once there is a need to construct subtrees separately, assuming the a Copy + Merge theory of movement allows an item (α) in one subtree (S_1) to be copied and then Merged into a separate subtree (S_2). Such an operation would create a movement dependency across otherwise unconnected phrase markers:[32]



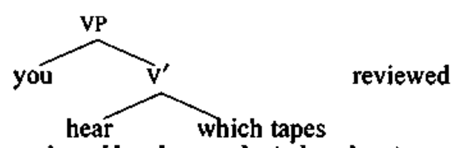
It has been argued that this sort of operation has in the past only been excluded by stipulation, and that removing this ban yields interesting conceptual and empirical gains. For example, such 'interarboreal' operations have recently been argued to render head movement unexceptional with respect to the Extension Requirement (by Bobaljik & Brown 1997, Uriagereka 1997); to underlie parasitic gap phenomena (Nunes 1995, Hornstein 1997a) and Control into adjuncts (Hornstein 1997a,b) and certain problematic cases of *wh*-movement (Uriagereka 1996, 1997). This is as it should be as pointed out by Bobaljik & Brown (1997: p345) either, "interarboreal operations should be admitted and exploited, or a principled reason to exclude them should be proposed." As mentioned in the introduction, *directionality*

[29] It will emerge that this problem arises for any complex non-complement.
[30] Or allow the possibility of 'overwriting'/'redefining' nonterminal nodes—i.e. abandon strict Extension.
[31] This same basic point is made by Bobaljik 1995, Bobaljik & Brown 1997, Hornstein 1996, Nunes 1995, Uriagereka 1996, 1998. It was first pointed out *to me* by Bob Frank and Joel Hoffman in Juan Uriagereka's Spring 1994 Seminar on Minimalism at UMCP.
[32] Note that this sort of operation requires that Move be dissociated from Chain Formation. I am unaware of any grammatical theory that allows CHAINS to have links that are not in a c-command configuration.

<u>FORM CHAIN</u>:
The syntactic objects α and β can form the nontrivial chain CH = (α, β) only if:
(i) α c-commands β;
(ii) α is nondistinct from β;
(iii) (at least) one given feature F of α enters into a checking relation with a sublabel of K, where K is the head (of the projection) with which α merges, and the corresponding feature F of β could enter into a checking relation with a sublabel of K; and
(iv) there is no syntactic object γ such that γ has a feature F' which is of the same type as the feature F of α, and γ is closer to α than β is.
<u>CHAIN REDUCTION</u>:
Delete the minimal number of terms of a nontrivial chain CH which suffices for CH to be mapped into a linear order in accordance with the LCA.

Since Copy and Merge are distinguished from Form Chain as distinct operations it is possible to copy an item from one subtree and merge it into a separate subtree. However, in this system a CHAIN need not and in fact cannot be formed until its members are in a c-command configuration. That is, Form Chain cannot apply until both copies are contained in the same subtree since c-command is not defined for items in unconnected structures. Only after the subtree containing the launching site of the moved item is merged with the subtree containing the landing site can Form Chain apply. This is what is claimed to underlie parasitic gap phenomena. Nunes argues that the derivation illustrated in (49) is responsible for examples such as the one in (48) (irrelevant structure is omitted).

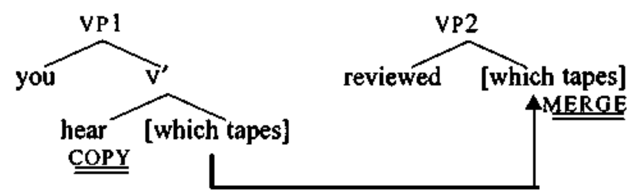
(48) Which tapes did you hear e before Bill reviewed **pg**
(49)
a) NUM = {which$_0$, tapes$_0$, **did**$_1$, you$_0$, hear$_0$, **before**$_1$, **Bill**$_1$, reviewed$_0$}



◊ <u>Note</u>: ***reviewed*** **has been selected and not merged**

b) NUM = {which$_0$, tapes$_0$, **did**$_1$, you$_0$, hear$_0$, **before**$_1$, **Bill**$_1$, reviewed$_0$}

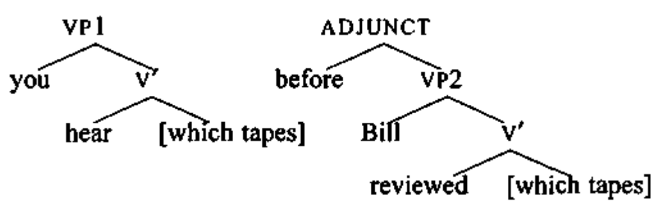
COPY([which papers])
MERGE([reviewed],[which tapes])



◊ <u>Note</u>: **Form Chain does not apply yet since the items are in unconnected subtrees**

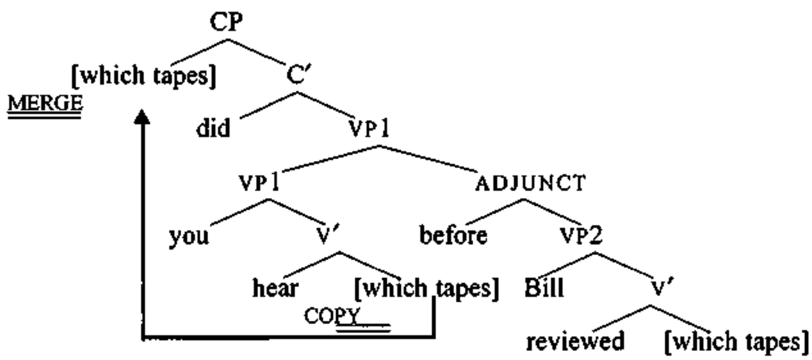
c) NUM = {which$_0$, tapes$_0$, **did**$_1$, you$_0$, hear$_0$, before$_0$, Bill$_0$, reviewed$_0$}

MERGE(Bill, VP2)
MERGE(before, VP2)



d) NUM = {which$_0$, tapes$_0$, did$_0$, you$_0$, hear$_0$, before$_0$, Bill$_0$, reviewed$_0$}

MERGE(VP1, ADJUNCT)
MERGE(did, VP1)
COPY([which tapes])
MERGE([which tapes], CP)



Notice that even after the two subtrees (VP1 and the adjunct) are merged, Form Chain still cannot apply since the copies of *which tapes* are in separate subtrees (i.e. they are still not in a c-command configuration). Once the application of Copy (in (d)) applies to create the item that merges to form the specifier of the matrix CP, then a c-command relation is established between the matrix CP and each of the lower copies of the *wh*-phrase. Form Chain then applies twice, creating two chains, one between the matrix Spec-CP and the complement of the matrix verb, and one between the matrix Spec-CP and the *wh*-phrase in the adjunct.

Now the tension between the Copy Theory of movement and the LCA comes into play. If the PF order of terminal items is mapped from the asymmetric c-command relations that exist in phrase-markers, then copies create a problem since they are items that literally c-command (other instances of) themselves. That is, given that the *wh*-phrase occupies multiple positions in the structure, which copy gets 'pronounced'? Nunes argues that the correct deletion of copies follows from economy—the upper-most copy of both chains has checked/erased features that the lower copies have not, thus the Chain Reduction procedure deletes the lower two copies yielding the (PF) output *which tapes did you hear before Bill reviewed.*

Notice again that it is the decomposition of 'movement' into the distinct sub-operations of Copy, Merge, Form Chain, and Chain Reduction that allows Nunes to exploit sidewards movement—once an item is copied, it is no different from an item that has been selected from the numeration in that it can then be merged anywhere else—even in a separate subtree. This is possible because of intermediate multiple-rootedness (that is possible or necessary to separately construct the adjunct and the matrix VP) and the division between operations that 'displace' items (Copy and Merge) and those that code the (representational) relation between the source and target of the 'displacement' (Form Chain). Note also that the separate construction of the adjunct allows the addition of this subtree to respect the Extension Condition. The account he presents crucially relies on intermediate multiple-rootedness and a formulation of movement that allows such operations to be potentially interarboreal.

2.2 Dispensing with Chains Recall from §1.5 above the discussion of chains and Hornstein's (1995) All For One Principle.

AFOP: Every link in a chain meets the morphological conditions satisfied by any link in a chain.

This principle allows for the tail of a CHAIN to have its CASE features checked if the head of that CHAIN has entered into a checking configuration. Put another way, CHAINS check features under this view, not individual copies/links of CHAINS—it is in virtue of having a copy/link in a certain (checking) configuration that a CHAIN has its features checked. Given the AFOP, *any* copy in a CHAIN can safely delete and Hornstein's A-movement analysis of quantifier scope ambiguities can be maintained.

It is against this background that Hornstein (1996, 1998) examines the question of whether or not CHAINS are dispensable. Given the above discussion of the AFOP, it would seem that CHAINS are absolutely crucial if one wants to hold onto the A-movement account of the quantifier scope facts discussed in §1.5, since what drives the analysis is the possibility of keeping *any* copy of a CHAIN and what allows that possibility is the AFOP. But this is arguably meaningless if the concept of CHAIN is abandoned—it is the very existence of CHAINS that allows all of the copies to have their features checked (in accordance with the AFOP) so that any of the copies can be retained to determine the scope possibilities. Nonetheless, Hornstein argues that CHAINS can indeed be dropped from the inventory of grammatical formatives given the assumptions that, (i) θ-roles are features that need to be checked (like CASE features) and that, (ii) lowering operations are permitted at LF.

Recall from above that in order to have the structural possibility of the object c-commanding the subject and therefore scoping over the subject, the lower copy of the subject (the one that is in its θ-position) must be the one that is retained while the retained copy of the object must be the upper one (the one that is in its CASE position). I repeat the selective deletion possibilities from §1.5 here:

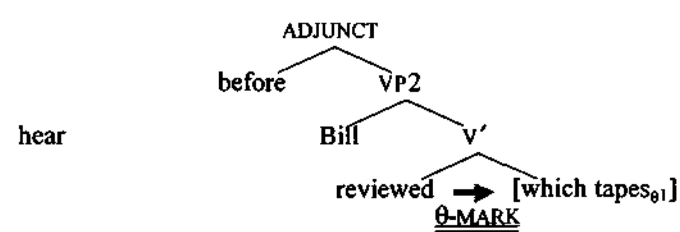
(42) (a) [**everyone**...[...**someone**...[...~~everyone~~...[...~~someone~~...]]]]
(b) [**everyone**...[...~~someone~~...[...~~everyone~~...[...**someone**...]]]]
(c) [~~everyone~~...[...~~someone~~...[...**everyone**...[...**someone**...]]]]
(d) [~~everyone~~...[...**someone**...[...**everyone**...[...~~someone~~...]]]]

