ABSTRACT

Title of Dissertation: ALTERNATIVE DIRECTIONS FOR MINIMALIST INQUIRY: EXPANDING AND CONTRACTING PHASES OF DERIVATION

John Edward Drury, Doctor of Philosophy, 2005

Dissertation Directed By: Professor Juan Uriagereka, Linguistics Department

This dissertation develops novel derivational mechanics for characterizing the syntactic component of human language — Tree Contraction Grammar (TCG). TCG falls within a general class of derivationally-oriented minimalist approaches, constituting a version of a Multiple Spell Out (MSO-)system (Chomsky 1999, Uriagereka 1999, 2002). TCG posits a derivational WORKSPACE restricting the size of structures that can be active at a given stage of derivation. As structures are *expanded*, workspace limitations periodically force *contractions* of the span of structure visible to operations. These expansion-contraction dynamics are shown to have implications for our understanding of locality of dependencies, specifically regarding successive cyclic movement. The mechanics of TCG rely on non-standard assumptions about the direction of derivation — structure assembly is required to work top-down. TCG draws a key idea from TAG; that is, *recursive structure* ought to play a direct role in delimiting the range of possible interactions between syntactic elements in phases of derivation. TAG factors complex structures into non-recursive *elementary* trees and recursive *auxiliary* trees that are

combinable via TAG's two operations (substitution/adjoining). In TCG the expansion of structure in the workspace is similarly limited to containing only *non-recursive* stretches of structure. In the course of a derivation, encountering "repeated elements" in the expanding dominance ordering forces *contractions* of the workspace (understood to happen in potentially different ways depending on the properties of repeated elements). In certain circumstances, repeated elements are *identified*, allowing information from earlier stages of derivation to be carried over to later stages, underwriting our (novel) view of successive cyclicity. Recursive structure is retained in the global "output" structure, upon parts of which we understand the workspace to be superimposed.

ALTERNATIVE DIRECTIONS FOR MINIMALIST INQUIRY:
EXPANDING AND CONTRACTING PHASES OF DERIVATION

By

John Edward Drury

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Committee:
Professor Juan Uriagereka, Chair
Professor Norbert Hornstein
Professor Howard Lasnik
Professor Colin Phillips
Professor Amy Weinberg
Professor James Reggia

Dedication

To my Father
Dr. Thomas Francis Drury
for fostering love of wisdom

And to Lisa
for patience and understanding

Acknowledgements

There is no way to adequately thank all of the people who made the completion of this dissertation possible. But this is the place where one is to make such an attempt, so here 'goes. I thank first my friend and committee chair Juan Uriagereka for leaving just enough bread crumbs behind him on that twisting and difficult to follow trail that leads to the Outside of The Box. Thanks as well go to my committee: Colin Phillips, Amy Weinberg, Howard Lasnik, and Norbert Hornstein, and to James Reggia for agreeing to fill the role as Deans Rep on short notice. To both Howard and Norbert I owe a special additional debt for last minute help above and beyond the call.

Standing on the shoulders of others is a job for an acrobat — which I am not. So, to the following people I must say that I am sorry for all the kicking and if I poked you in the eye once or twice, please know it was not intentional,...thank you for your patience and for whatever help you provided me at one point or another in conversations about linguistics in general or about issues in the vicinity of this dissertation: Mark Arnold, Stephen Crain, Norbert Hornstein, Howard Lasnik, David Lebeaux, David Lightfoot, Colin Phillips, Phil Resnik, Amy Weinberg, Juan Carlos Castillo, Kleanthes Grohmann, Cedric Boeckx, Max Guimarães, Jeff Lilly, Matt Kaiser, Tom Cornell, Robert Chametzky, Chris Wilder, Georges Rey, Ricardo Extepare, Itziar San Martin, Aitizber Atutxa, Lucia Quintana, Eli Murgia, Caro Struijke, Acrisio Pires, Peggy Antonisse, Julien Musolino, Viola Miglio, and Laura Benua.

I owe a serious debt to Maryland's Linguistics Department as whole, for supporting my work and my education, and for providing an excellent environment in

which to grow. Within this department, Kathi Faulkingham deserves extra special thanks for helping to ensure that my typically pathological response to administrative deadlines did not cause any unsolvable problems.

Thanks also go to Michael Ullman for generous support and encouragement during my stay in the Brain & Language Lab in Georgetown University's Department of Neuroscience, and to B&L Lab members past and present, especially Karsten Steinhauer and Matt Walenski.

I thank also my parents, Thomas and Margaret Drury for unconditional love and backing in all my endeavors. And last, to my amazing wife Lisa, thank you for putting up with my process and my drama over these years, and for keeping the faith that one day (today!) this would actually come to an end,... and thank you little O, for demanding that I complete this work so as to make more time for play and laughter.

Table of Contents

CHAPTER 1: Spelling Out the Workspace

the idea of the form implicitly contains also the history of such a form
Hallé, Oldeman, & Tomalinson (1978)

This dissertation develops novel derivational mechanics for characterizing the syntactic component of human language — Tree Contraction Grammar (TCG). The approach falls into the general class of derivationally oriented systems under development within the Minimalist Program,[1] and more specifically into a category of models that I will call here Multiple Spell Out (MSO-)systems (Chomsky 1999, Uriagereka 1999, 2002). MSO-systems, generally speaking, divide derivations into sub-derivations, the outputs of which may be independently evaluated at the interfaces to extra-grammatical systems and may play a special role in demarcating domains with import for understanding the locality of syntactic relationships.

I propose in this work a general way of thinking about how MSO-systems function, relying on a distinction between a syntactic WORKSPACE and a derived OUTPUT structure. It is within this context that the core theoretical intuition underlying TCG emerges. The approach is informed as well by Tree Adjoining Grammar (TAG).[2] In particular, as in TAG approaches, the fundamental notion of recursive structure is argued here to play an direct role in understanding the locality properties of syntactic dependencies. The general perspective developed in this work sees MSO-systems (and TCG in specific) together with TAG as a family of closely related approaches.

[1] Chomsky (1995, 1998, 1999), Uriagereka (1998, 2002), Lasnik & Uriagereka (2004).
[2] See Frank (1992, 2002), Frank & Kroch (1995) among many others.

TCG distinguishes itself in terms of the mechanisms it makes available for analyses of successive cyclic movement (SCM) phenomena in two ways that I argue to be of broad interest theoretically:

(1) a. **The Non-Existence of "EPP-/P-features":** if the key ideas are right, special features driving intermediate movements in SCM are not needed

b. **Derivational Directionality**: the mechanics of TCG derivations demand that structure assembly work "top-down", and not bottom-up as in Chomsky's (1994, 1995) widely adopted Bare Phrase Structure (BPS)

It turns out that (1)a and (1)b are related. Telling the story of how this might be so is the main mission of this dissertation.

A metaphor helps to get the general intuition behind our Workspace/Output-Distinction (WS/O-distinction): picture an arbitrarily complex syntactic structure upon which we might shine a spotlight beam which can illuminate only small portions of structure, leaving the rest in darkness. Construction of syntactic objects and the licensing of dependencies within such structures is understood to take place within the illuminated span, and not elsewhere (no *syntactic* work can happen in the dark). Thus, in order to expand structure beyond the maximal span of illumination, the spotlight beam must "move on". As a consequence, some of the previously established structure will necessarily have to be left behind outside of the illuminated zone.

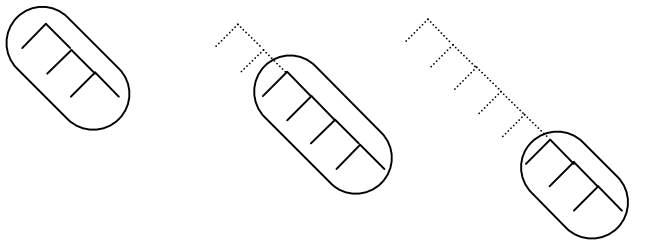
If derivationally later expansion of structure requires the spotlight to move-on before a required syntactic licensing operation has occurred, there is no backtracking to fix the problem. In such a situation the output structure is stuck with some illegible/unlicensed property which has been left outside the spotlight. (The interface

systems, unlike the narrow syntactic combinatorial system defined by the workspace, can see in the dark.)

The task of developing this metaphor into concrete proposals that can support reasoning about the syntactic component of the human language faculty is thus the task of specifying the principles that could be understood to govern the span of this spotlight beam, how and what operations happen within it, how and why it moves on to further expand structure, what happens to the old structure left outside the workspace boundaries as such later expansions occur, and so on. [3]

Consider a graphical illustration of the intuition informing our workspace/output distinction (WS/O-distinction). The following in (2) and (3) show two different partial derivations. The first of these schemas illustrates a system in which construction within the workspace creates the output structure in a top-down fashion; the second schema represents a similar process working bottom-up. Direction of derivation in this sense will be important in this work. As mentioned above, the present approach will be required to work roughly as pictured in (2) and not as in (3).

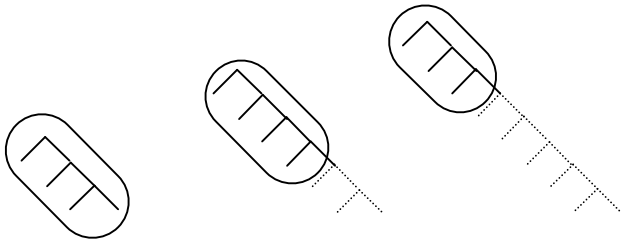
(2)



⇒⇒⇒⇒⇒⇒DIRECTION OF DERIVATION⇒⇒⇒⇒⇒⇒

[3] Note that any connection to the notion of "working memory" as this notion is deployed in the psychological literature is metaphoric. In particular, the limitations proposed here to govern the maximal spans of the derivational workspace are not expected to *vary* across individuals. The constraints proposed here are alleged to be "hard" architectural constraints on the combinatorial component. I do think there is room to relate the proposals here to a story about memory systems, but I won't be spending time on this issue in the present work.

(3)



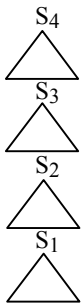
⇒⇒⇒⇒⇒⇒**DIRECTION OF DERIVATION**⇒⇒⇒⇒⇒⇒

The enclosures in these schemas represents the state of the workspace at earlier vs. later stages of derivation. Again, if this "spotlight beam" is limited in its scope then it must at certain points of derivation move on to expand further structure, thus leaving some of the previously established structure outside of its boundaries. Such abandoned spans are represented by the dotted-arcs for the portions of the structure outside the enclosures in (2) and (3). The notion of an OUTPUT then is the entire span of the established structure of which the workspace only allows us to see parts of at any given step.[4] I will show below that this way of thinking — our workspace/output distinction (WS/O-distinction) — is helpful for reasoning about the workings of MSO-type approaches generally. For example, in addition to having this distinction serve as a platform for the development of TCG, I will be discussing both Chomsky's (1998, 1999) view of spell-out as happening "by phase" in these terms, as well as Uriagereka's (1999) linearization-based view.

To begin to get a better feel for the WS/O-distinction, consider the old idea that there might exist special *cyclic* nodes defining domains in syntactic complexes within which some inventory of transformational operations $\langle T_1,...,T_n \rangle$ are applied and then reapplied to (recycled in) the next higher domain, and so on, as in (4)).

[4] With one exception: the first "phase" of derivation involving expansion of the structural description will, up until the first contraction of the workspace, correspond one-to-one with the output structure (see also fn6).

(4)



⇐(RE)APPLY $\langle T_1,...,T_n \rangle$ (*cycle*-❹)

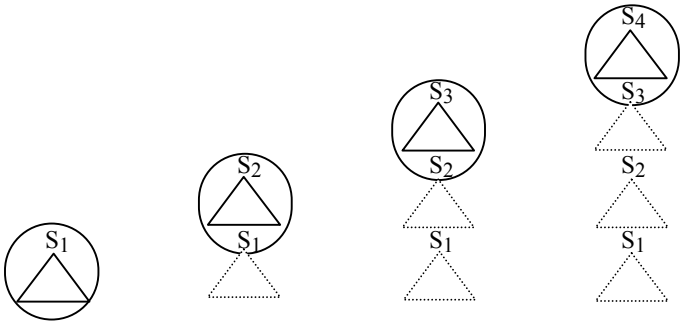
⇐(RE)APPLY $\langle T_1,...,T_n \rangle$ (*cycle*-❸)

⇐(RE)APPLY $\langle T_1,...,T_n \rangle$ (*cycle*-❷)

⇐APPLY $\langle T_1,...,T_n \rangle$ (*cycle*-❶)

This can be understood in present terms by identifying the maximal span of the workspace with such cyclic domains. As cycles are completed, a new workspace (cyclic domain) begins, leaving the previous cyclic domain stranded outside the borders of the workspace. This is pictured in (5):

(5)



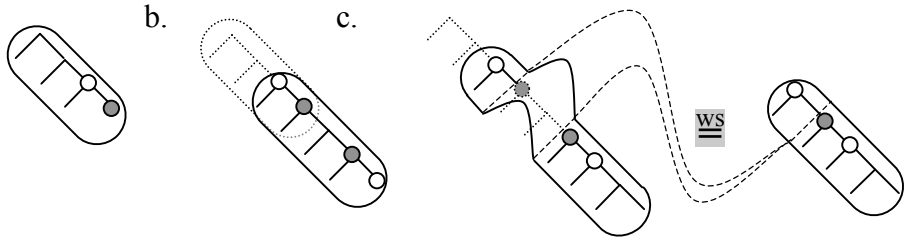
But this is just translating terminology — ordered nodes marking domains for applications of an inventory of transformations give way to a limited workspace that periodically expands up to a cyclic node and contracts, "clearing the buffer" to make way for the next domain expansion.[5] So what is the interest of the WS/O-distinction?

[5] My thinking of spell-out in terms an emptying/clearing of a buffer of sorts grows in part from numerous helpful conversations with Max Guimarães. See Guimarães (1999) for related discussion involving alternative derivational directionality and possible applications to thinking about syntax/prosody relationships.

Note that the schema in (5), and those in both (2) and (3) above, obey a general restriction. The structures in the workspaces at different steps of derivation in these examples are always *contiguous* (sometimes improper) parts of the output structure.[6]

I suggest that a principled view of how the contents of the workspace are regulated can result in situations where *non-contiguous* portions of the output structure are represented in the workspace. This will be key to our development of a basis for analyses of core cases of SCM. Consider a rather fancier illustration which conveys this alternative intuition:

(6) a. b. c.

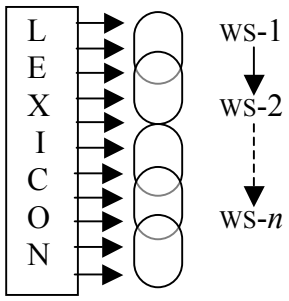


In (6) two types of workspace contractions are represented. Assuming for this example a top-down derivation, the first such contraction occurs between the first two structures in (6)a and (6)b. This is just like the contractions depicted in (2) in which we see expansion of the workspace at the bottom end forcing an abandonment of structure at the top (i.e., the spotlight beam moves on; of course the opposite happens with our schemas in (3) and (5) above where expansions at the top force abandoning structure at the bottom).

[6] That is, for any starting point of a derivation, up to the first contraction of the available span of structure visible in the workspace the workspace structure corresponds directly to the output. I say "corresponds" rather than "is identical to" since I will be viewing the workspace superimposition on outputs as a matter of *only* syntactic (F) properties being "in" the workspace, which will be understood to be associated with the PF- (π) and LF- (λ) relevant properties that will be understood to be what populates the output structure. However, the "structure" (ordering relationships between nodes) will be understood to be the same across the workspace/output division. I will unpack all this below.

However: between the second and third structure in (6)b and (6)c we have a contraction which is unlike the first. In this second contraction it is neither the top nor the bottom of the structure which is voided from the workspace, but rather some intermediate stretch. This results in bringing two nodes (the open/unshaded nodes in (6)) into a more local relationship within the workspace than existed prior to the contraction. It also results in some intervening material being "spliced-out" of the workspace, though we understand this spliced-out material to still be present in the output. Thus, the two rightmost objects (those in (6)c) are equivalent in terms of what is in the workspace (I mark this above as "$\overset{WS}{=}$") though the second abstracts away from the output structure to which it is connected (i.e., via the elements still in the workspace).

Our metaphor of a spotlight beam breaks down at this point of course, so let us kick that ladder away — the formal intuition should now be clear enough to proceed. The key idea is that non-contiguous portions of the output may be maintained in the workspace (WS). I will suggest below that the best way of viewing WS/O-distinction is not in terms of two levels of syntax, but rather to understand the connection between the two as a dynamic interface. WS-computations incrementally yield a structured object populated by only LF and PF relevant properties. But an interface is the meeting point of at least two different systems, if the WS is itself an interface system, then we should ask, "between what and what *else*?" On one side of the WS I have just suggested that we have a structured PF/LF object — what then is on the other side? Answer: the lexicon. The view of syntactic architecture can be visualized as follows:

(7) 

A better way of putting things then would be to say that the WS is itself the "interface" between the lexicon, on the one hand, and derived PF/LF output structures on the other. The lexicon feeds the WS which expands up to its limits (such limits are introduced and developed below), and then moves-on or *contracts*. The dynamics of WS expansion and contraction leaves in its wake a structured object — a tree — which is populated by only PF and LF relevant properties.

The interest of the WS/O-distinction within the TCG approach developed here is in the nature of the shrinking/contraction processes that yields a way of treating superficially non-local relationships as potentially reducible to more local domains. So the key question becomes: what drives contractions of the workspace, and how might we understand these to work in a way that can support analysis of more non-local-looking relationships?

I begin development of my answer(s) to this question in §1.1, introducing two constraints on categories and ordering in the workspace, and show how this yields a novel schema for analyses of successive cyclicity phenomena. In §1.2 I develop some sample general derivations for A- and A'-relations, and highlights some features that will be of interest in later discussion. §1.3 sums up the previous two. §1.4 backs up to consider MSO-systems and TAG focusing mainly on their general outlooks on successive cyclic movement. The discussion of these neighboring models is framed within our

WS/O-distinction, bringing forward its generality and highlighting some conceptual advantages of viewing MSO-systems in particular in this way. These discussions lead us to consider, in §1.5, some technical issues regarding the theory of movement chains, developing a version of an idea offered in Chomsky (1995) where it is suggested that movement chains be understood as sets of contexts/positions.[7] In §1.5 I also consider issues regarding categories/features and formal ordering — that is, the theory of local intra-/inter-phrasal organization — and settle on a "reduced" view adopting some ideas from Brody (2000, 2003). This reduced view is suggested to be a positive step in the direction of restrictiveness of the overall theory, though the central motivation for its adoption is its overall "good fit" with the key intuitions underlying TCG.