This is possible without CHAINS if the subject is allowed to *start* in its CASE position and *lower* to its θ-position. In order for such a movement to conform with GREED and/or LAST RESORT it must be the case that θ-roles are formal features that must be checked.[35] In addition, this movement must be covert (in the LF component) since it violates the Extension Condition.[36] Given these assumptions it is possible, the argument goes, to abandon the concept of CHAIN—items move to check features, either raising or lowering, and the copy that has all of its features checked is the one that is retained—all others delete to rid the derivation of their unchecked [-interpretable] features.

It is worth pausing for a moment to note that this discussion of Hornstein's enterprise is important to the present project. Consider the AFOP for instance. The *Coresidence* view of movement introduced above suggests a different view of this principle: the *All IS One Principle*. Movement, under the Coresidence view, is just (Re)Merge. All the 'copies' are trivially satisfied by the satisfaction of any one of the 'copies' since there *are no copies*—its all the *same item.*

Now let's turn to Hornstein's treatment of parasitic gaps and adjunct control given this agenda to eliminate CHAINS.

(50) Which tapes did you hear before Bill reviewed?
(51)
a) NUM = {which$_0$, tapes$_0$, **did$_1$**, **you$_1$**, hear$_0$, before$_0$, Bill$_0$, reviewed$_0$}



◊ **Note**: ***hear*** **has been selected and not merged;** ***which tapes*** **'picks-up' the θ-role**

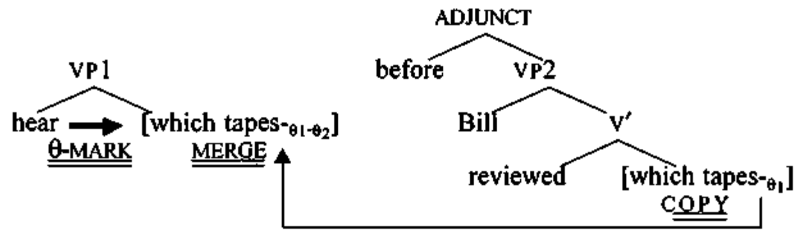
---

[35] Contra the assumptions of Chomsky (1995: chap 4): "Under any approach that takes Attract/Move to be driven by morphological features--whether Move F, Move α, and Greed, or some other variant--there should be no interaction between θ-theory and the theory of movement. θ-roles are not formal features..." (p. 312). This issue is of central concern since theAM/MSO system presented here requires 'lowering' operations to be quite general. Under standard assumptions about the separation of the domains of CASE assignment from those of θ-assignment, lowering of this sort could only be to θ-positions. To square such a system with the guiding ideas of the MP, it must be the case that the need to assign θ-roles drives movement.

[36] That is, it requires Copy and Merge to a position that isn't the root of the structure. Hornstein follows Chomsky in assuming the LF movement is not regulated by Extension. Note that this must be the case if the Spec-Head theory of Case checking is correct for English—if objects must covertly raise to Spec, Agr$_O$ (violating Extension) in order to check accusative case.

b) NUM = {which$_0$, tapes$_0$, **did$_1$**, **you$_1$**, hear$_0$, before$_0$, Bill$_0$, reviewed$_0$}

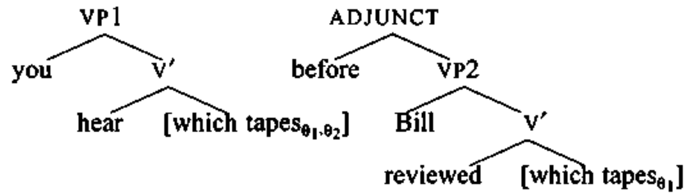
COPY([which tapes-θ_1])
MERGE([hear], [which tapes-θ_1])



◊ **Note: *which tapes* 'picks-up' a second θ-role from *hear***

c) NUM = {which$_0$, tapes$_0$, **did$_1$**, you$_0$, hear$_0$, before$_0$, Bill$_0$, reviewed$_0$}

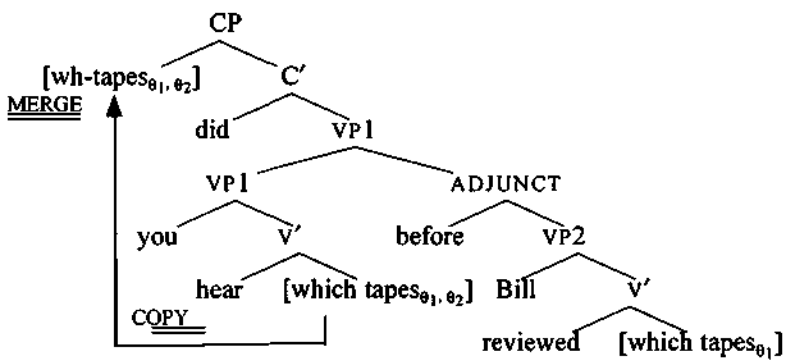
MERGE([you], VP1)



◊ **Note: The rest of the derivation goes as with Nunes (above) except that no CHAINS are formed and the *wh*-phrase 'carries with it' its θ-roles so that the 'top'/matrix copy has the θ-roles/features of both verbs**

d) NUM = {which$_0$, tapes$_0$, did$_0$, you$_0$, hear$_0$, before$_0$, Bill$_0$, reviewed$_0$}

MERGE(did, VP1)
COPY([which tapes$_{\theta_1,\theta_2}$])
MERGE([which tapes$_{\theta_1,\theta_2}$], CP)



The basics of the derivation are essentially the same as Nunes's analysis, except that the *wh*-phrase 'starts' in the adjunct, is copied and merged to the matrix verb, and then copied again to Merge to form the Spec-CP. Importantly, recall that Hornstein regards θ-marking as a process whereby items 'pick-up' θ-roles from θ-assignors. As a result, the upper-most copy of the *wh*-phrase contains all the θ-information necessary for its interpretation—that is, all the θ-information that *would be* associated with this item via CHAINS as in the Nunes analysis. And, like the Nunes story, the lower copies must delete because they have unchecked features while the upper-most copy is retained both at LF and for PF linearization purposes. But here, this is not a condition on CHAINS—this process applies deterministically for the phrase-marker—any item with unchecked features is simply deleted.

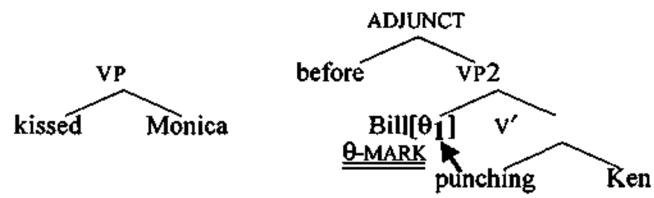
Notice that the θ-criterion has been essentially cut in half in that the bi-uniqueness of θ-assignment has been abandoned. While θ-assignors will be required to discharge their roles, there is no requirement in Hornstein's system that they do so such that there is a one-to-one correspondence with between roles and arguments. So, crucially, (i) θ-roles serve to 'drive' movement and, (ii) an item can receive, in principle, as many θ-roles as it likes.[37]

The same basic logic is extended to an account of Subject Control into adjuncts—Hornstein (1996) also turns a minimalist eye toward the Control Module of GB and examines the possibility of eliminating this sub-system of the grammar in favor of an A-movement account of the distribution of obligatory-controlled PRO.[38] Consider the following case of a controlled PRO in an adjunct:

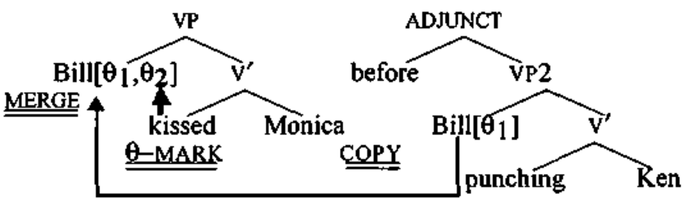
(52) Bill$_i$ kissed Monica$_j$ [before PRO$_{i,*j}$ punching Ken]

Hornstein treats this as a case of (sidewards/interarboreal) A-movement. The derivation is illustrated in (53). Like the parasitic gap cases, the adjunct and the matrix clause are separately assembled:

(53)
a) NUM = {kissed$_0$, Monica$_0$, before$_0$, Bill$_0$, punching$_0$, Ken$_0$}



b)
COPY(Bill[θ$_{punch}$])
MERGE(Bill[θ$_{punch}$], VP)



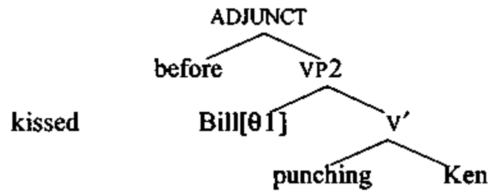
The VP is then merged with the adjunct and the derivation continues (the subject overtly raises further to check CASE, etc.). As with the parasitic gap construction, the sidewards moved item (*Bill*) 'carries with it' the θ-roles received in the adjunct and in the matrix. The copy of *Bill* which remains in the adjunct is deleted since it has not checked CASE, but this is unproblematic since the upper copy has derivationally 'acquired' all the thematic information necessary for interpretation. Further, Hornstein derives the impossibility of the object controlling into adjuncts from the economy condition PROCRASTINATE ("Avoid Overt Movement"—also discussed below).[39] Essentially this condition enforces a merge over move preference—consider the derivation of (53) at the point before *Monica* has been selected from the numeration and merged as the matrix object:

[37] To illustrate, consider:
(i) John expected to want to consider to promise Mary to go
(ii) John expected (J) to (J) want to (J) to (J) promise Mary (J) to (J) go.
(i) would follow from a derivation like (ii), where the (J)'s are deleted copies of *John*, which has A-moved from θ to θ position 'picking-up' all the external θ-roles in the sentence (leaving 'PROs' behind--now understood as mere copies = traces).

[38] For the details and accompanying argumentation, the reader is again referred to Hornstein (1996a,b). This view of Control manages to derive the core properties associated with obligatory-control PRO by identifying it as just an instance of A-movement. The basic intuition will be made clear in the discussion of the example that is relevant here--control into adjuncts (see previous footnote also).

[39] As Hornstein points out (personal communication) however, the generalization that only subjects can control into adjuncts might be incorrect depending on the status of examples like:
The police$_i$ arrested John$_j$ for PRO$_{*i,j}$ driving drunk.