Following this, Chapter 2 discusses empirical and theoretical issues regarding successive cyclic movement in some more detail, and raises some issues regarding so-called EPP-/P-features — what are called "Move/Merge-features" or "M-features" here — targeting them for elimination in the TCG approach. A number of other views regarding cyclicity are discussed as well.

Chapter 3 then turns to develop the TCG ideas in more detailed analytical discussion, focusing initially on Raising-to-Subject (RtS) and *wh*-movement. The approach is demonstrated to require a top-down implementation, and some contrasts with TAG are discussed. The approach is then explored in possible extensions to other

[7] To jump ahead a bit: I will suggest that Chomsky's particular view — which views the "context" for an element α as the entire derivation up to the point where α is introduced/merged (i.e., the "sister" of α is the context defining this occurrence of α, and the "sister" is itself viewed as the entire structure that this sister element dominates). I suggest in §1.5 that this is both too strong and too weak — it is too strong in that identifying/distinguishing occurrences requires reference to arbitrarily large stretches of previously established structure; it is too weak because we will see that reducing our understanding of contexts to just the *label* of the context α relates to will permit us to view some contexts as indistinguishable from others. This will turn out to be what underwrites SCM without the posultation of EPP-properties.

phenomena. I conclude with some open questions and general discussion regarding the architecture of syntactic theory.

1.1. WS Constraints & A Sketch of SCM in TCG

In this section I introduce two possible constraints on the workspace and examine some of their consequences. These notions conspire to yield linked-local relations of the successive cyclic movement (SCM) sort.

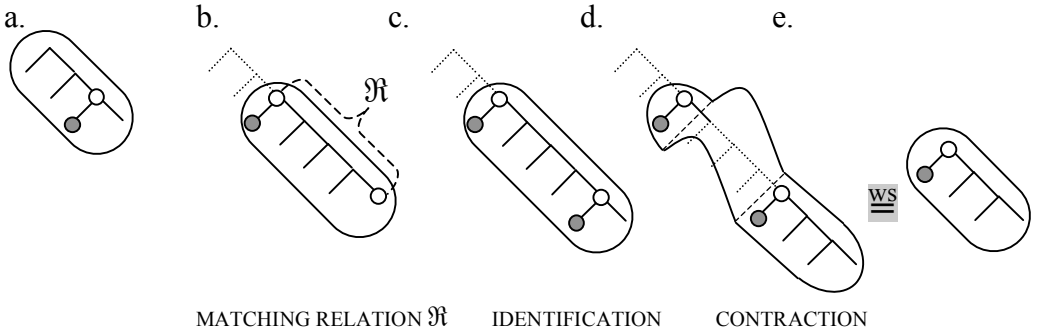
(8) **WORKSPACE ORDER:**
The elements in the workspace manifest a weak partial order (DOMINANCE)[8]

(9) **WORKSPACE DISTINCTNESS (ANTI-RECURSION):**
The workspace does not tolerate the presence of multiple tokens of type X

First, as mentioned, the system will be understood to work "top-down". I will return to explain why things *must*, in fact, work this way in the conclusion. Take the shaded node in (10)a to be a "to-be-moved" element and the open/unshaded node to be its initial structural context (housing the relevant licensing feature(s), e.g., *wh* or perhaps Case/ϕ information). Assume for the moment that the branching order represented by the tree structure manifests the traditional notion of *dominance* (i.e., a transitive, antisymmetric, reflexive relation):

[8] I will consider the possibility of a rather stronger statement regarding ordering in later discussion.

(10) a. b. c. d. e.



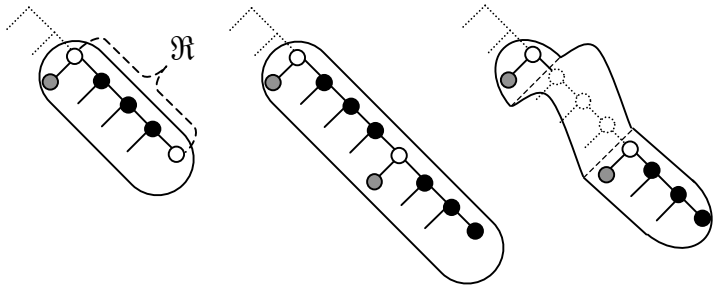
At the point in (10)b an element is introduced which satisfies a matching relation "ℜ". This relation will be further specified below, for now simply take ℜ*xy* to be satisfied if *x* and *y* are non-distinct. This situation thus violates the distinctness condition on workspace contents stated above in (9), so something must happen in response in order for this to be a well-formed workspace.

Suppose that the system responds by taking these non-distinct elements to be essentially the one and the same thing. If they are identified then whatever the higher element dominates, so does the lower one. This effects a copying/lowering of the shaded node ((10)c). Note that we do not duplicate elements in the workspace — what happens in the workspace is an identification of the open/unshaded nodes, so that subsequent to the contraction step in (10)d there are not two tokens or occurrences of either the "moving" element or its context. There are rather just single nodes for each in the workspace (note the WS equivalence is marked again as $\overset{WS}{=}$ between (10)d and (10)e above).

However, there are now, in virtue of this process, *two* such pairs in the output structure. Thus, what remains in the workspace subsequent to contraction is best understood in terms of the picture in (10)e, though (10)d captures the workspace/output structure correspondence. Thus, what is one in the workspace can be many in the output.

Observe that while identifying these open nodes in the structure could be understood to effect the equivalent of a lowering operation in terms of preserving the domination relations of the upper elements in the output, we clearly seem to require a way of blocking the similar copying of all such dominated elements. That is, why shouldn't this copying apply to the other dominated material (e.g., the dark nodes on the main path in (11)a, resulting in (11)b and then (11)c following contraction)?

(11) a. b. c.

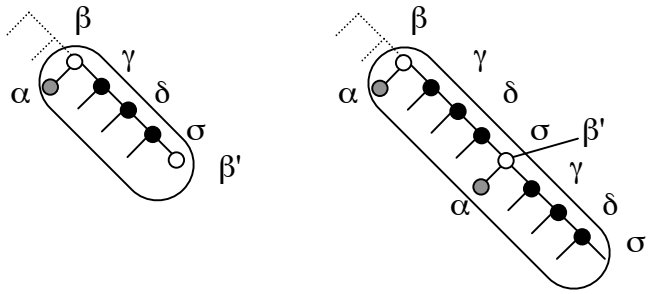


How is a regress of sorts avoided? What stops this process from copying all the domination relations of the upper instance of the two identified nodes (i.e., α dominates β dominates α dominates β,...*ad infinitum*)? How does this terminate?

I suggest that we regard the IDENTIFICATION and CONTRACTION steps pictured above in (10)c and (10)d as essentially a one-step operation governed by the general ordering restriction given in (8). Adding the "moved" element and its context to the bottom of the workspace structure in virtue of the identification of this lower node with the upper one adds new pairs to the dominance relationship in the workspace. Technically this will only be possible if elements in previous pairs in this dominance order that would introduce conflicts violating the antisymmetry of the dominance relation are *removed* from the workspace (though, importantly, *preserved* in the output). This is, essentially, the notion of *contraction* in the TCG framework.

Consider (11) again, with the nodes labeled so that we may refer to them in specifying the relevant formal ordering properties as in (12):

(12) a. b.



Prior to the introduction of β' (i.e. the step prior to (12)a) and the subsequent identification (β, β'), we have the following dominance order *D*:

(13) $D = \left\{ \begin{array}{l} \langle \beta, \alpha \rangle, \langle \beta, \gamma \rangle, \langle \beta, \delta \rangle, \langle \beta, \sigma \rangle, ... \\ \langle \gamma, \delta \rangle, \langle \gamma, \sigma \rangle, ... \\ \langle \delta, \sigma \rangle \end{array} \right\}$

The introduction of β' then adds the following pairs to *D*:

(14) $D = \left\{ \begin{array}{l} \langle \beta, \alpha \rangle, \langle \beta, \gamma \rangle, \langle \beta, \delta \rangle, \langle \beta, \sigma \rangle, \mathbf{\langle \beta, \beta' \rangle}, ... \\ \langle \gamma, \delta \rangle, \langle \gamma, \sigma \rangle, \mathbf{\langle \gamma, \beta' \rangle}, \mathbf{\langle \gamma, \alpha \rangle}, ... \\ \langle \delta, \sigma \rangle, \mathbf{\langle \delta, \beta' \rangle}, \mathbf{\langle \delta, \alpha \rangle}, ... \\ \mathbf{\langle \sigma, \beta' \rangle}, \mathbf{\langle \sigma, \alpha \rangle} \end{array} \right\}$

Assuming *D* is generally a weak partial order (transitive, antisymmetric, and reflexive), if we *identify* β and β' then we have ordering conflicts even if we do not copy all the nodes β dominates to the local domain of β' — for example: ⟨β, γ⟩ and ⟨γ, β⟩, ⟨δ, β⟩ and ⟨β, δ⟩, and so on. If everything the upper "occurrence" of β dominates is copied,[9] then we end up with the situation in (11)b/c and (12)b, and many more ordering conflicts would thus

[9] I will be refering to the notion of copying as a convenience. The idea here is that there is an operation (node identification) which results in the equivalent of copying, but that there is no specific "duplicating" operation which takes a single element α as an input and produces a pair of identical outputs.

arise in virtue of creating domination symmetries for the elements {γ, δ, σ} (so that we have both ⟨γ, δ⟩ and ⟨δ, γ⟩, *etc.*).

How might the system respond to the possibility of creating such ordering conflicts? Nunes (1995, 1999, 2004) addresses a similar problem as it arises in his development of the copy theory of movement set within the context of Kayne's (1994) Linear Correspondence Axiom (LCA). For Nunes the problem is that when an element α is copied and (re)merged in a c-commanding position, similar kinds of ordering conflicts emerge since α, in addition to now c-commanding "itself", both c-commands and is c-commanded by all the intervening nodes along the movement path. A Kayean view of structure/order correspondence requires there be no such conflicts in order to map hierarchy to precedence.[10] Nunes reconciles the conflicts between the linearization demands imposed by the LCA and the symmetric c-command relations in the structure resulting from movement as copying by positing a mechanism he calls Chain Reduction, stated as follows:[11]

(15) **CHAIN REDUCTION:**
Delete the minimal number of constituents of a nontrivial chain CH that suffices for CH to be mapped to a linear order in accordance with the LCA

A similar idea can be employed to fit with the idea of removing elements from the workspace (contraction/spell-out). The outcome we want is for β—α in (12)a to be reintroduced in the output structure so, for example, the elements {γ, δ, σ} will all dominate α in the output (this will be important for our treatment of certain connectivity

[10] See also Chomsky (1995) for some discussion of this point where the deletion of copies to satisfy the LCA conceived as a bare output condition on the PF side of the grammar is proposed.

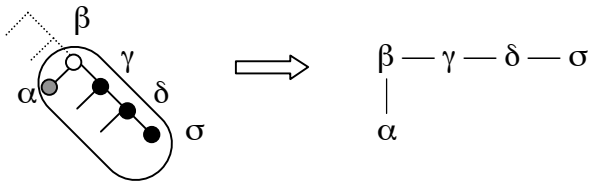
[11] Nunes additionally proposes a formal feature elimination procedure that is crucial to his analyses. I won't discuss this here.

effects — see Chapter 3). But we want these intervening {γ, δ, σ} elements to be spliced-out of the workspace so as not to introduce ordering conflicts.

Note that this requires some distinction between the workspace and the output to ensure that what is problematic with respect to ordering conflicts in the workspace is *not* problematic in the output. The nature of this particular difference relies on later developments, but I will offer a sketch below.

First consider what happens if we assume the following. Take the structure under discussion prior to the addition of the element β' which will match under relation ℜ with β, and let us prune away some of the notation to focus on the relevant elements and their ordering properties, as in (16):

(16)



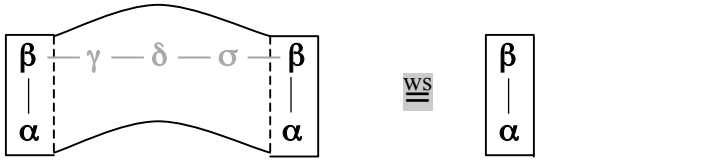
Now we add β':

(17) β — γ — δ — σ — **β'**
|
α

(β, β') satisfies ℜ, and the nodes are identified in the workspace. Since β and β' are now the same element, α comes to be dominated by the intervening elements:

(18) β — γ — δ — σ — β
| |
α α

This creates no ordering conflict since α was in no domination relation with the intervening elements prior to the identification. But the intervening elements *do* create ordering conflicts, and so the workspace must contract (splice-out the interveners) to respect the properties of the dominance order:

(19) 

WORKSPACE AND OUTPUT JUST THE WORKSPACE[12]

Although we have not yet specified the nature of the matching relation ℜ, the mechanics of workspace contraction as just discussed follow from our workspace constraints in (8) and (9) (together with ℜ). The ordering constraint in particular ensures that we will be able to add the new domination relationships for α, *but*: (i) we cannot add relations that cause ordering conflicts and (ii) any elements that *would* create such problems subsequent to the identification (β,β') must be spliced-out. And the addition of the new domination relationships that effect the "lowering" of α follows from the proposed response of the system to a potential violation of the distinctness condition.

What then of the output structure? Does it obey this (or any) ordering restriction or not? What about distinctness?

If we make the standard assumption that the items being combined are minimally triples of semantic (λ), phonological (π), and syntactic (F) information, ⟨λ, π, F⟩, then the

[12] So the workspace has just one β and one α. The output, however, has two β's and two α's (or, rather, as I will suggest in a moment, the output has two *correspondents* of β and two *correspondents* of α).

following line of thinking is available to us, and will in fact be central to our conception of the WS/O-distinction: the workspace manipulates only F-properties.

In fact, we can take this a step further: "being in the workspace" could be identified with "having F-properties". The general idea is another way of framing the key intuition underlying the TCG approach. That is, categorial F-properties are a limited commodity in the syntactic workspace. A given manifestation of the workspace can contain exactly many distinct elements as there are categorial distinctions in the system. There is no such limitation of this sort "outside the workspace" because being outside the workspace just means that these formal distinctions are no longer connected with $\langle \pi, \lambda \rangle$ information.

On the general view developed here, the output structure is a *syntactic* object in the sense of manifesting the formal ordering properties established in the workspace, but it will be populated with *only* π and λ properties. The output will in this sense be an object *of* the interface systems, with the PF-component inspecting only the π-vocabulary and the LF-component inspecting only the λ-vocabulary, but with *both* sets of vocabulary constrained by the same structure.[13]

The following illustrates the idea abstractly. Our general intuition of the workspace having to "move-on" to expand new structure is pictured first for an abstract domination order of elements:

[13] Having structure housing both the π and λ types of vocabulary also yields a venue for exploring primitive π/λ correspondences over such structures. For example, the well-known connections between prosody/intonation and the semantics of focus would be one such area to explore with these mechanics. These matters are not explored here. In general we will be concentrating mostly on what happens in the workspace, and how this might relate to the output. However some brief remarks will be made about how we might think about relationships established over output structures — these are suggested to be potentially truly non-local (examples include variable binding by a quantifier, long-distance obviation effects, so-called unselective binding, etc.).

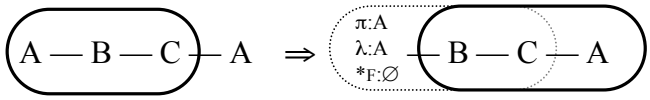
(20) 

Supposing then that {A, B, C} are the relevant formal properties, as the workspace moves on what will be left in the output structure are the associated π and λ properties of each formal element {A, B, C}, (e.g., $\{^{\pi:A}_{\lambda:A}, ^{\pi:B}_{\lambda:B}, ...\}$):

(21) 

Now when we say that the workspace "moves-on", we understand this to mean that the relevant formal properties which π/λ-pairs are connected to must be "reused" in establishing new expansions of structure in the workspace. That is, if the anti-recursion/distinctness condition in (9) holds, this means that such formal/syntactic information must be stripped away from earlier introduced elements so that it can be used to structuralize new/incoming ones.

We can now illustrate the situation described metaphorically above — where an unlicensed F-property is abandoned from the workspace:

(22) 

Assuming that there is no "backtracking" of the workspace, this will produce an anomaly as the interface systems are confronted with an illegible element. Here we have marked this offending property as "*F:∅", though note that above we suggested that an element's "being in the workspace" be identified with "having syntactic/formal properties". Below I will be suggesting specific roles for licensing properties like WH, Case/agreement, and θ, so the way this will actually be understood will be in terms of a failure of a formal

relation obtaining in the workspace leading to an illegible PF or LF property (see §1.2 below, and Chapter 3 for some discussion of features, valuation, and interface legibility).

Regarding our concerns about ordering and repeated elements in the *output*: this is now best viewed as a constitutive difference between the workspace and the output structure. The distinction resides in exactly whether it is possible to represent multiple tokens of a given type or not. In the output, this is possible. In the workspace, it is not. The systems supporting the representation/processing of PF and LF vocabularies, that is, are capable of handling multiple tokens of a given type; the narrow syntactic computation in the workspace, which is stated over formal features/properties, cannot do this. This is one of the central ideas underlying the TCG approach.

An important idea here, discussed in §1.5, is the idea of thinking of movement chains as sets of contexts/positions, though I will argue that we require a simpler view than the one presented in Chomsky (1995). There it is suggested that we view contexts as the entire structure derived up to the point where a moved/remerged item is (re)integrated. I argue that returning to a simpler view, where the context is simply the local *label*, and not the entire structure, allows us to view certain sets of contexts as indistinguishable, yielding SCM.

In the next section I develop some sample derivations for core cases of A- and A'-movement to get some technical ideas on the table.

1.2. Local & Linked Local Relations: Sample TCG Derivations

Now let us consider a pair of standard cases for which SCM analyses have been deployed, in particular *wh*-movement and raising-to-subject (RtS). First, some

simplifications regarding structure and category will be helpful — I will return to discuss these simplifications further in §1.5.