(54) NUM = {kissed$_0$, **Monica$_1$**, before$_0$, Bill$_0$, punching$_0$, Ken$_0$}



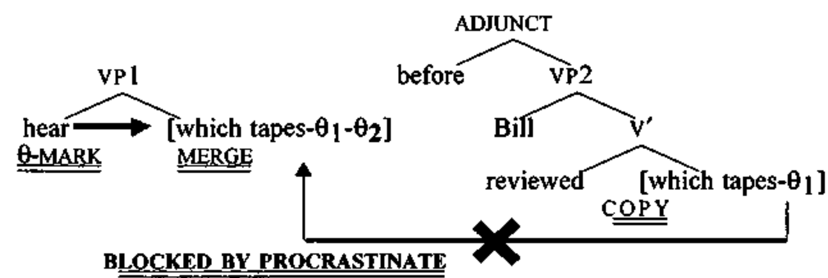
◊ Note: ***kissed* has been selected and not merged**

PROCRASTINATE forces *Monica* to be selected and merged, blocking the 'more costly' alternative derivation involving the copy and merge of *Bill*. The only possible continuation of the derivation that this economy condition allows at the point depicted in (54) requires that *Monica* become the object of the matrix verb—then, *Bill* can be copied and merged to form the matrix subject.

There is a crucial difference between Hornstein's analysis of adjunct control and his analysis of parasitic gaps in that the latter seems to entail a violation of PROCRASTINATE—it is exactly the object position of the matrix clause that sidewards movement targets in the derivation of parasitic gap constructions. Consider the relevant point in the derivation discussed above (51b):

(51)
b) NUM = {which$_0$, tapes$_0$, **did$_1$**, **you$_1$**, hear$_0$, before$_0$, Bill$_0$, reviewed$_0$}

COPY([which tapes-θ_1])
MERGE([hear], [which tapes-θ_1])



Since PROCRASTINATE forces Merge over Move the Copy and Merge of the *wh*-phrase should be blocked in favor of Selecting and Merging *you* to become the object of *hear*. Thus the derivations Hornstein proposes underlie parasitic gaps must allow a violation of PROCRASTINATE. Hornstein *does* provide a principled way to distinguish (with respect to economy) the adjunct control derivation from the derivation of parasitic gap constructions which permits the economy violation entailed by the latter. This will be returned to in §3 in the course of a discussion of how the AM/MSO system proposed in this investigation handles these cases.

A few points are worth highlighting before we move on. Like Nunes, Hornstein crucially exploits intermediate multiple-rootedness and interarboreal operations in his account of both parasitic gaps and control into adjuncts. His model manages without having CHAINS as grammatical formatives, but at an interesting consequence. Without CHAINS in the grammar Hornstein must abandon the traditional construal of θ-theory, and he must have the theory of movement crucially interact with θ-theory in that movement can be driven by thematic requirements.

Another consequence of eliminating CHAINS is that the resulting theory seems forced into taking the 'copies' in the Copy Theory of movement quite seriously in the following sense. The operation Copy creates items that have, as it were, their own independent derivational reality—they can check features, be assigned properties, and generally enter into relations independently of the items that these copies *are copies of*. Consider CHAINS and copies for a moment. One way of viewing the issue of whether or not CHAINS should exist as formatives is in the context of a strong reading of Chomsky's principle of INCLUSIVENESS—if we are unprepared to say that CHAINS are lexicon items, then they are an instance of the computational system *adding* something in the course of the derivation. Thus this strong reading of INCLUSIVENESS agrees with going the direction of Hornstein's 'chainless' system. But couldn't we level the same argument against 'copies'? Here arguably is also an instance of the system *adding* something in the course of the computation. But, the counter-argument would go, copies *are* lexicon items, they are simply a by-product of the basic operations of the system—a reflex of the need to have *one* thing enter into relations in *two* distinct positions. This line of reasoning underlies the Copy and Delete approach and has an consequence for how we view derivations with respect to numerations—that is, this view makes a crucial distinction between 'copies' and 'originals'. Two instances of an item α in a structure can correspond to either one numeration item (i.e. one of the instances of α is the result of Copy) or to two numeration items (Copy has not applied). I will return to this discussion below. For now, I simply want to call attention to the status of copies in Hornstein's (and Nunes's) analysis, and note the crucial role that they play in derivations. The worry in later discussion will be that a formal metaphor (copy) has been thrust into a substantive role in the grammar.

2.3 Problems We turn now to the status of economy/optimality principles in the Minimalist Program with an eye to how they might be effected given the derivations required by Chomsky's Extension-regulated derivations (§2.3.1). Then in §2.3.2, we turn to some troublesome cases given the economy considerations.

2.3.1 Economy Linguistic theory, specifically syntactic theory, has recently (re)turned to a careful scrutiny of its foundations. The agenda is "minimalist" in *two* senses (helpfully teased apart by Uriagereka & Martin, forthcoming). The first is essentially a house-cleaning mission—a rededication to a familiar positive heuristic—call it "Seek Parsimony". We witness instances of this in reasoning like, "if we need at least two levels of representation then we should have at most two" or "why have *both* X-Bar theoretic primitives *and* second-order structural relations like 'government'?" or for that matter "can X-Bar theoretic predicates be cashed out for something *more basic*?"

Hornstein (1997) suggests that this is the natural consequence of the success of GB in the 1980's—that although there is disagreement on the details, there *is* relative agreement on the overall 'shape' of what a theory that meets the requirements of explanatory and descriptive adequacy should like. So, the reasoning goes, *now* is a good time to think about what the *best* theory of that kind should look like. Whatever the reasons, there is a general push (unfortunately sometimes only in the rhetoric) to question fundamentals. These concerns are essentially meta-theoretic—they impact on our methodological decisions, suggest a spartan ontology of primitives, and in general serve as constraints on the sorts questions and answers we can/should entertain. Thus this first type of minimalism is what Uriagereka & Martin refer to as *Methodological Minimalism*, or "business as usual"—*ceteris paribus*, we take home the simpler theory. These sorts of considerations have driven most of what this paper has attempted to do so far.

The second sort of "minimalism" is more obscure. It is the substantive, empirical claim that the *system under investigation*—the computational system of human language (C_{HL})—is one that optimizes resources in some way. The claim is, for the moment, "quite vague", but the intuition is that C_{HL} minimizes 'computational effort'; that the basic operations of the grammar always function to take the lowest-energy route through the space of possible states.

One can imagine a number of ways to state grammatical principles that would be responsive to the economy intuitions. For example, we could think of economy/optimality principles as being sensitive to the number or relative 'cost' of the procedures used to transverse various regions of a space of possible derivations. One way to implement this idea is with the notion of comparison—economy conditions could be stated in a way that they evaluate various derivations (perhaps pair-wise) and enforce some criteria of optimality. Chomsky (1995) introduces just such a notion of comparison, with various restrictions on what sort of things count as comparable. For example, he proposes that economy conditions be restricted to only comparing *convergent* derivations. The reason for this is readily apparent. If the operations of C_{HL} are 'optimal' in the sense that fewer, less costly procedures are favored, then, as Chomsky (1995: p221) puts it, "the most economical derivation, then, applies no operations at all to a collection of lexical choices and thus is sure to crash." That is, if "Do Nothing" is a derivational option that is allowed to compete with other derivations which assemble lexical items into a syntactic complex and involve movement operations, then these latter more costly derivations will always be blocked by crashed derivations that incur zero cost in virtue of applying no operations. Thus, for example, a principle like PROCRASTINATE is proposed as a condition that only evaluates derivations that satisfy the requirements of the interface levels (PF, LF).

Chomsky further proposes that the class of derivations that can compete—call this class the Reference Set ($\mathfrak{R}$)—be even further restricted by only allowing competition between derivations that implicate that same set of lexical choices. This is implemented in Chomsky (1995) with the introduction of Numerations (NUM). As mentioned above, formally these objects are viewed as multi-sets of lexical items which encode the number of tokens of each lexical type that will enter the derivation.

> Let us take a *numeration* to be a set of pairs (LI, *i*), where LI is an item of the lexicon and *i* is its index, understood to be the number of times that LI is selected (p 225).

Importantly, this object is not viewed as a level of representation—there are no grammatical principles that are taken to hold over NUM—it is merely a formal device for delimiting the subpart of the lexicon that is involved in a given derivation.

A third restriction is imposed. For example, the exclusion of the possibility that the system could compare *Bill loves Monica* and *Monica loves Bill* for economy purposes is argued to be necessary.[40] Also, for a NUM with ***n*** lexical items there will be—in the limit—***n*!** possible SELECT&MERGE orders and therefore at least ***n*!** possible derivations to compare.[41] The space of possible combinations becomes quite intractable if MOVE and MERGE operations are allowed to interleave— "complexity too vast to be realistic" in Chomsky's words (p 227). Thus $\mathfrak{R}$ is more severely constrained by only allowing comparisons between derivations that have identical numerations *at each point*.[42] That is, at some derivational timestep **t**, two derivations are comparable if and only if their numerations are identical *at that timestep*. Thus derivations which enter into competition for economy purposes must involve: (i) the same lexical choices, and (ii) the same (derivational) order of selection of those choices. As Chomsky puts it:

> At a particular stage Σ of a derivation, we consider only continuations of the derivation already constructed—in particular, only the remaining parts of the numeration N. Application of the operation OP to Σ is barred if this set contains a more optimal derivation in which OP does not apply to Σ.

One hallmark empirical case involves existential constructions and the derivational timing of the merge of expletive *there*. Consider the familiar (55a) & (55b):

(55a) *There seems a man to be in the room
(55b) There seems to be a man in the room

These two examples are argued to represent the outcomes of two derivations that are alternate paths through the space of possible derivations which share the same lexical choices—both of which converge. Now consider how the comparison of derivations technology is deployed in an account of the contrast between (55a) and (55b). At timestep **t** the derivations and their respective numerations are identical:

(55a) & (55b) **NUM** = {$\mathbf{there_1}$, $\mathbf{seems_1}$, to_0, be_0, a_0, man_0, in_0, the_0, $room_0$}

[to [be [[a man] [in [the room]]]]]

[40] Or consider "That John likes sweets proves that John ate the pie" and "That John ate the pie proves that John likes sweets",...and one can imagine arbitrarily more complicated cases that arguably shouldn't enter into comparison. There is of course, the possibility that such cases *do* enter into comparison and are simply equal with respect to economy.

[41] See Johnson & Lappin (1997) for discussion.

[42] This is assuming, importantly, that we *should* care about the 'computational complexity' in the grammar—that such complexity is undesirable.

But at timestep **t+1**, Move applies in (55a) displacing *a man* while in (55b) *there* is Selected from NUM and Merged.

(55a) **NUM** = {**there$_1$**, **seems$_1$**, to$_0$, be$_0$, a$_0$, man$_0$, in$_0$, the$_0$, room$_0$}

[[a man]$_i$ [to [be [[e]$_i$ [in [the room]]]]]]

(55b) **NUM** = {**there$_0$**, **seems$_1$**, to$_0$, be$_0$, a$_0$, man$_0$, in$_0$, the$_0$, room$_0$}

[there[to [be [[a man] [in [the room]]]]]]

Assuming that there is a cost difference between Move and Merge (where Move involves Merge as a subpart), PROCRASTINATE insists that the less costly derivational route taken by (55b) blocks the more expensive option of (55a).