Consider the following case with multiple clausal embedding in (23)a, with the partial description in (23)b:

(23) a. Dave thought Mary believed John liked pizza
b. $[_{CP}\ C^0\ [_{TP}\ DP\ [_{T'}\ T^0\ [_{VP}\ V^0\ [_{CP}\ C^0\ [_{TP}\ DP\ [_{T'}\ T^0\ [_{VP}\ V^0\ [_{CP}\ C^0\ [_{TP}\ DP\ [_{VP}\ V^0\ DP]...]$

If we look at the "spine" of the clause as structured in (23)b — that is, the dominance ordering running from the root to the most embedded element that manifests the sequence of head-complement selection/projection relationships[14] — we see the following sequence of major categorial distinctions between types of elements as in (24)b (ignoring intra-phrasal projection level distinctions, thus collapsing any XP/X' to just X):

(24) a. $[_{CP}\ C^0\ [_{TP}\ DP\ [_{T'}\ T^0\ [_{VP}\ V^0\ [_{CP}\ C^0\ [_{TP}\ DP\ [_{T'}\ T^0\ [_{VP}\ V^0\ [_{CP}\ C^0\ [_{TP}\ DP\ [_{VP}\ V^0\ DP]...]$

b. C—T—V—C—T—V—C—T—V—D

This spine branches to include the external arguments in the specifier positions of the T-elements associated with each verb, which we add to this reduced diagram as follows (the branching, directional arc is superimposed here to clearly indicate the assumed dominance ordering):

(25) C—T—V—C—T—V—C—T—V—D
D D D

[14] On some views, the relation from functional-to-functional elements and functional-to-lexical elements is discussed in terms of *selection* (e.g., C^0 selects TP, T^0 selects VP, etc.), perhaps with a distinction made between "syntactic" and "semantic" selection (see, e.g., Abney 1987, 1991). On other views (Grimshaw 1991, 2002; van Riemsdijk 1991, 1998) functional-to-functional and functional-to-lexical relations are governed by the notion of (extended) projection, while "selection" is reserved for lexical-to-functional and lexical-to-lexical relations.

These reduced structures will be sufficient to make the points of interest here. Later on I will argue that this should be seen as more than expository convenience, but rather is a view of structure that makes available the "best fit" with our core constraints on the workspace (in (8)/(9)). [15] Now consider the following:

(26) a. John seems to tend to appear to like carrots
b. What did Dave think that Mary believed that John liked?

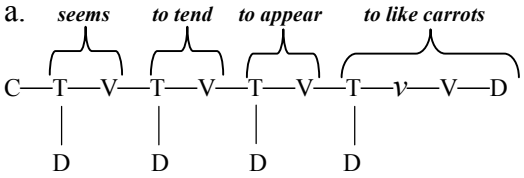
(27) a. John [seems [to tend [to appear [_ to like carrots]]]]
b. What [did Dave think [that Mary believed [that John liked _]]]?

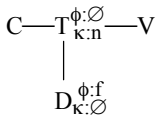
(28) a. John [seems [_ to tend [_ to appear [_ to like carrots]]]]
b. What [did Dave think [_ [that Mary believed [_ [that John liked _]]]]]?

There is something approaching a consensus in the literature that the examples in (26)a/b (raising to subject/RtS and *wh*-movement) are best viewed as involving linked local relationships of the sort pictured in (28)a/b, and not a direct "one-fell-swoop" relation as in (27)a/b. This is not entirely uncontroversial, though I will canvass a range of facts in Chapter 2 that are drawn from a variety of languages and which, taken together, strongly suggest that something like these so-called successive cyclic movements (SCMs) are real.

The TCG vision of these linked local dependencies can be schematized as in (29) for the RtS case (I will return to the *wh*-movement case below). (29)a gives a birds-eye view of the entire derivation, with the first stage building the matrix clause structure as in (29)b.

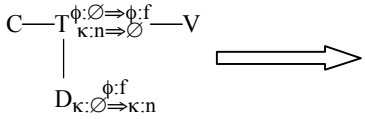
[15] I will be deploying reduced structures in this "horizontal" notation throughout this work. Structures of this reduced type are essentially those argued for in recent work of Brody (1999, 2003), and can be seen as related to more general efforts to downsize the array of label-types that analysis can appeal to. Collins (2001) is another such approach, but one which is incompatible with the central intuitions I will be developing regarding successive cyclicity (and "movement" generally). I will return to these issues below in §1.5.

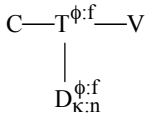
(29) a. 

b. 

Assume that T and D both enter the derivation with Case (κ) and agreement (ϕ) properties. T-ϕ is unvalued, requiring a relationship with another element with valued-ϕ (D-ϕ); assume the reverse holds for κ-properties (T-κ is valued, take κ to range over {∅, n, a, d}, for "unvalued", nominative, accusative, and dative/oblique, respectively; similarly, take ϕ to range over {∅, f, g, h} where *f*, *g*, etc. are stand-ins for more complex attributes and values like ϕ:NUM:*plural*, etc.).[16]

I assume that D and T enter into a reciprocal valuation, essentially swapping values, T retains ϕ and deletes κ, while D retains both valued properties, as follows (here and throughout, I will mark alterations of feature properties — valuation, deletion, etc. with transitions like "ϕ:∅⇒ϕ:f" = "unvalued feature ϕ gets value 'f'", or κ:n⇒∅ = "valued feature κ:n deletes" as in (29)c):

(29) c. 

d. 

At the next step of derivation a "like element" — T — is introduced. I am assuming that raising predicates (i) do not include a specification for an external argument (i.e., no

[16] This particular formulation of feature relations follows earlier proposals (Castillo, Drury, & Grohmann 1999; Grohmann, Drury, & Castillo 2000; Drury 2000).

small-*v*, though one is present to introduce the external argument of the most embedded clause, see (29)a above), and (ii) take defective T-complements (in roughly the sense of Chomsky 1999). Thus, the second T *could be* viewed as distinct from the first, since they differ in properties (the first/higher T has a ϕ-property that the second/lower T lacks):

(29) e. $C — T^{\phi:f} — V — T$
$\qquad\qquad |$
$\qquad\quad D^{\phi:f}_{\kappa:n}$

However, this is exactly the context in which we want the "reverse" of raising to occur. Suppose then that we assume the following as a first pass on our so-far unspecified matching relation $\mathfrak{R}$ from above:

(30) **MATCHING RELATION $\mathfrak{R}$:**
For two elements α and β, $\mathfrak{R}\alpha\beta$ iff:
(i) Either α dominates β or β dominates α
and
(ii) Either α subsumes β or β subsumes α

The condition makes reference to the notion of SUBSUMPTION, common in unification-based approaches to grammar deploying feature structures, specifically (from Shieber 1986:15):

(31) **SUBSUMPTION**:
A feature structure *D* ***subsumes*** a feature structure *D′* ($D \sqsubseteq D'$) if *D* contains a subset of the information in *D′*.

For example, given a node labeled X and a node labeled $X^{[F]}$, the former will subsume the latter since X contains a subset of the information in $X^{[F]}$. Subsumption is thus the "more general than" relation.

What we have above can be seen as a generalization of Chomsky's (1999) notion of AGREE, though (i) it introduces the possibility of both upwards and downwards valuation on the dominance order, and (ii) it extends the relationship to hold amongst *categories*. On Chomsky's view, in contrast, such relationships are asymmetrical, with unvalued elements ("probes") scanning their subordinate (c-command) domain for matching elements that can provide them with values ("goals").[17] Thus, asymmetry in valuation is taken to track asymmetry in formal ordering (e.g., goals can't typically value probes they c-command).

We will see later on some potential troubles with this statement of matching, in particular when applied to individual features it causes locality problems even for fairly simple examples (e.g., allowing valuation to go in either general direction on the dominance ordering will be seen to make it difficult to understand how *he saw her* can't mean *she saw him* — see §3.1 for discussion). For now however this way of thinking will allow us to give a sketch of how things work.

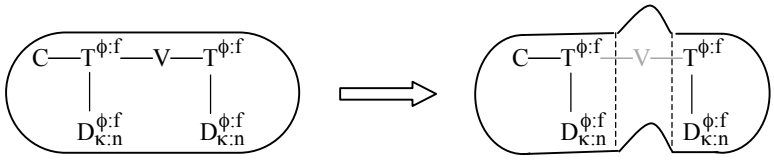
Taking the matching relation $\mathfrak{R}$ to involve categories, and not just features of them, might be taken to require some further comment. However, if there turns out to not be a good reason to have a fundamental division between categories and features, then this follows as a reasonably natural generalization of Chomsky's conception of

[17] However, a similar kind of κ/ϕ reciprocity as we have deployed here is present in a different form in Chomsky formulation (roughly: his idea is that κ gets valued as a reflex of agree with a ϕ-complete element; I won't comment further on this). However, on Chomsky's view the relation between the subject in a RtS construction is not directly associated with its surface/PF-output position. Rather, the traditional line of having this element orginate in its θ-position is assumed. This, I believe, holds onto a residue of D-structure. Though not coded in terms of a level of representation characterizing potentially unbounded objects (an infinite base component), it is nonetheless retained in the notion that items must enter the derivation *through* a θ-position. I see no minimalist motivation for this restriction, which is part of the motivation for exploring an alternative route regarding derivational directionality. However, we will see that the alternative top-down conception is acually demanded by the general view of contraction as applied to SCM phenomena.

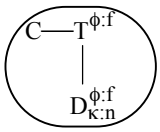
information flow and dependency-formation. What we will see rather is that features may be divided into classes which either serve to relate elements internal to a domain (e.g., ϕ-features) and potentially across such domains, while other features/properties (e.g., categorial distinctions like C, T, etc.) will serve to separate/distinguish elements within domains and across domains. I will return to unpack these ideas more explicitly below. The key idea to keep in mind is a feature-based view of domain boundaries — some properties are responsible for holding things together within domains, and others are responsible for keeping domains apart (or, as in SCM type relations, allowing limited overlaps between domains).

The general move that is being entertained here is to wed this generalization of AGREE with a version of Chomsky's (1995) discussion of CHAINS formalized as sets of context positions for an element α. I will return to elaborate on this point below, but note that what is being proposed here is essentially a "context" identification view of SCM (see §1.5). Before turning to these and other related matters in detail let us first complete the example derivation for RtS, and then take a look at one for *wh*-movement.

Returning to the derivation in (29): since ℜ holds of $\langle T, T^{\phi:f}\rangle$, these nodes are identified. Following the discussion above regarding identification and contraction and maintaining a coherent ordering in the workspace, this results in the following with the raising predicate itself (V) "splicing-out" of the workspace, and $D^{\phi:f}_{\kappa:n}$ being "reintroduced" at the bottom of the dominance order:

(29) f. 

And recall that this contracted workspace on the right-hand side here is really just:

(29) f'. 

The addition of the further raising predicates for the derivation of (29) goes exactly the same way, until the most embedded domain is reached. Prior to the introduction of the embedded *v*-element hosting external-θ, we would again have a workspace like that in (29)f/f'. Introduction of *v*, I assume brings with it a θ-feature:[18]

(29) g. C—T$^{\phi:f}$—v^{θ}
D$^{\phi:f}_{\kappa:n}$

θ-features, I will assume, correspond/connect to thematic predicates in a neo-Davidsonian sense (see, e.g., Parsons 1990, Schein 1993, Herberger 2000), relating a participant variable to an eventuality/situation. I suggest that the participant variables of such thematic predicates are inherently non-distinct and require valuation by κ/ϕ properties in order to be rendered locally distinct (not having these local properties around can result in the identification of such variables, as I will suggest is relevant for control and local anaphora, for example).

In the present derivation the step in (29)g involves a local A-relation that the superficially non-local relation to the matrix has been reduced to via successive contractions of the workspace — in effect "carrying-along" the matrix T κ/ϕ properties. It is this general property of these derivations which will allow us to dispense with

[18] Where the enclosure representing the workspace boundaries is not relevant I will simply leave it out.

reference to so-called "EPP-features" or their like (see Chapter 2 and 3). Intermediate specifier positions can exist, on this view, because (i) matrix ones exist, and (ii) intermediate positions involve an informational superset (more-general-than) relation with the corresponding matrix positions.

I assume that θ in (29)g takes the value of the dominating agreement feature (here ϕ:f) as in "θ[ϕ:f]". The suggestion is that θ-role assignment to the D-element is indirect, essentially importing a notion very similar in spirit to Williams (1994:33) notion of "vertical binding".[19] The outcome of this valuation then is as follows:

(29) h. $C—T^{\phi:f}—v^{\theta[\phi:f]}$

$$\begin{array}{l} C—T^{\phi:f}—v^{\theta[\phi:f]} \\ \quad\;\; | \\ \quad D^{\phi:f}_{\kappa:n} \end{array}$$

In general, I will be understanding A-relations and thematic discharge in this way — κ/ϕ exchange between T and D results in Case-marking of D and valuation of T-ϕ. The ϕ-property then associates with θ, which essentially takes this value as an index marking the participant variable (that is, $...\theta^{v}\langle \_ , e\rangle... \Rightarrow ...\theta^{v}\langle \phi{:}f, e\rangle...$). I will return to elaborate on this point of view.

A'-relations will be viewed somewhat differently. However, the fundamental notion of contraction and node/context identification will be understood to work in the same way for (e.g.) SCM involving *wh*-elements. Whereas the identifications for A-

[19] See also Williams (1983). A number of recent proposals of this kind have been entertained in the literature as implemented within a feature system alongside an adoption of an Agree-type relationship of Chomsky's (1999) sort. See Rizac (2004), Adger & Ramchand (2001), Butler (2004b), Koneneman & Neeleman (2003). Very similar notions have had a long tradition in frameworks that work exclusively with feature logics (e.g., HPSG; see Shieber 1985; Pollard & Sag (1994)). Williams proposal technically involves an indexing procedure connecting thematic roles with dominating projections, with predication then occuring between maximal projections as a matter of index sharing, thereby resulting in a connection to the lower (coindexed θ-role). See Castillo, Drury, & Grohmann (1999) for some earlier discussion of such features relations and the notion of VP-internal subjects; and see also Drury (2000).

movement involved the T-domain, the relevant relations in A'-movement will be between C-elements. Before turning to illustrate SCM of *wh*-elements, let us consider the local case of *wh*-movement:

(32) Who _ likes pizza?

We will illustrate a derivation for (32) down to *v* (remember: "top-down") to show how WH, κ/ϕ, and θ information will be understood to relate.

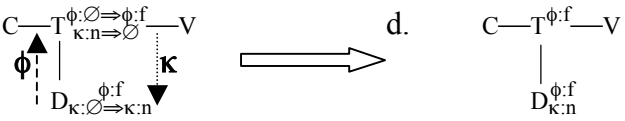
As A-relations serve to establish a set of feature-licensing relations resulting in an indirect view of θ-discharge, A'-relationships similarly provide a set of relations resulting in indirect κ-assignment. Assume that *wh*-elements come with a WH-property which (i) takes κ-features as values, and (ii) matches and deletes WH on D. Assume C is has unvalued ϕ as well, so we have the following in (33):

(33) a. $C^{\phi:\varnothing}_{WH}$—$T^{\phi:\varnothing}_{\kappa:n}$—$v^{\theta}$ (ϕ:f; $D^{WH}_{\kappa:\varnothing}$) ⟹ b. $C^{\phi:\varnothing\Rightarrow\phi:f}_{WH}$—$T^{\phi:\varnothing}_{\kappa:n}$—$v^{\theta}$ (ϕ:f; $D^{WH\Rightarrow\varnothing}_{\kappa:\varnothing}$)

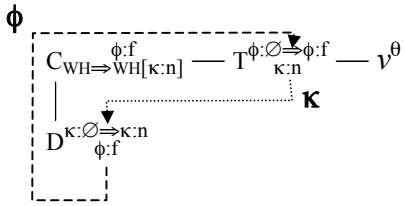
The now valued ϕ-property of C serves to value T-ϕ, and WH takes the κ-property of T as a value (WH[κ:n]). In virtue of these relations D-κ can now be valued by C:

(33) c. $C_{WH\Rightarrow WH[\kappa:n]}^{\phi:f}$ — $T^{\phi:\varnothing\Rightarrow\phi:f}_{\kappa:n}$ — v^{θ} ; $D_{\kappa:\varnothing\Rightarrow\kappa:n}^{\phi:f}$

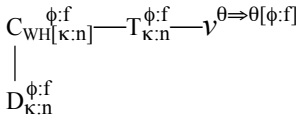
Thus, the presence of the WH-property serves to mediate the κ/ϕ swap of values. Like the A-relation case, there is some back-and-forth directionality to the flow of information in these feature-relationships. In A-relations, recall from above, we saw ϕ- and κ-valuation going in opposite "directions". Consider (29)c/d again:

(29) c. 

In the *wh*-movement case (A'-relation) the same holds, though the WH-property is alleged to be implicated in a mediating role (ϕ goes from D to C to T; κ goes from T to C to D):

(33) c. 

Now, as with the basic A-relation case discussed above, the ϕ-property of T will index the thematic position:

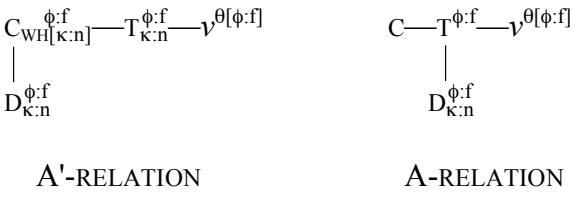
(33) d. 

The assumption here then is that indexing the participant variable of the θ-role with ϕ is to close-it off (saturate it) — the ϕ-property can only become connected with this θ-property if it has been valued in a way that has resulted in the assignment of Case. This happens in two possible ways now: (i) as in the A-relation, where ϕ is connected to an overt nominal marked κ, and so the θ-variable will be connected with the semantic properties of that element, or (ii) it is connected with a "bound κ", associated with the upper WH property — that is, connected with an "individuator" in our terms.

There is a version of a traditional view being implemented here. In GB-era terms (e.g., Chomsky 1981) we have the ideas that "κ-marked traces" are "variables" and that in

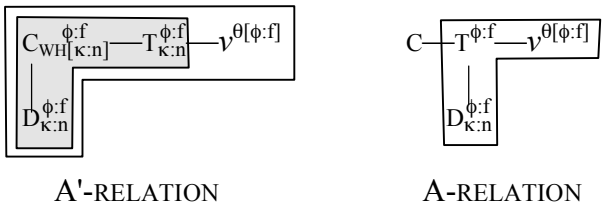
general κ/ϕ are intimately connected with θ-theory. I will return to these matters in further discussion in Chapter 3.

Let us consider now the picture we have of local A- and A'-relations side-by-side:

(34) 

A'-RELATION A-RELATION

In the A-relation, we have ϕ-features which form the connection between elements,[20] and in the A'-relation there is a mix. That is:

(35) 

A'-RELATION A-RELATION

This is one reasonable way of specifying the flow of feature-licensing information in local domains. ϕ-information flows from items that are specified to those that are not, filling in the values along the path; and κ-information does the same, though in the opposite direction on the path.

The suggestion is that once we have a valuation mechanism of the AGREE-sort that has recently been appealed to in elaborating minimalist syntactic theories (Chomsky 1999), then it seems there are some fairly straightforward ways to make it do most, if not *all* of the work.

[20] I am ignoring here any ϕ-properties that may be associated with C in the A-relation example. We might assume that C-ϕ can manifest an open clause, as with relatives, if we attribute a non-WH operator property to C. Later I will explore the idea that ϕ on non-finite C (without an operator property) is what allows indexing of non-κ-marked θ (control).

We need mechanism to "build", for example, extended projection sequences in the verbal domain. On a traditional movement story, we interleave the building of such sequences via merge operations with movement/remerge relationships involving nominal expressions as each of the relevant levels of structure is constructed. So, a θ-assigning element relates to a nominal, discharging its role to that nominal; further operations add higher projections specifying other licensing properties (Case/agreement), which we then need to relate to the nominal element as well (so we have an A-movement); further categories/features are added above that, which may provide yet further licensing properties, and so we relate the nominal expression yet again to the next highest layer (so we might have an A'-relation).

But we can view a local A'-relation complex in at least the following two different ways: (i) D^{wh}-V, D^{wh}-T, and D^{wh}-C, or (ii) C-T-V + D^{wh}. Below (§1.5.1) I will review Chomsky's (1995) discussion of chains as sets of contexts, and suggest that coupling a simplified version of that with an AGREE-type mechanism yields the following picture:

(36) a. In local sequences of categories (which will, in accordance with the anti-recursion provision in (9), not include repeated "like elements") like features co-value,

and,

b. Encountering like categories results in a similar "co-valuation" (like elements are identified, though, as with feature-valuation generally, only so long as one subsumes the other).

Intuitively, categorial differences in local domains prevents collapse of nodes — dislike elements "repel" one another. But this does not stop like features from identifying with

each other within such local domains (likes "attract" one another).[21] Across local domains, we have the possibility of interactions because the edges of such domains can become identified by keeping this same attract/repel logic in place (a like category is introduced, and this can result in identifications which allows a kind of domain expansion — as we saw above — a kind of copying/lowering).

Note also that the feature-relationships in our schematic A'-relation builds-in a useful property. Consider how a standard "copying" view of SCM of *wh*-elements works:

(37) a. Which picture of himself did John think Bill liked _ ?
b. [which...self] did John think [which...self] Bill liked [which...self]

It has been noted that on the copy view of such movements there must be some operation which ensures that the actual *wh*-operator does not appear in all the lower copies. As Safir (1999: 591) points out, quantifiers cannot bind other quantifiers, and somehow the lower copies must be understood as variable-like elements. Accordingly, one or another variant of the following sort of operation is typically taken to be in effect (Munn 1998 calls this "Make OP"; this particular illustration is taken from Safir's discussion):[22]

(38) a. Whose mother did Bill see _ ?
b. **whose** [**whose** mother] did Bill see [**whose** mother]
STEP❶: *"lift" the operator out*
c. whose [*x* mother] did Bill see [*x* mother]
STEP❷: *make variable/delete*-WH

This operation is built into the WH-licensing discussed above (see (33)a/b). The implementation is in terms of matching D and C WH-properties, with deletion of this

[21] See van Riemsdijk (1998) the working out of an intuition which similarly makes use of attract/repel, but in a rather different way from what I am entertaining here (his "Categorial Feature Magnetism").

[22] See Chomsky (1995:203), Munn (1994:399), Fox (1999, 2002), and Safir's (1999:591) discussions. See also Fox (2003) and the notion of Trace Conversion.

property from the D-element, but the effect is the equivalent of D projecting its WH-properties to C (there are several way that we could implement the idea, the one given above is simply one such way).

If we keep with our assumptions above, including the assumption of a top-down derivation, then we can understand the "operator" properties to be housed in C, leaving the D-*wh* phrase itself with a "hole" indexed by its κ-property. The result then is that the local structure provide above (repeated here) will have a logical form of the sort in (39)b:

(39) a. $C_{WH[\kappa:n]}^{\phi:f}$—$T_{\kappa:n}^{\phi:f}$—$v^{\theta[\kappa:n]}$
|
$D_{\kappa:n}^{\phi:f}$

b. WH[κ:n] ...[κ:n]... $\theta^{v}\langle\kappa{:}n, e\rangle$

(i.e.,...*wh*(x) ... [...*x*...] ... [...P*x*...]
whose *x* ... [...*x* mother...] ... [...P*x*...]; as in (38)c)

Now the top-down structure of this story makes it possible to understand the equivalent of a Make-OP sort of operation as happening on the first step (when D and C are integrated).

However, note that on longer distance *wh*-movement (the above example is local association with subject κ/ϕ-θ) the WH won't have a κ-value until it encounters a lower valued-κ. According the logic of category identification and lowering sketched above, we might take this to result in an operator being successively lowered to each new domain edge, along with the residue of the *wh*-element itself (e.g., roughly [*x* NP] = "$D_{\kappa:\varnothing}$" here):

(40) a. $C_{WH[\kappa:\varnothing]}$—….—$C_{WH[\kappa:\varnothing]}$—….—$C_{WH[\kappa:\varnothing]}$—….—
| | |
$D_{\kappa:\varnothing}$ $D_{\kappa:\varnothing}$ $D_{\kappa:\varnothing}$

This gives us part of what we may want for SCM, which is variable-like elements in all the intermediate positions, but it also gives us something that we don't want, namely the *wh*-operator at all the intermediate positions as well (i.e., recall the point from above that quantifiers don't bind other quantifiers).

Below I will suggest a way, relating to some ideas proposed by Uriagereka (1999) and from previous work of my own (Drury 1998), which appears to have the right properties to naturally yield the result that we *do* want, which looks more like this (in terms of what we want in the output structure):

(41) $C_{WH[\kappa:\varnothing]}$—....—C—....—C—....—
$D_{\kappa:\varnothing}$ $D_{\kappa:\varnothing}$ $D_{\kappa:\varnothing}$

I return this briefly below in discussing MSO-systems (§1.4.1), and again when we turn to analysis in Chapter 3.

1.3. Summary So Far,...& The Path Ahead

We have so far introduced a few key ideas. Let's sum up before proceeding. We have posited a workspace/output-distinction (henceforth: WS/O-distinction). In the course of elaborating on the key intuition we have suggested that the distinction be understood as a dividing line between the systems that manipulate elements by handling their syntactic properties only, versus those that handle the π- or the λ-properties. Moreover, *some* of the formal/syntactic properties (e.g. WH, κ/ϕ, and θ) have been understood to play a direct role in mapping out logical form distinctions. One way to look at this claim is to view the "workspace" as we have sketched it so far as "the interface" between the sound-meaning systems and *whatever system(s) are responsible for the general ordering properties of*

lexical/functional extended projection sequences. The suggestion above was that the workspace is a dynamic interface between "the lexicon" and the PF/LF systems.

Prior to this, we outlined some consequences of workspace restrictions stated over ordering and category distinctness, and showed how the combination of these ideas yields a schema for analysis that seems to provide for a novel view of successive cyclic type movement. The mechanics were suggested to follow on a natural generalization of an AGREE-type operation/relation of the sort studied in Chomsky (1999), broadened to include categorial identifications in a way that allows cross-domain interactions in virtue of a "carrying-over" of higher properties of elements into lower domains (via node-identification under subsumption). Some specific assumptions for A- and A'-relationships were sketched, providing a general (though reasonably detailed) outline of the approach to be further constrained and deployed below (Chapter 3).

The availability of the type of analyses relevant to SCM phenomena will be argued here to be extremely interesting in the minimalist setting. As we will see, the approach makes available a route for analysis which does away with any appeal to (what I argue are) spurious movement-driving features that have been taken to underwrite SCM in much current minimalist work (so-called EPP/P-features — which I will generally refer to as Move/Merge-features or "M-features").

There are, however, a number of component ideas in play here that require some further background discussion before proceeding. For example (i) the motivations for the reduced phrase-structure graphs deployed above, (ii) the idea of relegating all "movement" relationships to one or another type of category/feature relationship on the dominance path, and how this could be connected with other existing lines of thought

regarding movement/chains, (iii) the generality of the WS/O-distinction, (iv) the conceptual connections to other proposed MSO-type systems as well as to TAG approaches.

Additionally, one particular consequence of the TCG view, mentioned briefly above, is worth bringing up again here before heading into more detailed discussion. The general structure of the account of SCM effects demands that syntactic derivations be viewed as assembling structure roughly top-down, instead of bottom-up as assumed in Chomsky's widely adopted Bare Phrase Structure (BPS; Chomsky 1994, 1995). This move (inverting the direction of derivation) on its own teaches us nothing about successive cyclicity.[23] However, coupled with the right alternative views regarding structure and categories/features and how they might be generated by a derivational system, directionality can be seen to play a crucial role. This somewhat unorthodox outcome converges with the results of a number of other recent investigations which have

[23] Contra a discussion in Terada (1999), who suggests that successive cyclicity effects can be better understood within the incremental/left-to-right view of derivations proposed in Phillips (1996). While I agree that derivational directionality may be important, nothing in Terada's discussion establishes this conclusion. In Terada's proposal intermediate movements are stipulated to be driven by features as in many other minimalist approaches (positing what I call Merge/Move-features or M-features see Chapter Two below). Terada appears to think that having the ultimate licensing feature (e.g., +wh for question-formation) checked "first" helps in some way with "look-ahead" issues. But the logic relies on a spurious division between the 'top-most' licensing properties and lower ones (like Case/ϕ and θ-properties). The "look-ahead" problem is *symmetric*. In a bottom-up approach Case/ϕ and θ are locally licensed but a (e.g.) a *wh*-element must somehow eventually reach its corresponding licensing context, so where there is multiple embedding there is a look-ahead issue (the *wh*-element needs to "know" that the right licensing property will eventually show up). But the same goes on a top-down (or left-to-right/incremental) view, just the other way around (a *wh*-element needs to "know" that Case/ϕ and θ information will eventually show up, though its *wh*-property may be licensed immeidately upon entering the derivation). The mystery/problem/puzzle of SCM effects is rather about why there are ever any movement operations *other* than those which would connect these basic (*wh*, Case/ϕ, θ) properties. Why are there *intermediate* movements (chain-links/traces)? Positing M-features (e.g., EPP/P-features in Chomsky's parlance) to drive intermediate movements (as Terada does) is not a solution — that is the *problem*. I see nothing in Terada's discussion that adds to our understanding of why derivations ought to have a non-standard "direction" nor why successive cyclic movement ought to exist. On the present approach, in contrast, direction of derivation is demanded in order for things to work. I thank Cedric Boeckx for pointing me to Terada's paper.

reached similar conclusions regarding directionality and syntactic derivations (see in particular Phillips 1996, 2003).[24]

In addition to this difference with respect to standard BPS, the TCG approach differs as well from the structure of TAG derivations, which taken to obey a Markovian condition insisting that it be locally determinable whether a given pair wise combination of tree-structures is licit or not. One effect of this condition in TAG is an ordering freedom which for cases beyond pair-wise combination of trees allows the possibility of a many-to-one mapping of derivation structures to derived structures.[25]

I think a large part of the interest in the mechanics of the TCG system is that it has this fairly abstract general requirement regarding derivational directionality. What I am unsure about at present is what the ultimate significance of these ordering differences might be for the study of grammar *qua* "system of human knowledge" (i.e., as properties of a competence-level theory).

There are, however, some obvious points of interest to be made with respect to connecting grammatical theory and parsing (and perhaps production). The treatment of linked-local relationships here in effect introduces a way that displaced constituents can be in sense buffered as structure is expanded and then re-accessed as lower domains are constructed. The structure of the TCG account thus does not require the explicit add-on of a memory stack or related storage devices that have been appealed to in the past in

[24] Other work along this same line includes previous work of my own, Drury (1997, 1998, 1999a,b), and a number others including Boeckx (1999), Guimarães (1999, 2004), Richards (1999, 2001), Terada (1999).

[25] That is, as in some other approaches (like classical Categorial Grammar (CG) or Steedman's Combinatory Cateorial Grammar (CCG)), TAG derivations manifest a kind of so-called "spurious ambiguity". This label is somewhat of a misnomer both in CCG and in TAG, as both approaches have suggested that the relevant derivational ordering alternatives are not in fact "spuriously" ambiguous but rather do make linguistically significant distinctions. See Frank (2002) for discussion.

discussions of filler-gap dependencies in the context of parsing theories. The functional equivalent of such a device is, as saw in the sketch offered above, an essential component of the basic mechanics. I will not be concerned here with these issues, though its worth keeping in mind in the background.

In the next section I back up to consider the general structure of some proposed MSO-systems and TAG, looking at the structure of these approaches in terms of our WS/O-distinction. Following this I turn to some technical discussion further motivating some of the component ideas of implementation of TCG pursued here.

1.4. MSO-Systems, TAG, & Generalizing the WS/O-Distinction

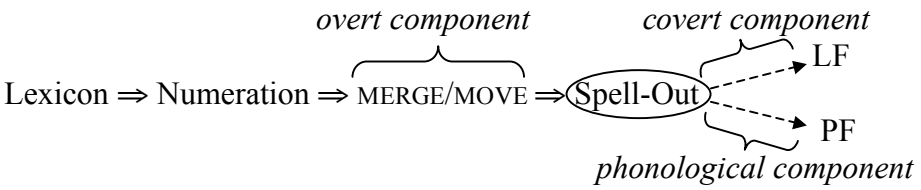
MSO-systems as they have emerged within the MP, generally speaking, are derivationally oriented models which parcel structure assembly into principled stages in virtue of applications of an operation called Spell Out (SO). Depending on assumptions varying across implementations, different sorts of MSO-systems effect different partitions of syntactic complexes (or stages of derivation) into local domains or chunks. Common across implementations is the general idea of SO as an operation that is periodically applied in the course of derivation resulting in a reduction or *contraction* of structural descriptions by shunting or transferring portions of structure to neighboring systems with which the syntax must interface. In this manner evaluation of certain aspects of well-formedness of syntactic complexes is thus suggested to be divided such that sub-parts of structure are independently inspected by the principles governing the interfaces.

The general idea of multiple spell out has a number of antecedents in earlier literature.[26] Within the context of the development of the Minimalist Program (MP) it arose in consideration of the architecture proposed in Chomsky (1993), which contained a weak residue of Government & Binding (GB) theory's level of S-Structure. Rather than a full fledged level of representation, this S-structure hold-over was simply taken to be a "point" of derivation as discussed in Chomsky (1995:229):

> 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***, [emphasis mine-JD] 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[umeration]→λ computation. We call the subsystem of C_{HL} that maps Σ to π the *phonological component*, and the subsystem that continues the computation from to LF the *covert component*. The pre-Spell-Out computation we call *overt*.

This passage characterizes the core properties of the minimalist Y-model:[27]

(42)



The development of MSO-systems in more recent work questions the idea of a single "point" of Spell-Out.

1.4.1. MSO and Linearization

Uriagereka (1999) was to my knowledge the first to propose within the setting of the MP that we ought to regard spell-out not as a single point in the syntactic derivation, but

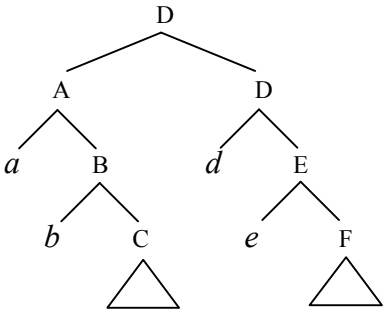
[26] See, e.g., Jackendoff (1972) and Bresnan (1971,1972).

[27] I have made no mention so far of the notion of "Numeration" in this model (as an intermediary between the Lexicon and the syntactic derivation). This object will play almost no role here, though see our concluding discussion in Chapter 3.

rather as a procedure that can apply more than once, perhaps limited by economy principles (e.g., perhaps of the general Last Resort variety, mandating that no operation occurs unless necessary to ensure convergence).

Uriagereka's proposal draws on the work of Kayne's (1994) proposed correspondence relation between linear precedence and c-command. Supposing with Kayne that asymmetric c-command relationships map to linear precedence, and taking Chomsky's view of spell-out as a process of stripping away "those elements relevant only for π" (see above), Uriagereka suggested that we identify domains for spell-out with sub-structures which constitute total/connected c-command orders. He argues that this allows a simplification of Chomsky's (1995) implementation of Kayne's LCA which avoids the need to state an induction step to cover the linearization of parts of complex structures with respect to parts within other such complexes. To illustrate, consider:

(43)



A version of Kayne's general idea would be to claim that where α asymmetrically c-commands β, α *precedes* β.[28] For the sub-structure dominated by A above this would yield the order ⟨*a*, *b*, ...⟩. But while we say that A asymmetrically c-commands the

[28] See Kayne (1994) for a conceptual argument that asymmetric c-command ought to be mapped to *precedence*, as opposed to *subsequence*. See also Uriagereka (1998) and Chametzky (2000) for important related discussions.

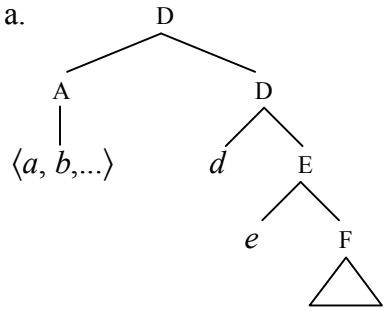
elements dominated by the two-segment category D, the elements dominated by A do not.[29] Therefore we need to add a step to the linearization procedure. That is, there are two separate c-command domains here which each independently constitute a total/connected order. But *b* and *e*, for example, are not so ordered. So we need a step to tell us that all the parts of the A-substructure are to precede all the parts of the D-substructure so long as A asymmetrically c-commands D (see Kayne's (1994) discussion for his handling of this issue).

Uriagereka's suggestion is that since independent c-command domains are trivially linearizable (i.e., they do not require appeal to an induction step of the sort informally sketched above), they independently undergo spell-out. The output of this procedure could be understood as a flattened structure which we regard as still "there" in the computation, but whose internal parts are frozen and therefore unable to undergo further interactions in any later stages of derivation. Alternatively, separately linearized substructure could be regarded as sent to the PF-component, leaving only a residual place-holder element "@", with some minimal specification of category/feature information relevant to the interaction of the spelled-out unit A with the rest of the structure.[30] These two options are sketched here:

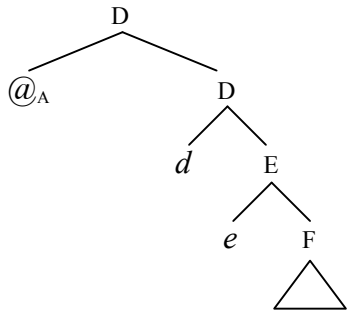
[29] I'm mixing in Kayne's assumptions regarding specifiers as adjuncts — this assumption is replaced in Chomsky's discussion by assumptions regarding intermediate-phrasal-level "invisibility" in order to get the right asymmetries for c-command to hold. This is inessential to the overview I am giving of Uriagereka's proposal in the main text, though I should make it clear that he otherwise follows Chomsky's BPS approach in his specific formulation.