One way to sum up Chomsky's proposed understanding of comparison and reference sets is to say that only derivations which correspond to identical sequences of Select operations can be compared for economy purposes. Non-identical sequences of applications of Select yield non-identical numerations and non-comparable derivations. But now recall the basic logic behind intermediate multiple-rootedness—the extent to which the system requires it is the extent to which the selection of lexical items from a numeration *must* occur without a 'derivationally immediate' application of Merge. Notice that the first step of *any* derivation involves an application of Select with no accompanying Merge operation. This can be said to follow from 'virtual conceptual necessity'—if Merge is binary (see Kayne 1984, Collins 1995. Groat 1995) Select must apply twice before there are two items to be merged—derivations have to 'begin'. The question is whether it makes sense to, or whether it is necessary, allow this dissociation to carry through to *any* application of Select at *any* point in a derivation. For instance, one could argue that while the first lone application of Select follows from 'virtual necessity', the most restrictive system wouldn't permit this to be a generally available option. That is, it doesn't follow from virtual conceptual necessity that Select and Merge are *generally dissociable*; nor does this possibility follow in any obvious way from bare output conditions—in fact, one might argue that immediate merge of selected items is mandated by a need to create a single-rooted object. However, as demonstrated above, a dissociation of these two operations *does* follow from the assumption that structure building proceeds bottom-up obeying the Extension Condition—intermediate multiple-rootedness is a necessary feature of such derivational systems. The question is, then, does anything force Select operations to generally be paired with Merge operations? The reason this is even an issue is that the system forces these operations to be derivationally dissociated to build complex non-complements—what happens if this dissociation is possible in the limit?

Well, one thing that happens is it renders Chomsky's proposed instantiation of reference sets completely useless. If Select can be generally dissociated from Merge, then there is nothing forcing the sequence of Merge operations to track the sequence of Select operations—two derivations, under this view, could exhaust their numerations via identical sequences of Select operations—therefore ensuring that they are in the same reference set, but Merge the selected items in completely different orders. Conversely, two derivations could differ in their sequence of Select operations but be identical in their sequence of Merge operations and *not* be compared. This suggests that either:

- It is not the sequence of Select operations that establishes comparability, OR
- There is some restriction on dissociating Select and Merge (that it only happens when necessary)

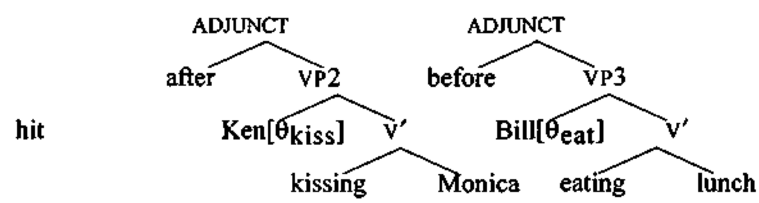
The former suggestion would amount to reformulating the notion of reference set so that it is actually *sequences of Merge operations* that determine comparability. This would remove the potential problem of the 'in the limit' dissociation of Select and Merge, but it doesn't banish the worries entirely. That is, nothing requires that, for example, a series of complex adjuncts be constructed in a certain order, or the matrix VP before a potential specifier, and so on. If this is so, then having sequences of Merge operations determine reference sets will have the effect of 'overpartitioning' the space of derivational possibilities, resulting in both (i) rendering desirable comparisons non-comparable, and (ii) making competitors out of derivations which shouldn't be. Consider again the discussion from §1 of the simple example—*the man saw the woman*. Recall that Extension insists on the separate construction of the specifier and the rest of the VP. Now consider two possible sequences of Merge operations—in one derivation, the specifier is built first; in another, it is built second. This should have the counter-intuitive and potentially problematic consequence of rendering these two sorts of derivations non-comparable for economy conditions. And of course, this is a perfectly general worry—it will arise every time the system needs to construct a complex non-complement.[43]

2.3.2 Constraining Sidewards Movement In addition to the above difficulties, allowing interarboreal operations complicates matters even further. For example, consider (56) and Hornstein's analysis of subject Control into adjuncts discussed above.

(56) Bill_i hit Ken_k after $\text{PRO}_{i,*k}$ kissing Monica before $\text{PRO}_{i,*k}$ eating lunch.[44]

Recall that for Hornstein, Control is just an instance of (in this case, 'sidewards') A-movement so that the 'PROs' in (56) are just copies of *Bill*. Recall the logic from which the necessity of subject control was taken to follow from—that the adjunct PRO's can *only* be copies of *Bill* since PROCRASTINATE prevents the movement of *Bill* to the object position of the matrix verb in favor of the select and merge of *Ken*. This of course is the desired result since the matrix object cannot 'control' the 'PROs' in the adjuncts. Now consider the derivation in (57).

(57)
a) NUM = {kissing_0, Monica_0, before_0, Bill_0, eating_0, after_0, lunch_0, hit_0, Ken_0}

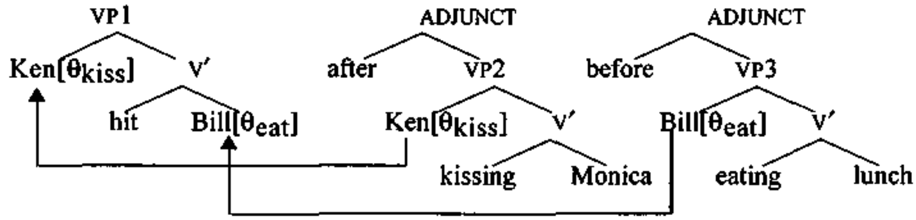


At this point both of the adjuncts have been separately constructed and the numeration has been exhausted (again, irrelevant functional material is omitted). Notice that *Ken* has been selected and merged in the first adjunct. Following the same logic that Hornstein deployed to account for the impossibility of the matrix object controlling into the adjunct, *Bill* cannot be copied and merged to receive the θ-roles in both adjuncts—this option is blocked by PROCRASTINATE, which favors the less costly option of selecting and merging *Ken*. There is nothing that prevents this derivation from next copying and merging both *Ken* and *Bill* to become the subject and object of the matrix clause as in (57b):

[43] What this might be suggesting is that economy conditions are not 'global' in the sense that what they make reference to are entire derivations. Economy conditions might be local conditions that insist on the 'least effort'/optimal construction of some natural class of subtrees. A likely candidate for this class of substructures are what Uriagereka (1996) calls 'command units'--those subparts of the phrase-marker for which c-command is a (near-) total relation for terminals. Also, there have been some recent suggestions that the way to properly state economy conditions is to place restrictions on either the construction of or access to numerations. Chomsky (1998) has suggested the latter view under which numerations would be cyclically accessed up to convergence. Uriagereka (1998, and personal communication) has suggested the former view which reintroduces something very close to the old notion of a 'kernel sentence' of *LSLT*. Castillo, Drury, and Grohmann (1998) argue for a similar view, but one that relativizes the notion of convergence to something akin to kernel sentences (what we called T-Domains) instead of placing restrictions on either numeration building or on numeration access.

[44] This example was brought to my attention by Juan Carlos Castillo and Matt Kaiser.

(57)
b) NUM = {kissing$_0$, Monica$_0$, before$_0$, Bill$_0$, eating$_0$, after$_0$, lunch$_0$, hit$_0$, Ken$_0$}



In this derivation the matrix subject and object each control a 'PRO' in the respective adjuncts yielding the impossible interpretation that *Ken* is the 'kisser of Monica' and *Bill* is the 'eater of the lunch'. Even if some alternative formulation of PROCRASTINATE were advanced so that *Bill* could legitimately be copied and merged to become the subject in both adjuncts, there is nothing that rules out the alternative derivation (57).

Further, consider the consequences of the worries discussed in §2.3.1. Consider an alternative derivation to (57′) in which the matrix verb and object were selected and merged *before* the adjuncts were constructed. Then the only possible continuation would be the desired one, where whatever becomes the subject of one adjunct must be Copied and Merged to become the subject of both the other adjunct and the matrix clause. But if there is no constraint on intermediate multiple-rootedness, this alternative (57′) won't outrank the derivation in (57) for one of two reasons. If (57) and (57′) are comparable, they will involve the same operations—no matter what there must be two sidewards movements. If these two derivations are not comparable, then either possibility should be permitted, and (56) is predicted to be ambiguous. Comparability, recall, is determined by the sequence of select operations. But if lexical material can be selected and not merged, then the numeration in (57) can correspond to both sorts of reference sets—those which have both (57) and (57′) as members, and those that do not. Either way, it seems that regardless of how the economy considerations are formulated, the possibility of having multiple subtrees in a workspace permits sidewards movements that yield the wrong results—there needs to be some restriction on how such derivations proceed.

Notice that this problem doesn't go away given the restatement of comparability in terms of sequence of Merge applications. Once an item has been merged into one of the adjuncts, it will always be cheaper to merge another item into the second adjunct than it will be to (sidewards) move into the other adjunct. The only route this derivation can take is to construct the matrix VP first, then there will only be one available item to become the PRO subject of both adjuncts and the matrix. But in virtue of requiring distinct sequences of Merge operations, this (desired) derivation will be non-comparable with the one in (50) and thus this case is still predicted to be ambiguous. But note that the only way for this derivation to even arrive at the correct structure requires building *the matrix* first.

Another complication that arises from interarboreal movement is how locality conditions are to be stated. Notice that if there is no constraint that insists that when an item is copied it must be immediately merged, then what prevents, for example, a *wh*-item *inside* a syntactic island from being copied and merged to a position *outside* the island? Call this the 'copy and wait' strategy. An answer that has been given is that this is ruled out by some representational requirement which regulates either the formation of CHAINS or the deletion of copies—for example, Nunes's Form Chain operation would rule out potential island violations of this sort. Or consider the adoption of Rizzi's Chain Condition suggested by Bobaljik & Brown (1997:fn.39) which imposes the same sort of representational requirement to prevent violations of the HMC (and which also rules out sidewards movement of the Nunes sort altogether).

This 'copy and wait' worry is especially problematic for the 'chainless' system proposed by Hornstein. What might prevent the circumvention of islands in his account? One possibility is to introduce another kind of economy condition that assigns a cost to maintaining multiple unconnected subtrees. This sort of constraint would be a generalized version of what Yang (1996) calls AS SOON AS POSSIBLE (ASAP). The idea would be that, although Select and Merge are necessarily dissociable, this dissociation is regulated by economy:

> ASAP: Applications of Select must be immediately followed by associated applications of Merge wherever possible; where immediate Merge is impossible, Merge as soon as possible.