[30] Perhaps simply core licensing properties like Case, agreement, *wh*, etc.

(44) a.



b.



The intuition in both implementations is that spelled-out structure functions like a derived terminal, allowing a trivial statement of structure/order correspondence (α precedes β ↔ α asymmetrically c-commands β). Precedence between elements which do not themselves enter into an asymmetric c-command relationship fall out from the structure of derivations involving separate linearization of c-command domains.

Uriagereka specifically proposed that *non-complements* might be understood as the structures that must undergo independent linearization in the sense just sketched, and further argued that Condition on Extraction Domain (CED) effects (Huang 1982) could be understood to follow from this. So the cases in (45) would be understood to be ungrammatical because the bracketed sub-structures would have to be independently spelled-out, making their contents "frozen" and thus inaccessible to further merge/move operations (the relative clause in (45)c would be out on the standard assumption that these structures involve adjunction and thus are also non-complements):

(45) a. *What do [explanations of *e*] bother you?
b. *What was Mary bothered [because Peter explained *e*]?
c. *What do you know [[the girl] [that _ explained *e*]]

Thus, on this linearization driven view of MSO, we have a potential account of at least these particular so-called strong islands.

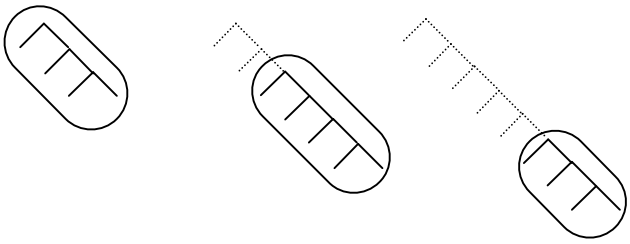
However, its not obvious why it should be that subjects and adjuncts need to be independently spelled-out, as opposed to the structures they associate with — either option would seem to permit the simplification of the linearization procedure.[31]

In Drury (1998, 1999) it is proposed that Uriagereka's linearization-based view of MSO be put together with incremental derivations of the sort proposed in Phillips (1996) (see also Phillips 2003). We can frame a version of this proposal within our WS/O-distinction as follows:

(46) **Workspace Connectedness (C-Command):**
The elements in a given syntactic workspace must manifest a connected c-command order (i.e., for every *x, y* in the set, either *x* c-commands *y* or *y* c-commands *x*)

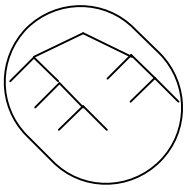
Recall our top-down schema of the WS/O-distinction from (2), repeated here as (47):

(47)



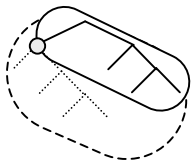
This derivation would have workspaces which all obey (46). The following workspace would not:

(48)



[31] See Drury (1998, 1999), Johnson (2000) for some discussion of this and related points. See also Uriagereka (2002) for criticism of Uriagereka (1999), and a working through of some alternative possibilities that for reasons of time and space I will not consider in the present work.

On a top-down view of structure expansion, the c-command path that was first assembled would have to "spell-out". We can envision spelled-out structure as being "ejected" from the workspace as follows, in the spirit of our proposed contractions discussed above (§1.1 & 1.2):

(49) 

The shaded node above would still be visible/present in the workspace, but the structure it dominates would be excluded (removed from the workspace = spelled out).

There are numerous details here that require elaboration (e.g., with respect to symmetry vs. asymmetry of c-command between sisters; see Kayne 1994, Chomsky 1995), but the basic idea would be that the workspace would be limited to only contain trivially linearizable structure as in Uriagereka's proposed simplification of a Kayne-type order/structure correspondence. This view then doesn't include anything that gets around the objections raised above however (e.g., regarding which of two sisters spells-out, etc). I put it on the table now to underscore the generality of the key idea of viewing spell-out of sub-structures as essentially being "kicked out" of the active workspace (we will see other illustrations below).

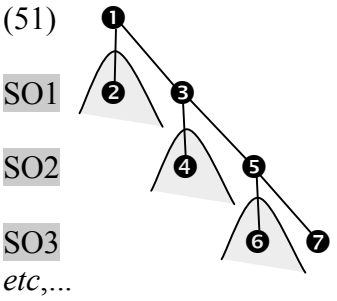
Note that the reduced structures introduced in the sample derivations in §1.2 were suggested to involve only a dominance ordering. Suppose that we were to modify the proposed restriction in (46) to refer to dominance in our reduced structures, as in (50):

(50) **Workspace Connectedness (DOMINANCE):**
The elements in a given syntactic workspace must manifest a connected dominance order (for every *x, y* in the set, either *x* dominates *y* or *y* dominates *x*)

At each branching point in the top-down expansion of the domination sequence, something would have to "spell-out" (be voided from the workspace).

To illustrate, take the following nodes to be introduced in the order indicated by their number. The initial sequence ❶—❷ would satisfy (50), but the addition of ❸ would add a domination relation between ❶ and ❸ but no such relation between ❷ and ❸, so we could take ❷ to be required to "spell-out" (be voided from the workspace). Subsequent spell-outs would occur for the same reason (SO1, SO2, etc.). The result of this condition is binary branching.

(51)



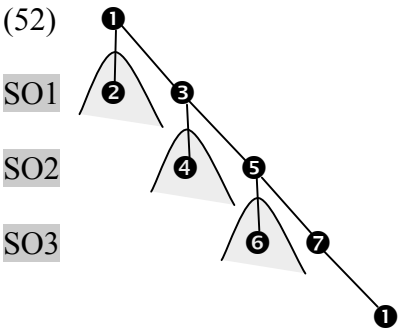
Note that these spell-outs would have to work differently than the general shape of Uriagereka's proposal. Since the connectedness requirement would be stated over the dominance relationship, the even-numbered nodes would literally have to be "gone" from the workspace. So this raises the question as to how they might interact with later structure.

However, recall from our sketch of A- and A'-relations above that the connection between items is mediated by various kinds of feature-exchanges (valuations, etc.). On this view, for example, ❶ and ❷ above might relate in such a way as to leave the appropriate feature-relationships visible on ❶ alone (and thus still "in" the dominance

path). For example, it was suggested that θ-assignment to a "subject" is mediated by the interrelationships between D and T with respect to κ and ϕ properties, with the ϕ-properties serving as a link to the thematic variable (θ) introduced by the verb. Following these valuation exchanges, D is marked for κ and D and T are connected by co-valued ϕ.

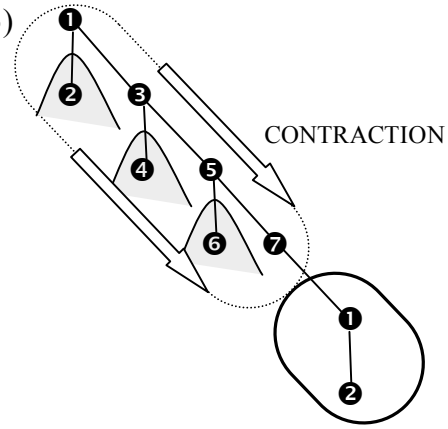
Given this general picture, we might consider the possibility of an element being spelled-out (e.g., like ❷ above), and then to *re-entering* the workspace in virtue of a later instance of node-identification. For example, suppose that ❼ introduces another instance of the ❶ type:

(52)



If (❶, ❶) meets our matching relation ℜ, then identification would occur. But, as argued above, this would require the splicing-out of all the intervening odd-numbered nodes in (52). But nothing would prevent the copying/lowering of ❷, as this would create no ordering conflicts, nor would it violate the connectedness condition:

(53)



Intuitively, this would have the effect of an element (here: "❷") leaving the active workspace (being spelled-out), and then "returning" again to the workspace as its context was copied via node-identification. Note that further additions to the structure from the point in (53) (e.g., associating a new element directly with ❶) would result in ❷ having to spell-out *again*, in order for the workspace to comply with connectedness.

But, ❷ could dominate arbitrarily complex structure, so its not obvious that we could, given our distinctness condition on the WS, simply reintroduce such complexes at the bottom of the domination order in virtue of the node-identification illustrated above. However, recall that the ❶-❷ relation has been understood to involve some feature-value exchange. This suggests the possibility that we could understand the copying/lowering as involving simply a reintroduction of a simple label, implementing the notion of a derived terminal in Uriagereka's (1999) sense. That is, the node identification (❶,❶) would result in reintroducing a simplex marker for ❷ above, facilitating the copying/lowering we require but not reintroducing all of the potentially complex structure dominated by ❷ into the workspace. This would still allow us to see the "left-branch" material of ❷ being

successively reintroduced into the output structure, in virtue of the initial feature-licensing connection established in the matrix position.

The "context" element itself (❶) would, in contrast to ❷, be a constant presence in the workspace (it does not spell-out, get reintroduced, spell-out again, as ❷ would). These mechanics are relevant to a discussion at the end of §1.2 regarding cyclic-*wh* movement and an operation of the "Make-OP" sort. There we referred to the difference between the following two sorts of structures, and suggested that the former introduces a copying of operator-elements that we do not seem to want; whereas the latter seems to have the right properties:

(54) $C_{WH[\kappa:\varnothing]}$—….—$C_{WH[\kappa:\varnothing]}$—….—$C_{WH[\kappa:\varnothing]}$—….—

$D_{\kappa:\varnothing}$ $D_{\kappa:\varnothing}$ $D_{\kappa:\varnothing}$

(55) $C_{WH[\kappa:\varnothing]}$—….—C—….—C—….—

$D_{\kappa:\varnothing}$ $D_{\kappa:\varnothing}$ $D_{\kappa:\varnothing}$

The node-identification procedure, plus the now strengthened condition on workspace ordering (in terms of connectedness of the dominance order) makes available a distinction between C and D of exactly the sort we want. C is constantly "there" in the workspace, while D must spell-out, be re-entered to the workspace, spell-out again, and so on as the local domains are established in a top-down fashion.

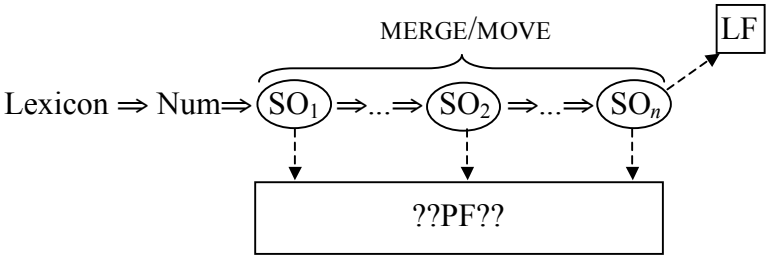
I will return to these general ideas in the course of developing some analyses to explore the matter a bit (see Chapter 3, especially §3.2.3 & §3.3.3), but the general suggestion is that we think of connectedness of the dominance ordering and the anti-

recursion restriction as working together to factor complex structures into natural local domains, creating major boundaries at both points of branching and points of recursion.

Every syntactic theory of which I am aware needs to say something about (i) a theory of types, and (ii) formal ordering properties. The suggestion here is that it may be possible to get these very basic notions to do quite a bit of work for us if we can seek out the right combination of conceptions of each (as suggested also in the work of Epstein, Groat, Kawashima, & Kitahara (1998), though with rather different conceptions pursued).

For now, observe that the general idea of MSO as proposed by Uriagereka removes a "point" of spell-out from the familiar Y-model in (42) above, in favor of a more dynamic looking system with multiple such points:

(56)

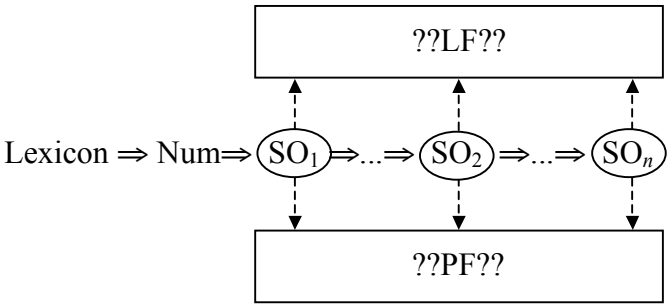


This kind of derivational architecture raises a number questions about the status of levels of representation. If spell-out is not a unitary point, do we need to amend our conception of PF as a unified object in the sense of "level of representation"? Uriagereka suggests that his dynamic view of spelling-out is compatible with a level-less conception in which there is no unified object that is subject to a single-step evaluation with respect to grammatical conditions. On his view, separately spelled-out sub-structures could be seen as being sent to the interface systems separately, leaving the grammar architecture with a PF-component but no level of this sort *per se*.

However, note that there is nothing in the MSO view that *requires* us to abandon levels of representation. It could simply be that MSO establishes the PF representation in the steps given by the independent instances of linearization, but that it still forms a coherent connected object that can be subjected to further (PF-system) operations. That is, we can simply regard levels as incrementally established. But it matters a bit what we take the "levels" to actually be. I will return to this point, but note here that this is roughly the content of the WS/O-distinction (a limited span derivational workspace that incrementally builds an output representation). However, the suggestion above was that the object which is incrementally assembled is "syntactic" in the sense of manifesting the formal ordering properties laid down in the workspace, but which is an object of the extra-grammatical PF/LF-systems in terms of what sorts of properties/features/categories populate this object (what sort of properites "decorate its nodes", if you will).

This PF-motivated view of MSO raises questions about LF too, in particular: are there reasons for thinking that SO is involved in similar kinds of divisions of derivations on the LF-side of the grammar? Asymmetric c-command, after all, is taken to be relevant for scope and binding and the like; are there thus reasons for thinking that SO sends material to both the PF and LF systems, leaving us with a model like (57)?

(57)



An architecture of this general sort — with both PF- and LF-relevant spell-outs (setting aside the staggered vs. uniform views) — has in fact been suggested in connection with Chomsky's recent proposals regarding phases (derivation by phase: DbP) which brings us to the DbP/MSO view of derivations and their handling of SCM.

Note that Uriagereka's linearization-based view of MSO on its own does not offer us anything immediately obvious in terms of helping to understand SCM phenomena. C-command domains are themselves potentially unbounded in depth, whereas the key point about SCM is that roughly clausal (or perhaps smaller) units form special domains that "punctuate" otherwise longer-distance looking relationships into linked-up local ones.[32]

However, this view might be interestingly fit together with something like Chomsky's phases which constitute a subset of the domains picked out by Uriagereka's linearization-based conception. In later discussion I will suggest that Uriagereka's central idea maybe best viewed in terms of general formal ordering restrictions on the workspace along the lines sketched above — that is, not specifically tied to linearization demands, but rather to the internal coherence of ordering properites in narrow syntax.[33] I turn now to a phase-based MSO-system and cyclic movement.

---

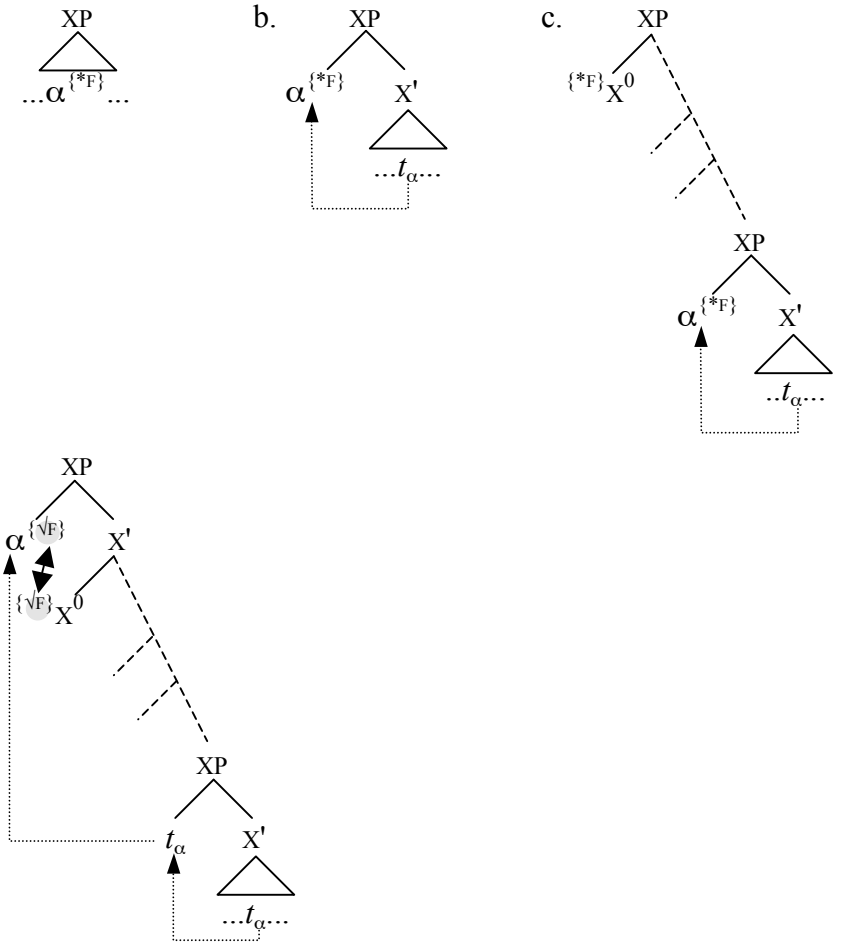
[32] I borrow the notion of "punctuated" relations from Abels (2003); see Chapter 2. See also Bošković (2002) for a discussion evoking ideas from Aoun (1986) and others regarding the notion of having certain phrase boundaries serve to "break" chains.

[33] Again, see also Uriagereka (2002) for critical discussion of his own previous proposals, and some alternative suggestions that I won't be considering in the present study. My own view here will involve a formulation akin to the notion of Workspace Connectedness offered in (46). Its worth noting here that Hornstein & Uriagereka (1999) appeal to a similar kind of interface motivation for spelling-out as this linearization based conception, though on the other (LF) side of the grammar. Briefly: they examine the possiblity that moved DP's may project their label to determine the type of their dominating category, essentially allowing the potential of taking clauses as "external arguments" of DP. They suggest this as the syntax supporting generalized quantifiers, and argue that, like the left-branch-type effects at PF, the projection of D-labels in their moved-to target positions creates an analogous kind of effect at LF. Technically, they argue that DP's do not so project their labels in the overt syntax, but in the covert component a "re-projection" occurs, essentially allowing specifiers to determine the types of their containing phrases.