A condition of this sort could potentially solve the problems that unconstrained intermediate multiple-rootedness pose for Chomsky's comparison mechanisms. It could also rule out the 'copy and wait' strategy for extracting items from inside islands if it was taken to hold over the potential dissociation between Copy and (derivationally immediate) Merge. Notice also that this ASAP strategy could potentially solve the multiple adjunct problem for the Hornstein system.[45]

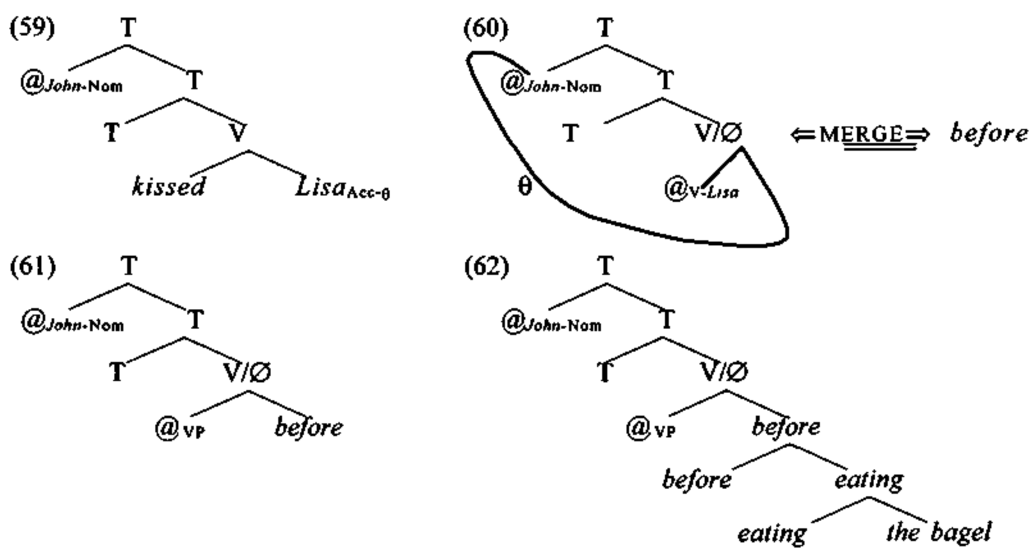
But, now a problem arises. If the problem raised above for multiple adjuncts/multiple adjunct control is solved by the introduction of something like a 'Merge ASAP' condition, the following should be impossible:

(58) Which women$_i$ did Ken$_j$ realize Bill$_k$ met t$_i$ [before PRO$_{j,k}$ kissing pg$_i$]

In (58) the 'PRO' in the adjunct can be controlled by either subject. This, given the ASAP strategy that would fix the problem in (57), should be impossible. If the adjunct is unable to 'wait around' for the rest of the structure to be built, then given the PRO in the adjunct should only be controllable by *Bill*. Note that the adjunct must be built and crucially must be unattached at the point where the *wh*-item moves from the adjunct to become the object of *met*. But then according to something like the ASAP constraint on derivations, it shouldn't be able to remain un-merged until the matrix verb is encountered. If this kind of 'wait and see' is possible, the illicit derivation for multiple adjuncts in (57) should be allowed. If not, then control of PRO by *Ken* should be ruled out—and it is not. In (58) the PRO in the adjunct can indeed be controlled by either subject.

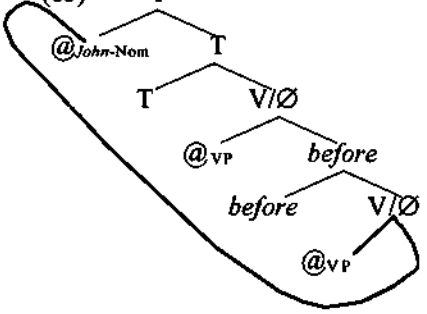
3 Reanalysis in the AM/MSO System

We turn now to a reanalysis within the AM/MSO system. Consider first the case of adjunct control. I will follow Hornstein's analysis and assume that the 'movement' treatment of these constructions (of Control in general) is essentially correct—that there is no grammatical formative 'PRO'. However, instead of sidewards movement out of the adjunct to the matrix subject position, the AM/MSO-syntax presented in §1 suggests that these constructions result from the (re)Merge of the matrix subject to the θ-assignor in the adjunct. The derivation for *John kissed Lisa before eating the bagel* goes as follows:



[45] Uriagereka (1998) discusses a similar idea under the label "Maximizing Minimal Resources".

(63)



In (59) and (60) the matrix is assembled (the DP *John* Spells-Out when it merges with T and the address/label is marked for Nominative Case; V is added and the object/complement is built, Case- and θ-marked). Then in (60) the subject re(Merges) with the verb-complement unit, and it is (compositionally) θ-marked (as discussed in §1.5). Step (60) depicts the Merge of the preposition *before* (perhaps complementizer, see Dubinsky & Williams 1995). The output of this operation is the Spelling-Out of the VP in (61). Then the adjunct is assembled as in (62), and the subject (strictly, the subject's address/label = $@_{John}$) (re)Merges to be assigned the θ-role discharged by the verb-complement unit in the adjunct.

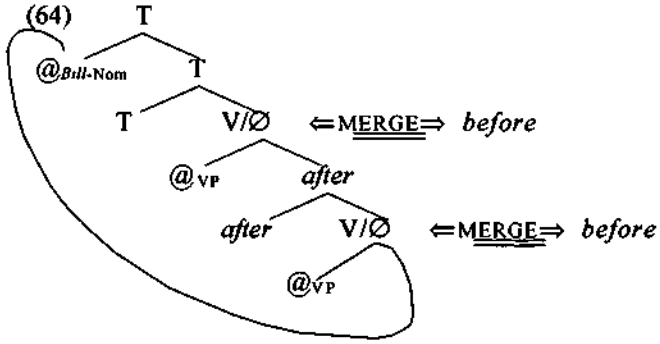
There are two important features of this derivation. First, given that the addition of the adjunct material causes the verb-complement unit to Spell-Out, it is expected that the object cannot (re)Merge to the θ-position of the adjunct. Because it has Spelled-Out with other material (i.e. the verb) it cannot enter operations as a unit excluding the verb from that point on. Thus this derivation predicts the impossibility of the object controlling the 'PRO' in these situations, exactly as desired.

Now consider the situation where there are multiple adjuncts as in the troublesome case (56) discussed above in §2.3.2, repeated here:

(56) Bill$_i$ hit Ken$_k$ after PRO$_{i,*k}$ kissing Monica before PRO$_{i,*k}$ eating lunch.

The crucial property is that *both* adjuncts contain PROs that *must* be controlled by the subject. Recall from above that in order to ensure this in Hornstein's system, something like a 'Merge ASAP' constraint had to be posited, forcing adjuncts to be integrated into the structure as soon as possible. Consider now the AM/MSO derivation for (56). In (64), the derivation has reached the point equivalent to the above derivation in (59)-(63), where the first adjunct has already been constructed:

(64)



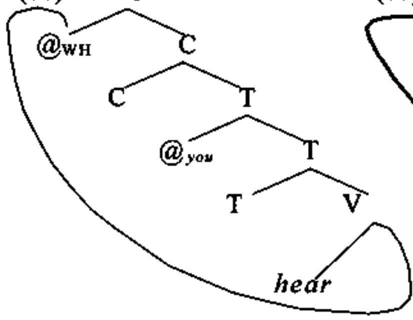
The step depicted in (64) is at the point where *Bill hit Ken after kissing Monica*, and as in the derivation (59)-(60), only the matrix subject *Bill* was able to (re)Merge to receive the θ-role assigned by the verb-complement unit *kissing Monica*. As in the simple case, the matrix verb-complement unit (*hit Ken*) has been forced to Spell-Out when the first adjunct was constructed. Now, in (64) the second adjunct can begin construction with the addition of the second preposition *before*. There are two conceivable places in the structure this item could target—the matrix VP (yielding something like a "after X AND before Y" reading) or the adjunct VP (yielding an embedded reading: [*after kissing Monica*[*before eating lunch*]]). But, **either way**, the only item that can (re)Merge to take on these θ-roles in the adjuncts is the matrix subject, for the same reason as with the simple case. Thus additional (economy) constraints on parallel sub-derivation interaction are unnecessary.

There is a trick here though. What the above derivations do, in effect, is allow the equivalent of A-movement out of an adjunct—a flagrant violation of Huang's (1982) Condition on Extraction Domains (CED).[46] Since this is a property shared by the AM/MSO derivations for parasitic gaps as well, we review these first and then return to the status of the CED in this system.

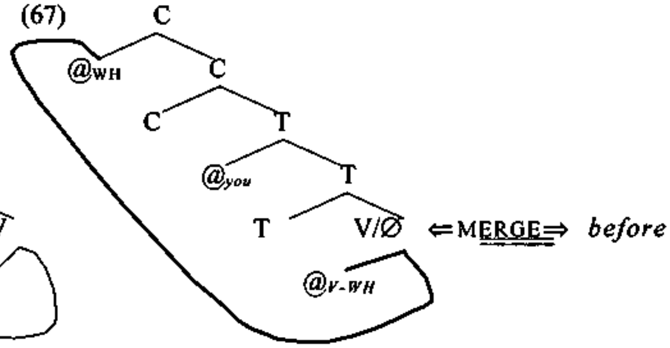
As with the adjunct control cases, I assume that the movement account for the parasitic gap constructions is generally correct insofar as these will be viewed as *movement*/(re)Merge and not something else (e.g. some sort of null resumptive strategy as in Cinque 1990). Consider first a basic case in (65) and the corresponding AM/MSO derivation in (66):

(65) Which tapes did you hear **t** before Bill reviewed **pg**

(66)

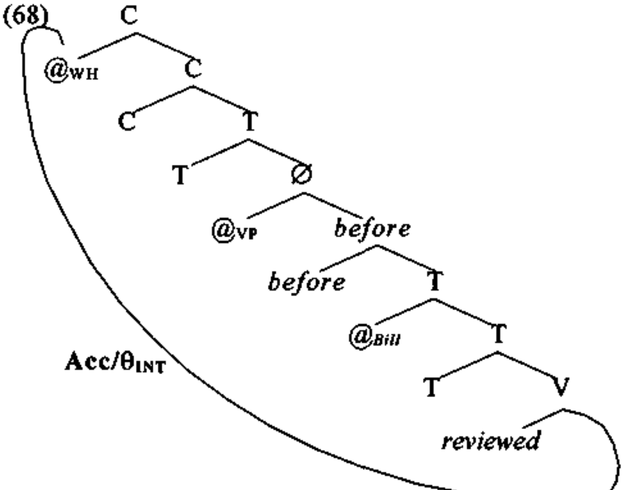


(67)



This far into the derivation, the matrix has been constructed—the *wh*-phrase has Spelled-Out and merged to C, the subject DP has Spelled-Out and Merged to T receiving Nominative Case, and the *wh*-phrase has (re)Merged to receive object Case and θ in (66). The next step involves the (re)Merge to the verb-complement unit by the subject label/address, thereby forcing the verb-complement unit to Spell-Out, generating the structure in (67). Then, as in the adjunct control case, *before* target the verb phrase (in (67)) causing the entire VP to Spell-Out. The adjunct is constructed as in (68) below, and the wh-phrase (re)Merges to the object position in the adjunct.

(68)



As with the AM/MSO derivation for the adjunct control cases, the CED is apparently violated. But there is an additional puzzle here, one that this system has in common with Hornstein's approach—it appears that *one* item is getting/checking *two* Cases. This is worrisome given the logic of Chomsky's (1995) checking theory—that checking essentially erases features so that items are expected to be 'frozen' in Case-checking configurations

[46] Or its equivalent. Whatever it is that disallows movement out of adjuncts generally. Here I will adopt a θ-theoretic condition, but essentially leave the matter open. I deal with the locality issues in more detail in Drury (forthcoming).

(unless they contain some other feature that requires checking elsewhere). But in the accounts offered by both Nunes and Hornstein, the *wh*-item that is copied presumably must satisfy/check both the matrix and embedded Cases in parasitic gaps constructions. The worry doesn't arise in the adjunct control cases since by hypothesis the infinitivals in the adjuncts don't assign any Case. But the worry stands for these parasitic gap constructions. Call this the TWO CASE PROBLEM.

Now let's return to the status of the CED in this system. I propose to adopt a component of the *Barriers*-style theories: L-marking. Recall from the discussion of Hornstein's system and the discussion of Uriagereka's Multiple Spell-Out, that what is allowing Hornstein to Copy & Merge out of an adjunct is the fact that the adjunct hasn't attached and therefore, according to Uriagereka's approach to strong islands, hasn't been forced to Spell-Out by linearization requirements. The question that we might ask is what the difference is between complements and adjuncts under this view. That is, *why* isn't it the case that adjuncts are built first in Extension-based systems, where the material they modify is added at the root? Put another way, what is the reason that it makes sense to build adjuncts 'separately'/not as complements? This is because adjuncts are typically not θ-marked. If they were, the reasoning would go, they wouldn't be separately constructed and wouldn't be forced to Spell-Out when they are attached in bottom-up, Extension-regulated systems. The point I'm approaching here is that there is a θ-theoretic condition lurking behind the linearization story alluded to above. That is, the structural motivation for Spelling-Out adjuncts is clear, but we might ask why they 'attach' the way they do. The answer is simply the argument/adjunct distinction. Thus, I propose here to embrace a θ-theoretic interpretation of this distinction and state the CED restriction *directly* as follows:

θ-Licensing: (re)Merge into a non-θ-marked domain is prohibited.

This bars movement into both subjects and adjuncts. The adjunct case is obvious, the subject case less so. Consider that under this θ-licensing constraint (again, really just the L-marking component of Chomsky 1986) subject, given that in an AM/MSO derivation these items aren't θ-marked until *after* they've Spelled-Out, (re)Merge into a complex subject should be impossible.

But now the AM/MSO derivations above are ruled out as they each involve (re)Merge into non-θ-marked domains. For now, I introduce the following stipulation:

θ-Tax: An item can (re)Merge into a non-θ-marked domain iff that item is already θ-marked.

This is to be thought of as a cousin to Richard's (1997) 'Subjacency Tax'. The idea is that a θ-assignor inside a non-θ-marked domain cannot see any higher than the roof of this domain in its search for an item to which it can discharge its θ-role. However, items outside this non-θ-marked domain are rendered visible to the θ-assignor if they are already θ-marked.

Thus, in the adjunct control derivation above, the only possible item that (re)Merge into the adjunct—the matrix subject—can do so because it has already been θ-marked by the matrix verb-complement unit, and is thus 'visible' to the θ-assignor in the adjunct. No issue arises with respect to Case, as already mentioned, since by assumption the [-finite] T in the adjunct does not assign any Case.

Now consider the parasitic gap derivation. The same logic with respect to the penetrability of the non-θ-marked domain holds here as well. The matrix *wh*-item, in virtue of it having already been θ-marked is a visible target outside this non-θ-marked domain, and thus can licitly (re)Merge into the adjunct.

But what about the TWO CASE PROBLEM? Suppose that under circumstances where CASE and θ are in one-to-one correspondence the CASE features can unify provided the two CASES *are identical*. Recall that the Root First derivations discussed above assumed that there is a difference between subjects and objects (in English) with respect to CASE and θ-theory—that complements receive their CASE and their θ-role (as a bundle) *in the same position*, but that the CASE and θ-requirements for subjects are met in distinct positions. Thus there should be a difference between objects and subjects with respect to the possibility of the CASE unification process suggested above. That is, *it shouldn't be possible for subjects*. This yields the desirable prediction of the impossibility of subject parasitic gaps. We will take CASE

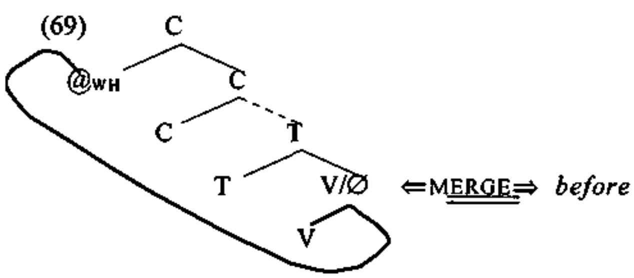
Unification to be stated as follows:

> <u>CASE UNIFICATION CONDITION</u>: If α is CASE- and θ-marked, α can relate to an additional CASE assignor β iff:
> (i) β assigns α a CASE that is identical to the CASE α already bears;
> AND
> (ii) β θ-marks α

Another way of putting this is to say that only identical inherent/thematic CASES can unify. The central intuition, following Uriagereka (1997), is that the function of CASE in the grammar is essentially to establish local 'distinctness'. Requiring this identity condition between CASES for 'additional' CASE assignment/checking is a kind of coherency condition in the spirit of UNIFORMITY of Chomsky (1995). If having received a CASE distinguishes an item from others, this parallelism/unification requirement amounts to saying that an item that has received two conflicting CASES is being distinguished from itself. This might be rendered coherent in a system that has adopted the Copy Theory of movement, but recall that the Re-Merge process assumed here results in a kind of *transpositional* object that co-resides in distinct structural locations—"movement" in this theory is just a process that creates such objects. What is viewed as the source versus the target of the movement, or the head versus the tail of a CHAIN is simply viewed as two structural 'vantage points' of the same lexical item—two places where the item 'participates' in the structure. Given this view of displacement, CASE conflicts are simply situations where an item has received impossible distinctness markings—additional CASES must be assigned such that they are consistent with the marking that the item already has. Condition (ii) of the CUC is rationalized as follows: I've been supposing that both CASE and θ can trigger Re-Merge. ("drive movement"). Assume that movement (Re-Merge) for θ-assignment is motivated by the needs of the target ("Enlightened Self-Interest") but that movement for CASE reasons is essentially motivated by the needs of the item to be Re-Merged ("Greed"). Once an item has received a CASE, it is satisfied with respect to this requirement. It can still be 'attracted' by an item that needs to discharge a θ-role, but not by an item that merely has a CASE to discharge. So essentially the CUC follows as a descriptive generalization from the hypothesized nature of one difference between θ and CASE—that the former is falls under Enlightened Self-Interest, while the latter falls under Greed.

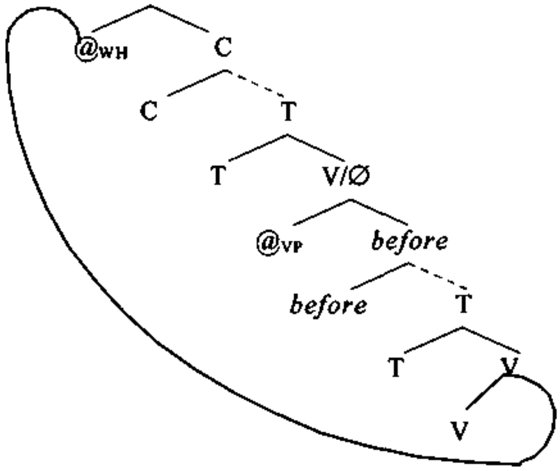
Let's examine again the AM/MSO parasitic gap derivation more closely with these considerations in mind (abstracting away from the earlier details):

(69)



This first instance of (re)Merge of the *wh*-item is licensed by the presence of the matrix verb and the θ-role that it has to discharge/assign. In virtue of this (re)Merge, the wh-item has received the (internal) θ-role discharged by the matrix verb and, importantly, Accusative CASE. When the first item of the adjunct (*before*) is added, everything below its merge-site spells-out. Then the adjunct is constructed as follows:

(70)



Now the generalization regarding potential CED violations becomes important—the wh-item, in virtue of its association to the CASE/θ position of the matrix verb, is licensed to associate inside the non-θ-marked domain of the adjunct—that is, as stated above, an item can only 'lower' into a non-θ-marked domain if it already has a θ-role.

Now consider the 'TWO CASE PROBLEM' again given the unification condition proposed above. The idea is that one and the same item can check both cases if and only if that item has been associated with a θ-role for each *additional* CASE that it checks. Notice that this is *not* claiming a bi-uniqueness of CASE and θ—an item can check as many θ-roles as it likes—but this bi-uniqueness does hold for such additional CASEs that an item checks—if that item checks one 'additional' CASE, then it must get one additional θ-role; if it checks two additional CASES then it must have two additional θ-roles, and so on. The CASE Unification Condition is essentially asserting that where an item has more than one CASE it must have more than one θ-role—multiple CASE requires multiple θ. And, given the discussion above regarding Greed/Enlightened Self-Interest, this essentially follows from the mechanics of Re-Merge. To illustrate a little better how the above system works consider some typical properties of parasitic gaps (adapted from Nunes 1995):

(71)
(a) Parasitic gaps are licensed at S-structure (Chomsky 1982)
(b) Parasitic gaps cannot be c-commanded by the 'real' gap (Taraldsen 1981, Engdahl 1983)
(c) Parasitic gaps are subject to island effects (Kayne 1983, 1984, Chomsky 1986a)
(d) Parasitic gaps can only be NPs (Chomsky 1986a, Cinque 1990)

The first property (71a) of these constructions follows immediately from the architecture assumed above. The relevant sort of cases are the following:

(72) *John filed every report without reading
(73) *Who filed which report without reading?