1.4.2. Phases/MSO & Successive Cyclicity

Illustrated in (58)a-d is a general schema for a fairly standard derivational approach to such linked-local relationships familiar from the MP; below we locate this general line of thinking within Chomsky's (1999) approach.

(58) a. b. c. d.



In the context of the MP, the moving element α is understood to have some property {F} which requires that α enter into a licensing relationship that cannot be established in its initial position within the subtree marked XP in (58)a (so we mark {F} here as {*F} until

licensed, and as {√F} afterwards).[34] For the case of local *wh*-movement (e.g., *who did John see* _ ?) we understand the *wh*-element to be directly displaced (perhaps copied and/or remerged) to the surface position where the licensing/checking of this feature {*F} can occur via a *match* with a corresponding feature housed in the target position (pictured in (58)d).

Of interest here is that the dependency between the top and bottom positions implicated in this sort of movement is alleged to not be "one-fell-swoop" but rather to involve linked local relations. That is, for reasons which vary across specific models, α may be required to move to an *intermediate* or *non-target* position in which its unlicensed property {*F} is not satisfied — pictured in (58)b — on its way to the final/target position where this licensing can (in fact, *must*) occur. Some other property may be satisfied by this intermediate movement, perhaps some property of this intermediate position or perhaps in virtue of constraints built into the movement operation itself.

Or, it may be that both sorts of motivations are in play. For example, in Chomsky's DbP formulation, elements must move to intermediate positions in order not to be stranded in an independently spelled-out domain. The idea of the DbP approach is that structures may be evaluated piece-meal, so unlicensed elements must be displaced from such locally evaluated structures in order not to crash the derivation. Chomsky motivates these "escape hatch" type movements from locally evaluated domains by positing features/properties of potential intermediate movement landing-sites to play the

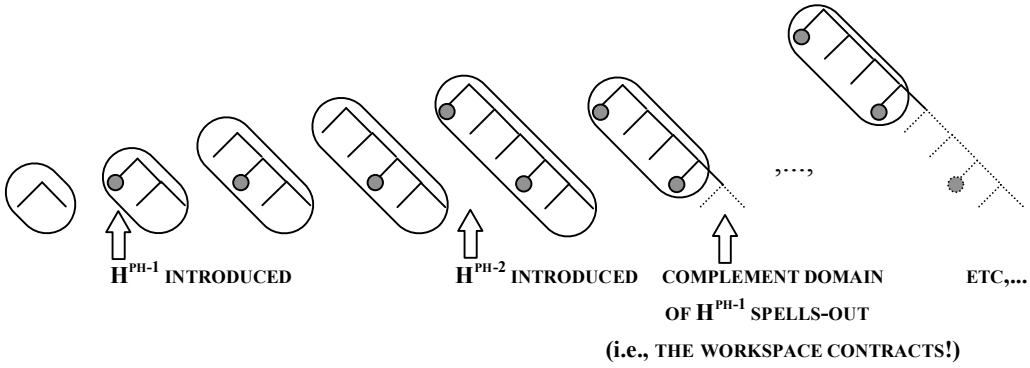
[34] I don't care for present purposes about any deletion/erasure procedures that may be part of such licensing/checking.

role of the local licensor for such operations (I will later on refer to these putative features as "Merge/Move-features" or "M-features").

Chomsky suggests that certain syntactic categories are phase-inducing, and that when multiple such heads are introduced into the derivation this results in systematic limitations on what remains "in active memory" versus what material is spelled-out and thus no longer accessible to computation. His Phase Impenetrability Condition (PIC) insists that for a given phase head H^{PH-1}, when a second such head H^{PH-2} is introduced the *complement domain* of H^{PH-1} spells out.

Abstractly then we have a derivation (on Chomsky's assumptions, a bottom-up one) with periodic applications of spell-out that in our workspace formulation we can picture as follows (phase-inducing heads are marked):

(59)



Spelled-out structure under the WS/O-distinction, as outlined above for the linearization-based view of MSO, is simply structure that is no longer in the workspace. Again, the WS/O-distinction as I am understanding it here is extremely general, and it is intended to be so. This gives us a common platform to discuss these different (otherwise technically rather different) proposals. It is, however, more than just another "way of talking". There is a substantive claim implicit here which I am carrying across the discussions of the

TCG approach as sketched above, Uriagereka's linearization-based MSO, and now the derivation-by/spell-out-by-phase view as proposed by Chomsky. The central claim revolves around the technical assumption that has been built-in here, which is that what is in the workspace is a piece of the output structure itself, matched up with formal properties which allow the establishment of ordering properties and syntactically significant relationships over such structures.

The general outlook avoids some questions we might ask of the informally presented notions (in both Uriagereka's and Chomsky's work) of the syntax "handing-over" or "transferring"/"shunting" structure to other systems in a piecemeal fashion. That is: what ensures pre-/post-spell-out *coherence*? How are the independently "handed-over" chunks related in these other systems? Do they need to respect the ordering properties established in the workspace? Or not?

Note that there are several ways that workspace ordering might be "respected". For example, we could think of the mapping between individual nodes in the workspace to the output structure as being *mirror theoretic* for (e.g.) the π-vocabulary (as in the seminal work of Mark Baker 1985, and as adopted in Brody 1999, 2003). I won't, however, be pursuing this particular point in this work (though I think its the right one to pursue given the overall architecture), it is the *general* point about pre-/post-spell-out coherence that I wish to stress here. The WS/O-distinction as we have conceived of it here provides a straightforward model regarding pre/post-spell-out (there should be a conservation of ordering properties — syntactic ordering should continue to constrain post-syntactic relationships). Nothing of course *rules out* the possibility of post-syntactic operations that would deform structure in ways that would result in loss of information

(so relationships across the interface wouldn't be trivially/transparently reversible). The present point is not that any of these things are impossible, but rather just that the WS/O-distinction provides a format within which to frame the issues.

Now asking questions about what spells out (and when) in the derivation is rephrased in terms of asking how/why/when the derivational workspace contracts.

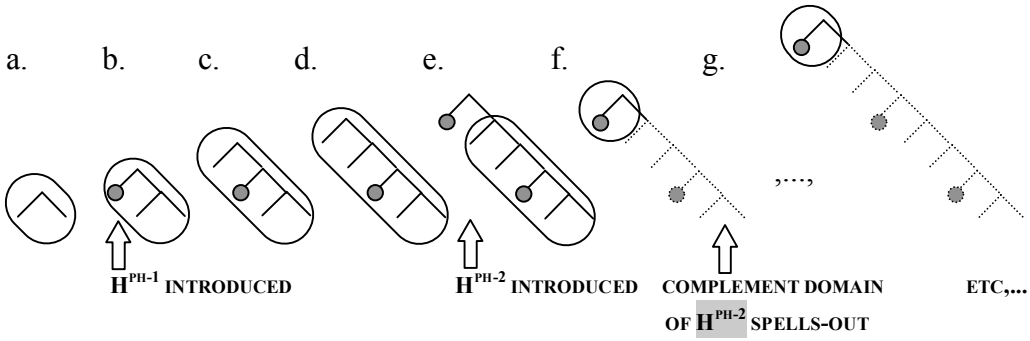
Thus, returning now to Chomsky's phase-based conception: Why should the workspace contract to exclude the complement domain of Chomsky's putative phase-inducing heads? Why not the entire sub-structure the phase-inducing head projects? Or, with our WS/O-distinction in play, why not retain the complement structure and simply spell-out the edge of the phase (i.e., the head and its external dependents)?

Note that on Chomsky's view of phases the domains circumscribed by the borders of the workspace *overlap* from some steps in the derivation to others. When the second phase-inducing head (H^{PH-2}) is introduced the first such phase-head and its external dependents are still in the workspace even though the complement domain is understood to spell-out (= "voided from the workspace").

So one important point illustrated by the discussion so far is that having successive stages of derivation in which some elements survive in the workspace despite further expansions of structure yields a kind of overlap (illustrated below in (61)a). This overlap is crucial to elaborating the notion of "escape hatches" for cross-domain dependencies. That is, escape hatches are possible because some position(s) constitute the top of certain workspaces that survive at the bottom of later workspaces — this allows an element moving to positions residing in such domain overlaps to be visible to elements higher in the structure.

There is, however, nothing inevitable about such possible overlaps between domains. We could, for example, imagine an architecture which would not permit them. In such a system, when a second phase-inducing head is introduced we could take this to signal the start of a whole new workspace. Consider:

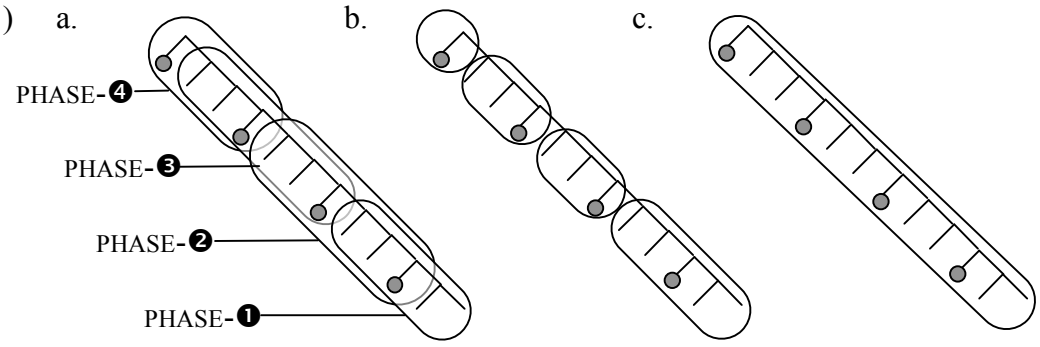
(60) a. b. c. d. e. f. g.



On such a view the step between (60)e and (60)f would result in the establishment of a brand-new workspace signaled by the introduction of a new phase-inducing head. This would yield a theory with non-overlapping phases.

Successive snapshots the derivational workspace for a longer stretch of derivation can be seen to yield maximal stretches of workspace structure (maximally expanded workspaces prior to any particular contraction steps or "spell-outs") as follows for Chomsky's view ((61)a) versus our hypothetical non-overlapping system ((61)b). Also, we include here for consideration a third possible state-of-affairs which imposes no restrictions on the workspace whatsoever ((61)c):

(61) a. b. c.



PHASE OVERLAPS **NO PHASE OVERLAP** **NO PHASES**

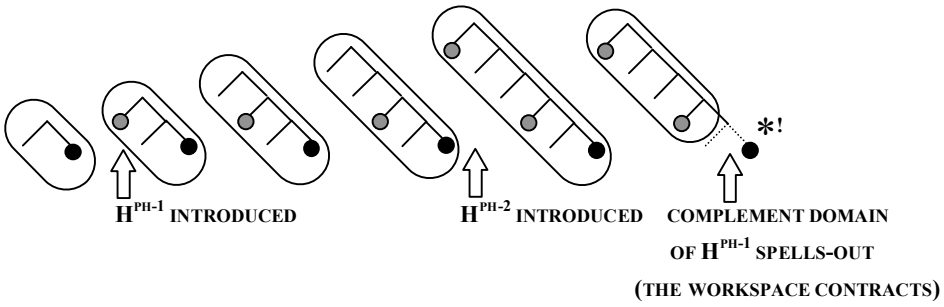
On the (61)a-view we expect the possibility of limited kinds of interactions between elements in a structure if we understand the operations responsible for connecting elements in substantive dependency relationships to be limited to what is present in the workspace at a given stage of derivation. On the (61)b-view we expect *no such interactions*. On the (61)c-view, if nothing more is said by way of constraining operations, interactions of all sorts are expected.

The (61)c-view would simply identify the workspace and output for all steps of derivation and would thus need to appeal to something other than the dynamic sort of domain-demarcation under discussion here to understand locality. For example, operations might be limited to only being able to relate two elements α and β in a structure if there is no intervening element δ that could enter into the same type of relation (e.g., Relativized Minimality; Rizzi 1990).

Its easy to see that having a workspace which limits the reach of dependency-forming operations in the grammar could be redundant with distance restrictions of the minimality sort that are often appealed to in the literature. (It may perhaps already be obvious given the heavy emphasis I have placed on the notion of workspace which direction I will suggest we go in removing any such redundancy).

Now let us consider a derivation involving successive cyclic movement in these terms. Take the darkly shaded marked node in the following to be an element bearing our {*F} property that requires licensing not available in its initial position, and first consider what happens if there is no SCM (as above, the grey-shaded nodes are the phase-inducing heads):

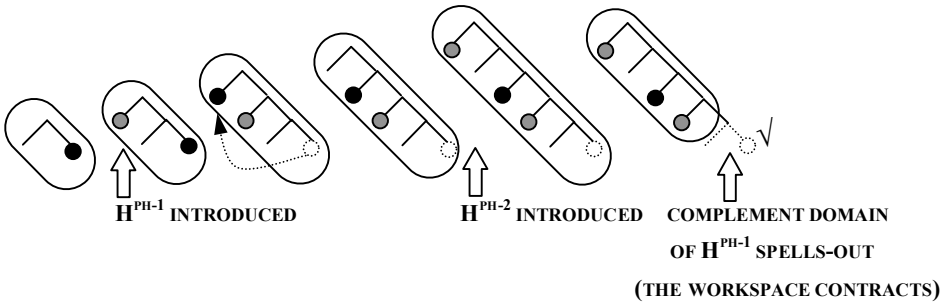
(62)



On this view then, if the {*F}-bearing element does not move from the complement domain of the first phase-inducing head (H^{PH-1}) it will be stranded in the abandoned (spelled-out) portion of (the output) structure. On the assumption that such unlicensed properties are uninterpretable by or illegible to the interface systems, this derivation would crash.

However, the structure of this account makes available the possibility of the {*F}-bearing moving to some position outside the complement domain of H^{PH-1} prior to the spell-out of that domain and thus managing to stay within the workspace (remaining active/visible for later steps of derivation). To illustrate:

(63)



As such a derivation continues, introduction of further phase-inducing heads would thus drive further contractions of the workspace (spell-outs) and would thus require additional local movements of the {*F}-bearing element until it reached a domain within which it could associate with an appropriate licensor.

There are some interconnected technical matters that require attention in such an approach. First, what happens to the problematic {*F} feature of the lower element (trace/copy)? Second, what motivates the movement out of the relevant complement domain?

Regarding the first point, if we regard the displacement operation as "copying" then why is the {*F}-feature not problematic for the lower copy when its containing domain spells-out (falls outside the workspace)? If it is *not* a copying operation — and instead involves a literal *re*-merger resulting in a multi-motherhood structure, then why doesn't the same worry hold (i.e., the {*F} property should still reside in both structural contexts)? On either the copy or remerge view it seems that we do not have a way to avoid the outcome that held in the *non-movement* situation if sub-parts of structure are undergo cyclic evaluation of the sort just sketched. We might suggest that the copy left by movement can have its {*F} property freely deleted, but then why couldn't this happen in the non-movement case? The answer to this question might be taken to involve appeal

to some later stage of derivation where a matching {F}-property would go unchecked, but its not clear that this response would be correct (e.g., what about *wh*-in situ?).

Regarding the licensing of the intermediate movement, there are two obvious possibilities. First, there might be some property of the intermediate landing-site that serves to license the movement but crucially *not* to license the {*F} property (i.e., it must "move-on" to some other position to license this property). Second, we might motivate the movement as not being driven by the landing-site properties, but rather by some combination of the local context and the {*F}-property itself. Movement out of locally evaluated domains might be possible just in case failure to do so would result in a crashed derivation at that point.[35] This relates to the question above regarding the properties of the *wh*-element itself, and how/why it does not crash the derivation even if it *does* move (in virtue of leaving behind a copy) and what to say about situations where we *do not* want the element to move (e.g., in *wh*-in situ cases). Again, I will return to these issues in later discussion. Before we move on to consider how TAG derivations work in comparison, let us sum up some key questions raised in this section. I will borrow from a discussion in Falser (2004) and refer to these as the questions of TRIGGERING (64)a and CONVERGENCE (64)b:

(64) Assuming SCM exists,...

a. What motivates it? Properties of the moving element? Properties of the intermediate target? Both? Neither? (i.e. something *else*)? (**TRIGGERING**?)

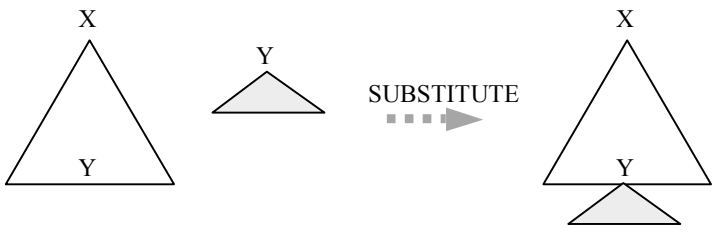
b. What are the mechanics of movement like such that unlicensed features {*F} do not remain to cause convergence problems in spelled-out domains (either on the copy, or on the remerge view)? (**CONVERGENCE**?)

[35] This basic line is suggested in Lasnik & Uriagereka (*forthcoming*).

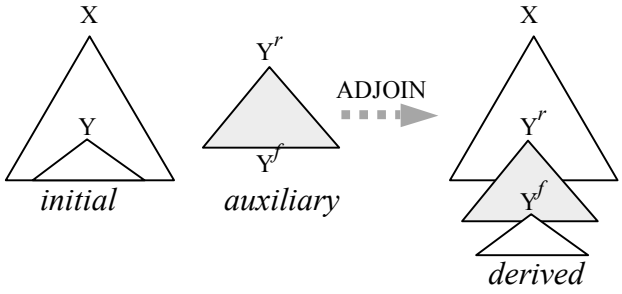
1.4.3. What Goes for Cyclicity in TAG

The categorial distinctness condition on the syntactic workspace introduced above (see (9)) relates to ideas from work in Tree Adjoining Grammar (TAG), though the insights will be implemented rather differently here. The key TAG idea is that we might understand the fundamental notion of recursive structure to play a central role in understanding the range of possible interactions between phases of derivation (more neutrally: between chunks of structure). As a matter of its basic architecture, TAG factors complex structures into non-recursive *elementary* trees and recursive *auxiliary* trees that are combinable via TAG's two main operations (substitution and adjoining). These two operations can be pictured as follows:

(65)



(66)

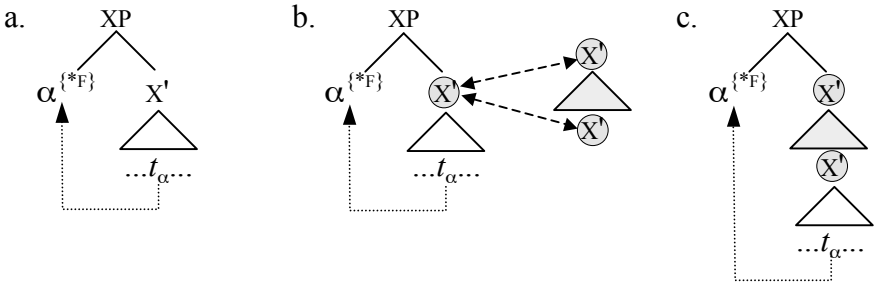


Auxiliary trees in TAG, as pictured above, are special in that they are taken to have related top (root) and bottom (foot) nodes (e.g., Y^r and Y^f in (66)) which enables the complex of relationships which they "sandwich" to be spliced-in for an some equivalent atomic element within another structured object (e.g., Y in the initial tree in (66)). Thus,

trees without this top/bottom characteristic are elementary; those with this characteristic are auxiliary. As Frank & Kroch (1995:113) put it, "the recursive character of auxiliary trees provides [...] a domination-preserving expansion of a single node in a piece of phrase structure into a larger structure.

Now consider what happens in place of successive cyclic movement in TAG. I say "in place of" because in TAG syntactic dependencies like *wh*-movement are argued to be localized to elementary trees — movement *across* such structures of the GB/MP sort illustrated above is ruled out on general architectural grounds. Thus the movement of the element α in (67)a targets what will in fact be its final landing-site, crucially within the bounds of the elementary tree.[36] This is the only movement operation that there is in this approach. This movement (/chain) relation between the base and final/target position is then stretched as a consequence of the *adjoining* operation illustrated above, which splices-in intervening material as pictured in the step from (67)b to (67)c.

(67)



[36] I will discuss later on some ideas about limitations regarding the "size" and "shape" of elementary and auxiliary tree structures, following among others the work of Frank (1992, 2002). I am leaving to the side the issue of the licensing of the feature {*F} for this illustration of TAG-mechanics. How this licensing works requires a bit more detailed and subtle discussion. This isssue is *important* however, as it turns out that the TAG approach I discuss here (i.e., from Frank's 2002 discussion) requires formulating the adjoining operation to allow "checking" across elementary trees. See Chapter 2 for some relevant discussion.

Further adjoining operations can then splice-in more auxiliary structures, yielding the effect of a long-distance relationship between α and its base trace position. Note that there are no "intermediate traces/copies" on this view.

A key aspect of TAG is the identification of the root/foot nodes of auxiliary structures with a corresponding/matching element in an elementary/initial tree in the adjoining operation. The TCG approach developed here exploits a "matching" relationship of roughly this kind as well, though such matching is understood to effect the *opposite* of TAG-theoretic adjoining as we saw in our introductory sketch (contraction).

Consider how we might view TAG derivations in the context of our WS/O-distinction. In the development of one specific TAG approach, that in Frank (2002), it is suggested that syntactic derivations are divided into two major stages. The first stage involves the merge/move mechanics familiar from Chomsky (1995). However, departing from Chomsky's "one-stage" system, Frank suggests that the merge/move portion of derivations is limited to only being able to generate structures that meet the following general condition:

(68) **CONDITION ON ELEMENTARY TREE MINIMALITY (CETM):**
The syntactic heads in an elementary tree and their projections must form an extended projection of a single lexical head

Reference to "extended projections" comes from the work of Grimshaw (1991, 2002) and others. The effect of this condition is that the merge/move portion of syntactic derivation only create objects that are roughly clause-sized or smaller. These merge/move-derived objects then, in Frank's system, are fed to a *second stage* of derivation which deploys the TAG-theoretic operations of substitution and adjoining sketched above.

Translated into our WS/O terms, we can understand Frank's CETM as a condition on syntactic workspaces, and then assume that there can be *multiple* such workspaces corresponding to the basic tree structures that form the input to the second, TAG-theoretic stage of derivation. This TAG portion of derivation is then conceived as a component which effects combination of such workspaces to form derived (output) structures (I take it that this basic picture is clear enough to not require a diagram).

Of course, this amounts to a rather different conception than the views sketched above. I mentioned the possibility earlier of having a system which would impose *no* restrictions on the syntactic workspace and thus would require that locality be stated in ways other than the dynamic view of local domains we sketched at the outset. The TAG view on this general outlook would be an entirely different approach, but one that would view syntactic workspaces as always coextensive with outputs. However, instead of introducing locality constraints on operations within the workspace, we rather have a limitation on workspace size (i.e., of the CETM sort), plus the major architectural division of derivation into (i) a first stage of local structure creation with multiple workspaces, and (ii) a second stage that handles combination of workspaces into larger complexes.

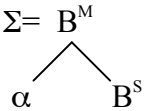
The TCG view developed here maintains a "one-stage" view in the sense of not positing two distinct stages of *syntactic* operations. I turn now to elaborate further.

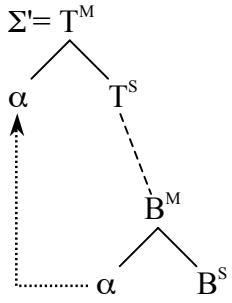
1.5. Implementation of TCG

This section backs up to consider some technical issues regarding the notion(s) of movement chains (§1.5.1). This leads us to consider, when coupled with the sketch of contraction offered above, a "reduced" view of categories and structure (§1.5.2).

1.5.1. On Relating Positions in Structure

This section discusses a possible view of movement chains based ideas from Chomsky (1995). In derivational terms we think of an item α as *first* associating with some other, independent element to form a structure Σ constituting the initial/base position B (as in (69)a). Later on, derivationally speaking, some operation causes α to enter into a second set of relationships with some target element T to form the structure Σ' constituting the derived/target position, where the T-elements dominate the B-elements (as in (69)b):

(69) a. 

b. 

Take the superscripted 'M' and 'S' in (69)a/b to stand for the mother and sister elements respectively, which together form a merge-derived structural context for α (e.g., B^S = base position sister, etc.). I will return to the issue of whether one or the other of these may suffice — or whether both are somehow required — for identifying the contexts for a moved/displaced element α, but note here that Chomsky (1995:252) understands the

relevant element for defining the contexts of α as the sister or co-constituent of α (i.e., B^S and T^S in (69)b). Consider:

> Suppose that α raises to a target [T] in Σ, so that the result of the operation is Σ' [...]. The element α now appears twice in Σ', in its initial position and in the raised [target] position. We can identify the initial position of α as the pair ⟨α, β⟩ (β the co-constituent of α in Σ [i.e., B^s in (69)a/b—JED]), and the raised position as the pair ⟨α, K⟩ (K the co-constituent of the raised term α in Σ' [i.e., T^s in (69)b—JED]). Actually, β and K would suffice; the pair is simply more perspicuous. Though α and its trace are identical, the two positions are distinct. We can take the chain CH that is the object interpreted by LF to be the pair of positions. [...] C-command relations are determined by the manner of construction of [the object in (69)b above—JED]. Chains are unambiguously determined in this way.

Following through on this view for our example in (69)b we see that there are two such relevant positions in Σ', POS_1 and POS_2, where $POS_1 = \langle \alpha, T^S \rangle$ and $POS_2 = \langle \alpha, B^S \rangle$.[37] As Chomsky puts it, these two positions are distinct, but together they constitute the chain $CH = \langle POS_1, POS_2 \rangle = \langle \langle \alpha, T^S \rangle, \langle \alpha, B^S \rangle \rangle$; or, if we adopt the "more austere version", then CH is simply $\langle T^S, B^S \rangle$ since "α and its trace are identical".

Observe however that there are situations in which the *context* positions are themselves identical. In particular, consider again the following standard cases of the putative cyclic A- and A'-movements in (70):

(70) a. John [seems [_ to be likely [_ to appear [_ to like carrots]]]]

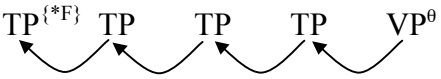
b. What [did Dave think [_ [that Mary believed [_ [that John liked _]]]]]?

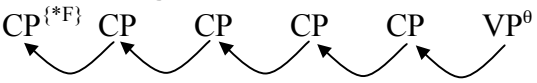
Intermediate A- and A'-movements involved in such cases are classically taken to have more-or-less the following general abstract shapes:[38]

---

[37] I am taking a shortcut here for expository purposes. The co-constituent forming the sister context for the "raised" position in (69)b would not be just "T", but rather a more complex set-theoretic object in Chomsky's general Bare Phrase Structure approach. Here we take "T" to "stand in" for this more complex object.

[38] There are, as we will discuss later, views which take movement to be "more cyclic" than this, as well as less (i.e., "one-fell-swoop" views). I will discuss these matters in the next Chapter.

(71) a. 

b. 

So while it may be true that a moving element α is "identical" within each of these contexts, at least *some* of the relevant context pairs ought to yield identity (or non-distinctness) as well. T-to-T and C-to-C movement could yield a chain CH = {⟨α, C'⟩, ⟨α, C'⟩}. Is this a problem?

The suggestion inherent in the TCG view of cyclicity sketched above can be understood as the claim that such a state-of-affairs is not only "not a problem", it is in fact crucial to understanding linked local relationships. As we saw earlier, this is at the heart of the TAG architecture as well. And as we will see briefly later on, these relationships have been argued to be central in some MP work (e.g., see Bošković 2002, Grohmann 2003; both of whom stipulate that A-movement is T-to-T and A'-movement is C-to-C).

However, it is important to note that the sketch above regarding Chomsky's view of chains and contexts overlooks a key feature of his view. For Chomsky, "contexts" are *not* simply the local label of the element that a moving "α" relates to, but rather the entire structure derived up to that point. So, there would on his view be no issue which could arise in terms of distinguishing the contexts in CH = {⟨α, C'⟩, ⟨α, C'⟩}, since the contexts would always be *unique* (they are distinguished by the differences in the structure they dominate). I will turn to this in a moment.

The suggestion here (as sketched in §1.1 & §1.2) is that the natural relationship is <u>not</u> between a moving element and these various intermediate positions which just happen to share the super-category specifications of the sought-after target landing site;

rather, the natural relationships are between the contexts themselves. That is, the natural basic relations that the "moving" element can/should be understood to enter into are the substantive "core" licensing properties (e.g., *wh*, κ/ϕ, θ, etc., what Fukui & Speas 1986 called "Kase" properties). The generalizations about the "movements" *other than these* are most elegantly and naturally stated by positing direct relationships between the contexts. For this, we need only to specify a notion of like/unlike within an architecture where such differences could matter for derivation and representation. This is the aim of TCG.

Our sketch of the TCG approach to SCM presupposed that the "contexts" which can be identified (resulting in lowering) were understood to relate to the "moving" element in terms of a local dominance relation. Note Chomsky's suggested view above states things in terms of sisterhood. Below I will show that the sisterhood view can't support what we would require of it within the TCG approach, and that we in fact require the relevant relation to be motherhood/domination.

However, before heading down that road it of some interest to probe Chomsky's discussion of movement chains a bit further to consider two important technical issues. In particular, the notation used above to mark the SCM's in (70) does not accurately capture one the versions chains discussed in Chomsky (1995), though as we will see, it seems to be demanded by the more recent work proposing that derivations work by phase (depending, as we will see, on how we view "spell-out" — our WS/O-distinction turns out to be helpful in this respect — see below).

The two important issues involve (i) how we understand what contexts are, and (ii) how contexts are tracked/connected in the course of the derivation. Specifically, Chomsky (1995:300) makes the following remarks about (72) (= his (88)):[39]

(72) We are likely [t_3 to be expected [t_2 to [t_1 build airplanes]]]

He writes:

> Here the traces are identical in constitution to *we*, but the four identical elements are distinct terms, positionally distinguished [...]. Some technical questions remain open. Thus, when we raise α (with co-constituent β) to target K, forming the chain CH = (α, *t*), and then raise α again to target L, forming the chain CH' = (α, *t'*), do we take *t'* to be the trace in the position of *t* or α of CH? In the more precise version, do we take CH' to be (⟨α, L⟩, ⟨α, K⟩) or (⟨α, L⟩, ⟨α, β⟩)? Suppose the latter, which is natural, particularly if successive-cyclic raising is necessary in order to remove all -Interpretable features of α (so that the trace in the initial position will then have all such features deleted). We therefore assume that in [(72)] the element α in t_1 raises to position t_2 to for the chain CH_1 of [(73)], then raises again to form CH_2, then again to form CH_3.

(73) a. $CH_1 = (t_2, t_1)$
b. $CH_2 = (t_3, t_1)$
c. $CH_3 = (we, t_1)$

CHAINS are ordered pairs of contexts where a particular context for a given element α is understood to be its *sister* or *co-constituent*. If we take this to mean that the "context" is the *entire structure* dominated by the sister element, then the more complete version of the relevant objects in (73) for the derivation of (72) are those in (74):

(74) a. CH_1 = (⟨*we*, [to *we* [build airplanes]]⟩, ⟨*we*, [build airplanes]⟩)
b. CH_2 = (⟨*we*, [to be expected [*we* [to *we* [build airplanes]]]]⟩, ⟨*we*, [build airplanes]⟩)
c. CH_3 = (⟨*we*, [are likely [*we* [to be expected [*we* [to *we* [build airplanes]]]]]]⟩, ⟨*we*, [build airplanes]⟩)

[39] Chomsky's example in the discussion referred to in the text used the token "*we are likely to be asked to build airplanes*". I have switched out *ask* for *expect* in my discussion here. It seems clear from the context that Chomsky intended to have an passivized ECM verb in this example, as pointed out to me by Howard Lasnik.

Below I will suggest that we adopt this idea regarding contexts, but reject the view of contexts as understood as the entire structure up to the relevant step of derivation.

The "technical questions" Chomsky raises in the quoted passage above amount to the choice between the following two options regarding how contexts are connected in the course of derivation. From the derivational stage depicted in (75), we can consider the result of the next movement of *we* to be (76)a or (76)b:

(75) ...[to be expected [*we* [to [*we* [build airplanes]]]]

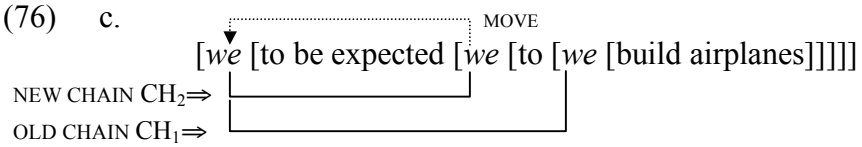
(76) a. [*we* [to be expected [*we* [to [*we* [build airplanes]]]]] MOVE

b. [*we* [to be expected [*we* [to [*we* [build airplanes]]]]] MOVE

It is interesting to note that if the technical option Chomsky pursues ((76)a) is correct, that we appear to have cases for which the concepts of MOVE and CHAIN would be dissociable — compare (76)b where the any notation for the resultant chains would transparently recapitulate the derivational history of movement (see also (76)b' below for illustration). These two options turn out to not be equally compatible (at least not equally straightforwardly compatible) with derivation by phase.

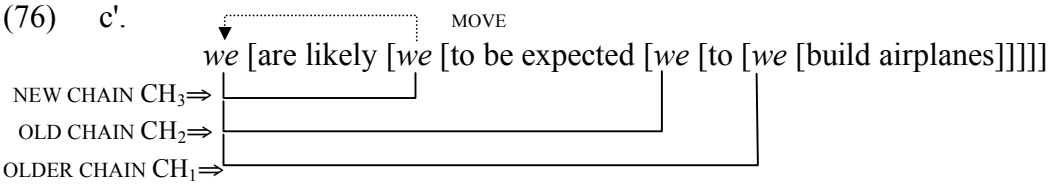
Before turning to this point about chains and phases, note there is at least one other technical option which Chomsky does not consider. This third option would regard movement as extending the *initial* (derivationally prior or "older") chain and forming a new one as in (76)c (in contrast Chomsky's version creates a new base-position-tailed chain, leaving the older one intact):

(76) c.



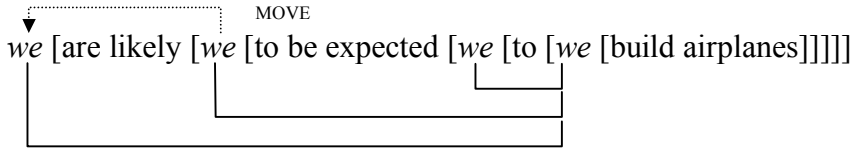
The next step on this alternative view would involve the formation of a third chain (CH_3) and a kind of stretching of the previous two chains, as in (76)c':

(76) c'.



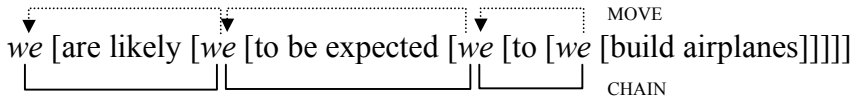
Contrast this with the next step for (76)a (Chomsky's approach) in (76)a':

(76) a'.



Note that on both Chomsky's alternative technical view ((76)a/a') and the alternative ((76)c/c') we can understand move and chain as dissociable to some extent, compared to (76)b where the relevant movements and chains are essentially the same, as mentioned above; consider (76)b' in this regard:

(76) b'.



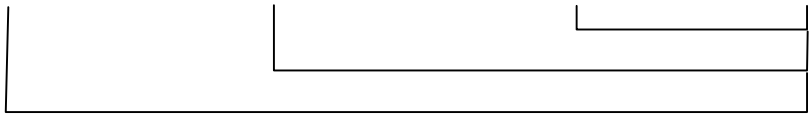
What is the difference between the choices in (76)a-c? One obvious point is that only (76)b/b' seems straightforwardly compatible with the notion of spell-out by phase of Chomsky (1999) (this is the notation used informally above to illustrate our basic successive-cyclic A- and A'-movements).

Phase theory, as introduced above, is a recent version of the general idea of cyclic domains for rule application. In Chomsky's recent work the suggestion is that CP and *v*P (and possibly others) constitute special domains which, upon derivational completion, require that their complement domains be shunted/transferred to the interpretative systems for evaluation. Above we suggested a way of viewing these transfer steps of derivation within our workspace/output structure distinction. Consider however the impact that viewing spell-out/transfer as a literal "handing-over" or "removal" of sub-parts of structure has on our discussion of chains in Chomsky's terms.