The point is that if parasitic gaps were licensed covertly (post S-structure/post Spell-Out/in the LF component) then the covert raising of the quantifier or the wh-item in the above examples should 'license' the gap following *reading*. Given the AM/MSO system proposed here, the objects of the matrix verbs in the above examples must Spell-Out when *without* is merged to the structure to begin the construction of the adjunct—thus this item will not be able to Re-Merge to become the object of *reading*. These basics also partly account for property (71b) above, the merge of the adjunct will always cause the verb and its complement to be Spelled-Out so the 'real' gap will always be in a separate command path as the parasitic gap. (71b) is also partly accounted for in terms of the conditions that insist that the gaps have the same CASE.

The reason that such gaps are 'parasitic' recall, is that the only way to associate into an adjunct (i.e. a non-θ-marked domain) is if the item that is trying to associate into the adjunct is already θ-marked. The fourth property also follows from the θ-theoretic nature of the account presented above—only D/NPs can have θ-roles, so they are the only things that could possibly be parasitic gaps. Since I do not have a full theory of locality developed for this system, property (71c) is currently not accounted for, but this is feature of the analysis that I have in common with the other approaches mentioned in this paper.

Now reconsider the puzzle we left for the Sidewards Movement analysis in §2. The relevant case was the following:

(74) Which women did John realize Bill met t [before PRO kissing pg]

Recall that the 'PRO' in this adjunct can be controlled by either *John* or *Bill*. Here this is handled straightforwardly. At the point in the derivation where the adjunct is added, *before* can either target the matrix verb *realize* or the embedded verb *met*. If the former, then *Bill* will be part of the unit that must Spell-Out and thus cannot reassociate into the adjunct, while both the *wh*-phrase and the matrix subject can associate into the adjunct—yielding Control and the parasitic gap. If *before* targets the embedded verb *met*, then *Bill*, being the closest potential already θ-marked target, will be the controller and the *wh*-item will (re)Merge as in the simpler cases to form the parasitic gap. Thus the puzzle that arose for Hornstein's system follows straightforwardly in the AM/MSO system without contradiction.

This section does not hope to have offered a complete account of parasitic gap phenomenon. What I do hope to have accomplished is rendered plausible a route toward a more complete account within the AM/MSO system. The concern in trying to construct a convincing sketch of an analysis here is to make sure that the cases which are handled by the interarboreal movement analyses have a potentially satisfying treatment without these operations. A fuller treatment of these analytical questions is offered in Drury (forthcoming).

It seems that there are some reasons for thinking that the possibility of tree-to-tree movement and the need to construct command units separately have undesirable consequences that the AM/MSO system does not. Notice that the problems raised for the Sidewards Movement account (in §2) that seem to arise from intermediate multiple-rootedness and interarboreal operations *do not arise* in the AM/MSO system—this is because the system does not require intermediate multiple-rootedness and thus does not require the necessary precondition for tree-to-tree movement—having multiple, unconnected subtrees during the course of the derivation. The system requires neither economy restrictions (like Procrastinate or Yang's ASAP) nor representational 'output' filters (such as chain formation conditions assumed by Bobaljik and Brown and by Nunes). In general, there is no dissociation between operations in the AM/MSO system. Select and Merge can be viewed as simply Merge, with none of the undesirable consequences with respect to economy conditions that were raised above. And, since movement in this system is viewed as just another instance of Merge (with no attendant operations of Copy or Delete or Chain Formation) the 'copy and wait' possibility also doesn't arise—and thus there is no reason to introduce mechanisms or conditions to rule it out—it follows from what is arguably the maximally simple assumption about structure-building operations—there is only *one* operation—Atomic/Simple Merge.[47] Of course, there is also Spell-Out, but this is a feature of any system that assumes anything like the T-model.

Conclusion

This has been a somewhat long story. At the risk of having buried the lead, here's what I think has been accomplished. The initial technical question regarding the difference(s) between

[47] This re-raises the issues surrounding Procrastinate--in this system there is no obvious Merge vs. Move 'cost' distinction, and therefore no obvious way to implement the intuition behind this economy strategy. I am pursuing the implications of various formulations of this economy condition (and the nature of economy conditions in general in joint work with Kleanthes Grohmann and Juan Carlos Castillo (Castillo, Drury & Grohmann, 1998). For reasons of space I will not pursue this discussion here. However, note that the removal of the cost distinction between Merge and Move that this system suggests might be the right thing to do. After all, there are only a small range of empirical cases which are responsive to this sort of treatment. One of them, Hornstein's explanation for the difference between parasitic gaps and adjunct control (and the impossibility of object control into adjuncts) has been dispensed with here. Another case in (55) above (§2.3.1) has been the subject of intense scrutiny the result of which has been any number of alternative analyses of those existential constructions which doesn't rely on economy arguments. To sum, its not entirely clear that Procrastinate is the right sort of economy condition to have. Put another way, if it *is* the right sort of condition, it doesn't have a natural implementation in the AM/MSO system that has been discussed here.

two kinds of derivational systems—those regulated by Extension versus the AM/MSO ("local merge") system proposed here—has driven a discussion about what sorts of features a derivational system could have that wouldn't have obvious 'recodability' in representational systems. The big suggestion here is that the real interest of derivational systems lies in the potential to create and then alter relations/structure from step to step. The "now-you-see-it-now-you-don't", dynamic view of constituency introduced by Colin Phillips is one such example; the factoring of derivations into sub-derivations introduced by Juan Uriagereka is another. This paper has suggested that these two ideas should be combined. Further, within such a system, there is a strong temptation to explicitly pin the asymmetric properties of command orders on the asymmetry of linear orders—that the hierarchical asymmetries are in a way 'parasitic' on the 'order of formatives'. Here the move has been to identify such orders with the order of merge/derivational combination.[48]

The status of 'movement'/'displacement' has been given a novel twist—the only operations of the system are Merge and Spell-Out. Copy, Deletion, and Form Chain have essentially been abandoned. Copy + Delete is translated here as (re)Merge following Spell-Out.

The differences between the chainless Copy-view of movement, representational Chain conditions, and the Coresidence view advanced here are unclear. The subpart of the theory dealing with displacement is a difficult place to find salient differences between alternatives. My suspicion is that the Coresidence view advocated here is really closer to the representational understanding of displacement—that these are cases of simple operations creating a 'transpositional object' (i.e. a chain). But here the creation of such objects has been crucially derivational and different from most representational formulations by following Hornstein and allowing multiple θ-marking (and multiple Case-marking).

Although not explicitly discussed above, the systems which exploit the intermediate multiple-rootedness and allow movement to mediate between disconnected subtrees are also examples of what I'm taking to be salient differences between derivational and representational systems. And, by incorporating Uriagereka's MSO, this system has essentially agreed that it is sensible to so factor derivations into sub-parts—but the system here manages to capture this property of derivations without ever having more than one object in a derivation. That is, in the bottom-up systems regulated by Extension, multiple roots are a consequence—you must construct subtrees separately and then merge these complex objects. This allows the potential—give a Copy+Merge view of displacement—to have operations that form dependencies across these distinct, unconnected objects (i.e. interarboreal operations/sidewards movement). What has been accomplished here is essentially a retention of such relations, but with a crucial difference—the movement dependencies are ***formed first***, then the object is 'factored' into subtrees (by Spell-Out). The Phillips-style derivations plus Uriagereka's MSO allows us to view things this way. Crucially, movement dependencies are never across unconnected objects. They will always be between c-commanding positions because in this system that's all there is. At any one point, there is only a single 'command unit'. In a sense, this paper ***agrees with*** the investigators who have noticed the intermediate multiple-rootedness and the potential for interarboreal/sidewards movements—but here, the establishment of the relation occurs first, then the factoring of the tree into separate sub-derivations. This allows us to state the familiar and sensible restrictions on displacement compared to the Copy+Merge view, where its not obvious how to "license" a 'copy' that is essentially targeting a projection that is part of a separate subtree.

A couple of final notes regarding things that should be looked into. Given the view on order proposed above, this system would do well to investigate instances of movement relations that traditionally have been viewed as 'rightward' movement. The worry for the present system is that given the Coresidence view of movement, cases where an item gets pronounces lower than its original Merge site raise interesting problems. Also, wh-in-situ and

[48] Note there is an unclarity above whether the orders of interest is a property of the input or a property of the operations. That is, things could be stated such that the input is essentially a set and not a sequence, but the 'seriality' of derivations imposes a sequence, which has been suggested corresponds to precedence (by axiom essentially). That is, here we've introduced a Derivational Correspondence Axiom. OR, there may be reasons to think of the input as being inherently ordered, and identifying that order with the 'order of formatives'. Here the differences are, I think, difficult to tease apart.

other phenomenon that have been argued to involve covert raising need to be examined to see how these cases would fit into the present system. Given MSO, connectivity/reconstruction facts should be looked at from the perspective of this system since nothing has been said above regarding the status of the semantic features of lexical items with respect to Spell-Out and Re-Merge. Last, the status of economy conditions requires some serious rethinking here. As mentioned above in a footnote, this system abandons a cost distinction between Move and Merge and thus abandons a plausible rationale for various 'timing' effects (e.g. expletive insertion). I leave these and other serious and interesting issues for future investigation.

References

Bach, E. 1988. "Categorial Grammars as Theories of Language". In Oerhle, et al. (eds.) *Categorial Grammars and Natural Language*. Kluwer, Dordrecht.
Barker, C. and G. Pullum. 1990. "A Theory of Command Relations" *Linguistics & Philosophy* 13: 1-34.
Bobaljik, J. 1995. "In terms of merger" In R.J. Pensalfini & H.Ura (eds) *Papers on Minimalist Syntax: MIT Working Papers in Linguistics,* vol. 27.
Bobaljik, J. & S. Brown 1997. "Interarboreal Operations: head movement and the Extension Requirement" *Linguistic Inquiry* 28.2, 345-356.
Bobaljik, J. & H. Thráinsson 1997. "Two Heads are Better than One" Talk presented at the University of Maryland College Park.
Brody, M. 1995. *Lexico-Logical Form: A radically minimalist theory.* MIT Press, Cambridge.
Castillo, J.C., J. Drury, and K. Grohmann, 1997. "Cyclic Evaluation as a way to acheive Local Economy." Ms. University of Maryland, College Park.
Chametzky, R. 1996. *A Theory of Phrase-Markers and the Extended Base*. SUNY Press, Albany.
Chomsky, N. 1970. "Remarks on Nominalization" In R. Jacobs & P. Rosenbaum (eds.) *English Transformational Grammar.* Ginn, Watltham, MA.
Chomsky, N. 1981. *Lectures on Government and Binding.* Foris, Dordrecht.
Chomsky, N. 1986. *Barriers.* MIT Press, Cambridge.
Chomsky, N. 1993. "A Minimalist Program for Linguistic Theory" . In K. Hale & S.J. Keyser (eds.) *The View from Building 20.* MIT Press, Cambridge.
Chomsky, N. 1994. "Bare Phrase Structure". In Gert Webelhuth (ed.) *Government and Binding Theory and the Minimalist Program.* Basil Blackwell, Oxford.
Chomsky, N. 1995. *The Minimalist Program.* MIT Press, Cambridge.
Chomsky, N. 1998. "Minimalist Inquiries: the framework" Ms. MIT.
Clark, R. 1990. *Thematic Theory in Syntax and Interpretation.* Routledge, New York.
Collins, C. 1997. *Local Economy.* MIT Press, Cambridge.
Drury, J.E. 1997. "Top-Down Derivations" Talk given at the University of Maryland Linguistics Department Student Conference, October, 1997.
Drury, J.E. 1998a. "Root-First Derivations: Mutiple Spell-Out, Atomic Merge, and the Coresidence Theory of Movement.": Generals Paper (defended April 1997) University of Maryland, College Park.
Drury, J.E. 1998b. "Precedence and Derivational Combination in the Minimalist Program: Order of Merge and Order of Formatives." Ms. University of Maryland, College Park.
Drury, J.E. Forthcoming. "The Challenge of Parasitic Gaps in the Minimalist Program." In S.Aoshima, M. Diab, J. Drury, & T. Neuvonen (eds.). *University of Maryland Working Papers in Linguistics.* Volume 8.
Drury, J.E. In progress. "Sequences of Representations and Projection as Unification." Ms. University of Maryland, College Park.
E-Kiss, K. 1985. "Parasitic Chains." *The Linguistic Review* 5, 41-74.
Epstein, S. 1995. "The Derivation of Syntactic Relations." Ms. Harvard University.
Ernst, T. 1994. "M-command and precedence" *Linguistic Inquiry* 25, 327-335.
Guimarães, M. 1998. "Repensando a Interface Sintaxe-Fonologia a partir do Axioma de Correspondencia Linear". Universidade Estadual de Campinas. MA Thesis.
Higginbotham, J. 1983. "A note on phrase-markers." *Revue Quebecoise de linguistique* 13: 147-165.
Hornstein, N. 1995. *Logical Form: From GB to Minimalism.* Blackwell, Cambridge.
Hornstein, N. 1996. "Movement and Chains" Ms. University of Maryland at College Park. [To appear in *Syntax*].
Hornstein, N. 1997. "Adjunct Control & Parasitic Gaps" Ms. University of Maryland College Park.
Huang, C-J.T. 1982. "Logical Relations in Chinese and the Theory of Grammar" PhD. dissertation, MIT, Cambridge.
Johnson, K 1990. "Object Positions" *Natural Language and Linguistic Theory* 9: 577- 636.

Johnson, D. & S. Lappin 1997. "A Critique of the Minimalist Program" *Linguistics & Philosophy* 20, 273-333.
Kayne, R. 1984. *Connectedness and Binary Branching.* Dordrecht, Foris.
Kayne, R. 1994. *The Antisymmetry of Syntax.* Cambridge, Mass.: MIT Press.
Kitahara, H. 1995. "Target-α: Deducing Strict Cyclicity from derivational economy" *Linguistic Inquiry* 26, 47-77.
Kitahara, H. 1997. *Elementary Operations and Optimal Derivations.* MIT Press, Cambridge.
Koizumi, M. 1993. Object Agreement phrases and the split VP hypothesis. In *MIT Working Papers in Linguistics 18: Papers on Case and Agreement I,* 99-148. MITWPL, Department of Linguistics and Philosophy, MIT, Cambridge.
Koster, J. 1987. *Domains and Dynasties.* Dordrecht, Foris.
Langacker, R. 1969. "On Pronominalization and the Chain of Command" In D. Reibel & S. Schane (eds.) *Modern Studies in English.* pp. 160-186. Englewood Cliffs, NJ. Prentice-Hall.
Lasnik, H. 1995. Verbal Morphology: *Syntactic Structures* meets the Minimalist Program. In *Evolution and revolution in Linguistic Theory: Essays in honor of Carlos Otero,* ed. Héctor Campos and Paula Kempchinsky, 251-275. Washington D.C.: Georgetown University Press.
Lasnik, H. 1997. On feature strength: Three minimalist approaches to overt movement. Ms. University of Connecticut. [To appear in *Linguistic Inquiry*]
May, R. 1985. *Logical Form.* MIT Press, Cambridge.
Munn, A. 1993. *The Syntax and Semantics of Coordinate Structures.* PhD. dissertation. University of Maryland at College Park
Nunes, J. 1995. *The Copy Theory of Movement and the Linearization of Chains in the Minimalist Program.* PhD. dissertation, University of Maryland, College Park.
Partee, B., A. ter Meulen, & Wall 1990. *Mathematical Methods in Linguistics.* Kluwer, Dordrecht.
Pesetsky, D. 1985. "Morphology and Logical Form" *Linguistic Inquiry* 16.2: 193-246.
Phillips, C. 1996. *Order and Structure.* PhD. dissertation. MIT.
Rizzi, L. 1990. *Relativized Minimality.* MIT Press, Cambridge.
Speas, M. 1990. *Phrase Structure in Natural Language.* Kluwer, Dordrecht.
Stowell, T. 1981. *The Origins of Phrase Structure.* PhD. dissertation, MIT, Cambridge.
Uriagereka, J. 1996a. "Multiple Spell-Out" Ms. University of Maryland, College Park.
Uriagereka, J. 1996b. "Minimal Restrictions on Basque Movement" Ms. University of Maryland, College Park.
Uriagereka, J. 1997. "Formal and Substantive Universals in the Minimalist Program" In C.Wilder, H.M. Gärtner, and M. Bierwisch (eds.) *The Role of Economy Principles in Linguistic Theory.* Berlin: Akadmie Verlag, p170-204.
Uriagereka, J. 1998. *Rhyme and Reason: an introduction to minimalist syntax.* MIT Press, Cambridge.
Watanabe, A. 1995. "Avoid Redefinition" In *MIT Working Papers in Linguistics 18: Papers on Case and Agreement I.* MITWPL, Department of Linguistics and Philosophy, MIT, Cambridge.
Weinberg, A. 1997. "Minimalism and Parsing" Ms. University of Maryland, College Park.
Yang, C. 1996. "Minimal Computation" Ms. MIT, Cambridge.

John Edward Drury
Department of Linguistics
1401 Marie Mount Hall
University of Maryland
College Park, MD 20742
jdrury@wam.umd.edu

differences between derivational views of syntax are of special interest to this investigation. While this intermediate multiple-rootedness is a *necessary* feature of Chomsky's BPS ('necessary' to the extent that the Extension Condition is necessary), it is not a required feature of the AM/MSO syntax introduced above. Thus, the current project will follow the second of the strategies suggested by Bobaljik & Brown (in the above quote) by advancing both a negative and a positive thesis:

- Intermediate multiple-rootedness (and the interarboreal operations this makes available) have unwanted conceptual and empirical consequences which require undesirable additional constraints to ensure restrictiveness
  AND,
- If a current assumption about structure building is inverted—in particular, if we view derivations as in the AM/MSO system instead of 'bottom-up' then intermediate multiple-rootedness is excluded for a principled reason (it is not *required*).

Since interarboreal operations rely on the possibility of intermediate multiple-rootedness, a system that eliminates the latter also excludes the former.[33]

2 Sidewards Movement: Parasitic Gaps and Adjunct Control

This section (§2) reviews some of the accounts that use deploy interarboreal operations, and then the following section (§3) returns to try out the AM/MSO system and sketch of analyses of the parasitic gaps and adjunct control data. In addition to the technical point that AM/MSO obviate the need for intermediate multiple rootedness (and therefore the possibility of interarboreal operations), §3 attempts to render plausible a stronger suggestion—that not only *can* derivations be understood in this way, they in fact *should* be.

2.1 Copies, Chains, and Linearization Nunes (1995) offers a set of proposals that articulates a way to derive the general mechanics of movement operations from the (PF) need to optimally linearize the terminal items of phrase markers. He is engaged in a minimalist exercise which aims to arrive at an understanding of the primitive operations of the grammar that allow one to derive the manner in which CHAINS are linearized. The center of the issue is the interaction between the Copy Theory of movement (Chomsky 1993, 1995) and Kayne's LCA. If movement involves the copying of the feature-contents of one item and merging this copy elsewhere in the structure, the question arises as to how these copies are mapped to a linear order given the LCA. Nunes proposes a system that effectively ensures that the head of a CHAIN will be the link that 'counts' for linearization purposes since all other links/copies in a CHAIN will be deleted to meet interface requirements.[34] What is of interest to this discussion is his use of intermediate multiple-rootedness and sidewards movement.

Nunes views movement not as a primitive operation, but as a descriptive label for a complex of distinct derivational sub-processes—Copy, Merge, Form Chain, and Chain Reduction. Thus, contra Chomsky (1994, 1995: chap.4), the creation of a CHAIN is not an 'automatic' result of the operations of Copy and Merge—it instead requires the additional operation Form Chain to apply. Crucially, note that it is this factoring of the operation 'Move' into distinct suboperations that allows the possibility of interarboreal operations. Nunes argues that a subcase of such operations—so-called Sidewards Movement—provides a natural account of parasitic gap constructions. First, consider his formulations of Form Chain and Chain Reduction:

---

[33] The conclusion will turn out to not be so strong as this, as we will see by the end of this paper. The AM/MSO syntax presented here has the virtue of not *requiring* intermediate multiple-rootedness, but nothing in the current system but stipulation says that its impossible. The obvious point will be that interarboreal operations of the sort to be discussed in this section are excluded once the precondition for their existence (multiple roots) is removed. Although, as Hornstein points out (p.c.) the subcase of interarboreal operations underlying his analyses (to be discussed below) actually doesn't rely on unconnected subtrees. What is crucial is the denial of the θ-criterion, thus allowing the movement into a θ-position. Once this is done, its a technical matter hinging on whether the definitions under something like the Barrier system are stated in terms of domination or containment. If the latter, and if the θ-criterion is abolished, "sidewards movement" is permitted.

[34] This is essentially because the head of a chain will be the only link that has all of its features checked. The details of how this is done and the involvement of economy conditions in this account aren't relevant to this discussion. The assumptions and mechanisms of Nunes's theory that are relevant will addressed as necessary.