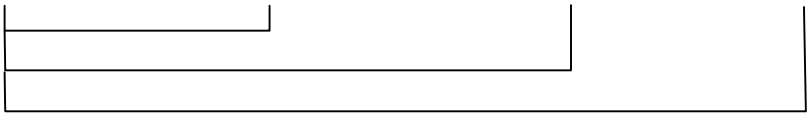
We can make the point with reference to the case of successive cyclic *wh*-movement. For this example, we will consider only C^0 as constituting relevant phase-inducing category, as the point regarding the formal shape of chains remains the same even if we consider additional narrower domains for spell out (like *v*/VP).

The relevant structures and chains would look like (77) for successive cyclic *wh*-movement on Chomsky's view, and like (78) on the alternative discussed above:

(77) What [did Dave think [_ [that Mary believed [_ [that John liked _]]]]]?



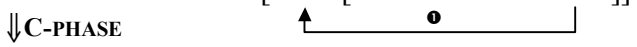
(78) What [did Dave think [_ [that Mary believed [_ [that John liked _]]]]]?



But, suppose that we understand phases (here: CP*s*) as spelling-out their complement domains upon reaching the next higher phase-inducing head, as Chomsky (1999) proposes. Technically, the proposal embodied in Chomsky's PHASE IMPENETRABILITY

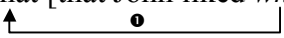
CONDITION (PIC) has it that when the next highest phase-inducing head is reached, all of the substructure constituted by the previous phase-head's *complement domain* spells-out, leaving a residue (roughly equivalent to the "checking domain" of Chomsky's earlier proposals, see Chomsky 1993). This means that the first movement of the *wh*-element in our example will have its "head" visible and will therefore be able to be moved to the next CP, as this element will occupy the "edge" of the previous phase. But what happens to the initial chain (❶) when the substructure containing its "tail" spells-out?

(79) a. [what [that John liked *what*]]



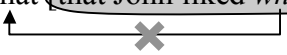
⇓C-PHASE

b. [that Mary believed [what [that John liked *what*]]



⇓C-PHASE SPELLOUT⇒⇒⇒⇒⇒∅

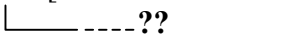
c. [that Mary believed [what [that John liked *what*]]



BROKEN CHAIN???

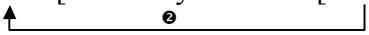
⇓C-PHASE

d. [that Mary believed [what [.....∅.....]]



⇓C-PHASE

e. what [that Mary believed [*what* [.....∅.....]]



f. *etc.,...* (by phase)

The spell-out by phase idea, regardless of the grain or size of structures considered to constitute such phase-domains, is not straightforwardly compatible with the idea of having the kind of view of chains in (77), nor the variant introduced above in (78).[40] The

[40] It could be that Chains are "real", and work as suggested in Chomsky (1995) (as discussed above), but that they are fundamentally *not* "syntactic objects". Maintaining the reality of "chains" in a derivation-by-phase architecture appears to require the postulation of a kind of cross-dimensional object that exists across sub-stretches of syntactic computation and the interpretative components or that chains are fundamentally objects of the interpretative system(s) which the syntax in some sense creates but cannot itself handle/manipulate (the system only sees particular elements "α" which can merge and remerge).

problem with both of these conceptions is that they involve the postulation of a *syntactic* relationship which is maintained to the base position, but on the derivation by phase view these lower positions are understood to be in some sense "absent" at the relevant later stages of derivation in virtue of the spell-out operation.

Even the idea of having a composed or linked chain appears to not make any sense on the derivation by phase view, as there is no stage of derivation over which we could describe such objects. We appear to either need chains to be non-syntactic entities — e.g., objects of the interface system (maybe plausible) — or we need to regard chains as objects somehow superimposed over a dynamic derivational history (i.e., still "syntactic" but "higher order").[41]

Note as well that it is not entirely clear how to maintain chains as *syntactic* objects on a phase-based view that adopts as well the view of contexts as the entire structure of derivation up to the relevant point (i.e., everything dominated by the sister/context for α). The view of contexts as the entire structure to which α relates seems to require that we can refer to such structures — but if portions of such contexts are dynamically shunted/transferred by phase, its not obvious how this should work. At best, contexts could be defined down to the previous phase-inducing head, and not below.

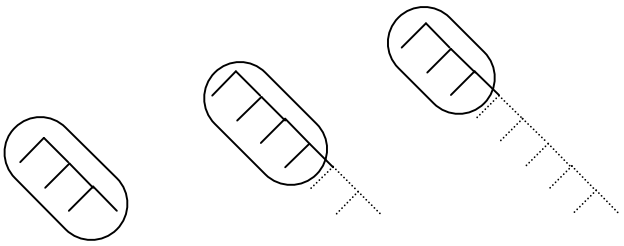
These are interesting consequences it seems to me. Put another way, suppose we take the conditions we want to hold of chains to be syntactic conditions. If we can't refer to chains themselves (since there are no structural contexts over which we can capture the relevant relationships in the multiple spell out view), this means, for example, that

[41] See Uriagereka (1998) for a discussion of such a view.

whatever properties are associated with the *wh*-element in virtue of having entered into the complement (θ) position of the embedded verb *like* in our example above, these must be somehow maintained as properties of the element itself (e.g., θ-marking of and element α could be understood as α receiving or being marked somehow with a θ-feature — see Hornstein 2000 for an extensive development of this approach).

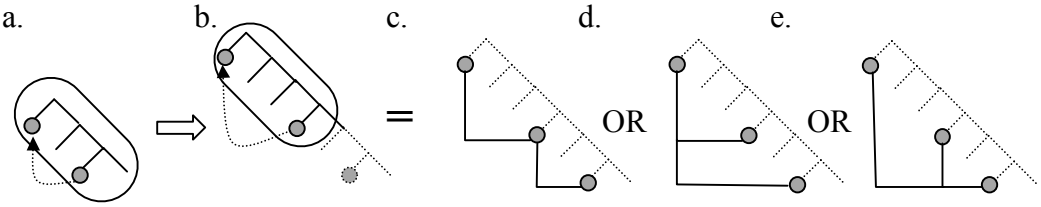
However, note that the workspace/output distinction as we have introduced it sidesteps these issues. I mentioned above that we might deploy our view to avoid issues that might arise regarding pre-/post-spell-out coherence. This is exactly such a situation. Consider again our earlier schema (i.e., the bottom-up version), repeated here in (80):

(80)



However it is that we might choose to view CHAINS, this schema allows us to straightforwardly maintain the overall coherence of the derivation in virtue of maintaining the output structure in the way pictured above. Thus the initial and subsequent movement pictured in (81)a/b could be seen to yield any of the objects in (81)c-e, depending on how we sort out Chomsky's two technical options ((81)c and (81)e) or our additional one ((81)d):

(81) a. b. c. d. e.



So we have a format available for considering all of the possibilities discussed above regarding how contexts are tracked/connected throughout the course of a derivation. Moreover, this view is neutral as it stands on the question of whether we define contexts as just the *sister*-label, or in terms of the entire structure dominated by the sister/co-constituent. I will return to this issue in a moment.

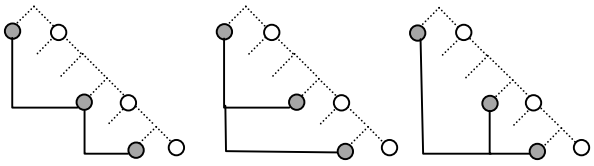
This underscores again the generality of the workspace/output structure distinction and the new idea it brings to discussions of MSO-systems. It helps us here because it points to a way of conceiving of spell-out which does not involve a literal "handing-over" of structure from the syntax to the interface systems, as spell-out is sometimes characterized informally. Or, rather, the WS/O-distinction offers a concrete formulation of the content of "handing-over"/"transfer" under which the technical questions raised above do not arise. So whatever relations we establish as part of the syntax can still be "there", but simply not within the active stretch of syntactic computation. This makes it possible to conceive of chains in any of the ways pictured above, with potential stages of derivation that might have workspaces in which only *parts* of a given chain might be visible. This is another instance of the WS/O-distinction providing a way to understand pre/post-spell-out coherence.[42]

Let us now put this discussion back together with Chomsky's idea that chains are fundamentally connections between contexts. Take the unshaded nodes below to be the

[42] McGinnis (2004:64fn18) raises the issue of how c-command is supposed to be understood as holding across phases. For example if c-command is understood as derivational in the Epstein *et al* (1998) sense, its not obvious that when α and β merge they come to c-command everything each other dominates if some of the derivationally *previous* domination relations are literally no longer "there" in the narrow syntactic computation. Her exposition presupposes the intuitive notion of "handing-over"/"transfer", which is why her raising this question makes sense. Again, these issues do not arise given our WS/O-distinction.

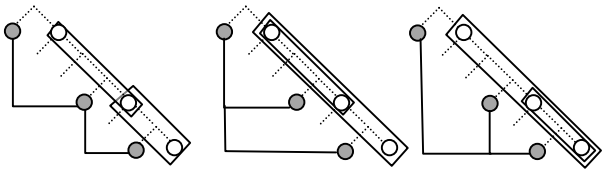
sister or co-constituent elements defining the contexts (the "chain") for the moved element α (occurrence of α represented by the shaded nodes).

(82)



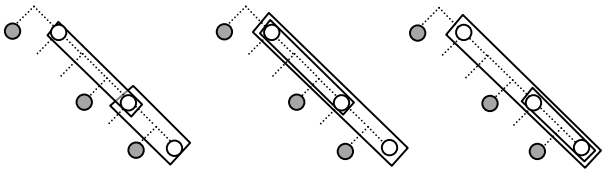
Viewing these context elements as independently relating along the dominance sequence — a relationship that is "there" in any event, whether we view at as manifesting a "dependency" relationship or not —yields the following three possibilities in (83) corresponding to those in (82):

(83)



We can now dispense with the extra-structural arcs yielding:

(84)

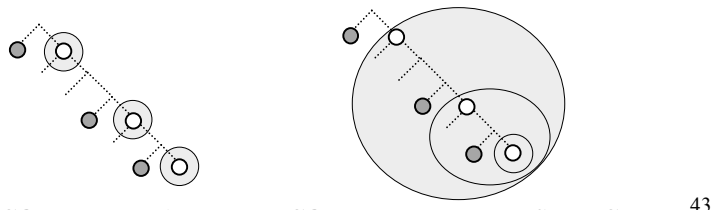


If the relationships between context elements are in some sense independent of the element(s) that relate to them (i.e., that "move through" the positions they in part define), then we might consider the possibility that a given element α might relate only once to such chain structures, perhaps targeting different *parts* of such complexes.

We might also entertain a different conception of contexts, in two senses. First, as pointed out by Chomsky (1999) and Lasnik (2000), there are two possible relations on a merge-based view that might be the relevant for implementing the context-view of chains. So far we have considered only sisterhood/co-constituency, but there is also motherhood or immediate domination/containment. We will require this latter conception (see below).

Second, we have the following possible difference between two ways regarding how to understand what contexts actually are, which is independent of the sisterhood/motherhood distinction:

(85)



CONTEXT=LABEL **CONTEXT=ENTIRE STRUCTURE**[43]

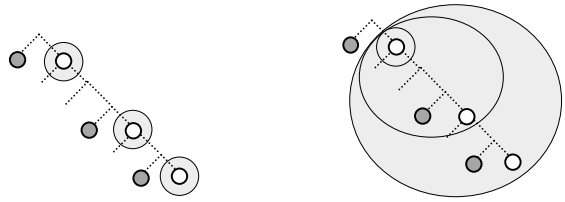
On the righthand side we have a picture of contexts as suggested in Chomsky's example (see (74) above). This is a view cast in hiearchical terms that is similar in spirit to Chomsky's (1955) definitions of contexts in terms of strings (where occurrences of an element α are uniquely defined by the left-to-right content of a string up to a given occurrence — picking up on a notion present in Quine (1960) for formalizing variable occurrences in logic). On this view, we cannot exploit the possibility of having the system be "unable to distinguish" between contexts, since contexts are derivationally

[43] Howard Lasnik (p.c.) points out that this distinction is technically between two different ways of conceiving of labels — either as local head/phrase information or as encoding the "entire derivational history" up to the relevant point where α is integrated (whether on "first" or some subsequent *re*-merge). I will retain the notion of label for the local category/feature information view, using the notion of the "entire structure/derivational-history" to refer to Chomsky's (1995) view.

unique. On the lefthand side, however, we have a descriptively less powerful view which identifies contexts by just the local label of α's sister (or, perhap instead: α's mother).

These two views shake out somewhat differently if derivations work top-down. For example, if α's is initially integrated in the top-most position, and particular points of derivation are what is relevant for identifying contexts as in Chomsky's view, then α's initial context will be just the sister or mother node characterizing this initial position. The natural extension of the idea of contexts as the entire derivation up to the relevant point where α is integrated (or remerged) to the top-down view would then see intermediate positions as identified by all of the structure that dominates them. Howevere, the local-label view remains the same on a top-down view. That is:

(86)



CONTEXT=LABEL **CONTEXT=ENTIRE STRUCTURE**

"TOP DOWN"

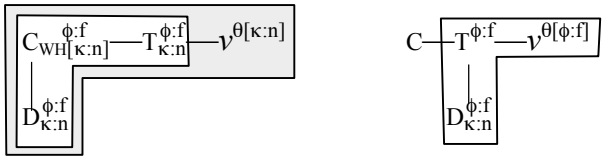
It is the weaker notion of contexts — viewing them as simply the local label, and not the entire structure up to the relevant point of derivation — that our view of SCM requires. This could perhaps be motivated on minimalist grounds appealing to simplicity and locality — the local label view does not require that we keep track of arbitrary stretches of derivation in order to keep track of occurrences of a given element α. However, the suggestion here is actually a bit stronger than this. That is, rejecting the descriptive power inherent in the "entire structure" view of contexts yields a system that is weaker in precisely the way that we require to understand SCM. It is, in fact, another way of stating

the key idea being developed here to say that it is because contexts are narrowly/locally defined that situations can arise where they are not unique, and it is *this* state-of-affairs that underwrites SCM type relationships (that is, situations in which contexts in adjacent domains cannot be uniquely identified).

So, to sum, we adopt a context-based view of movement chains, but limit the defining contexts to just the immediate local relationships (here: dominance relations realizing feature-licensing connections).

In addition, we might consider the possibility that our three technical options regarding how contexts are connected to each other in the course of the derivation are actually not in fact technical/theoretical options for characterizing a single sort of relationship, but rather three different species of chains — different *constituency structures* of chains if you like (see Uriagereka 1998:399 for a related discussion). Recall from our discussion of some schematic TCG derivations for A'- and A-relationships in §1.2 that we pointed to a "grouping" defined by stretches of agreeing properties on the dominance ordering, particular stretches of the path with shared ϕ and/or κ values. In particular, we pointed to the following:

(87)

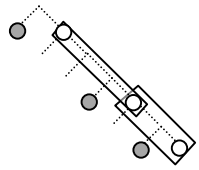


A'-RELATION A-RELATION

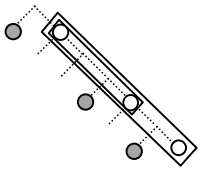
In Chapter Three I will suggest that these ideas about co-valued properties along the dominance path can in fact be helpfully viewed as a kind of "chain" constituency. Note that the verbal projection path in the A'-relation in (87) manifests a grouping of the sort

seen in the middle schema in (84), repeated here in (88) with some possible descriptive labels suggesting a typology of relationships that I will return to.

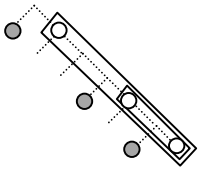
(88)



Binding/Control
(connections between
A-relations)



A'-relations



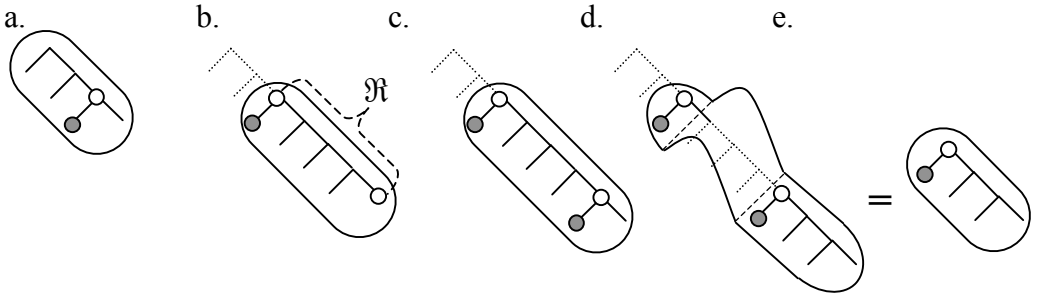
Some A-relations

It will be beyond the scope of the present work to pursue these divisions (and the possibility of others perhaps) in great detail, but in the course of developing some analyses I will again return to these schemas to point out some of the patterns which emerge on the specific implementation of the general TCG view being proposed in this work (see our concluding discussion in Chapter Three).

Let us return now to some technical possibilities regarding the node-identification process that I suggested might be useful in understanding SCM. This will lead us to a discussion regarding labels and structure and a particular set of assumptions regarding these concepts that I will be adopting here.

Recall the following schema from above (repeated here as (89)) in which we suggested that the structural context for a "moving" element might be understood (under some relevant matching relation $\Re$) to collapse/become-identified-with some lower like element. The suggestion was that such identification results in the equivalent of lowering.

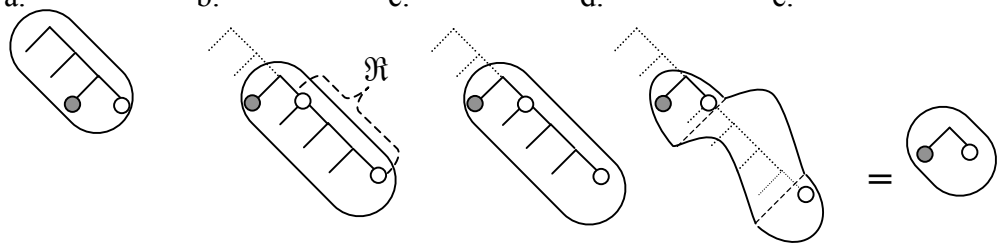
(89) a. b. c. d. e.



MATCHING RELATION ℜ IDENTIFICATION CONTRACTION

Although we did not mention this earlier (as we had not yet discussed the notion of chains and contexts), this view demands that the relevant contexts for α be understood in terms of motherhood. Note what happens if we view the relevant matching to occur with the *sister* element of the shaded node as in (90):

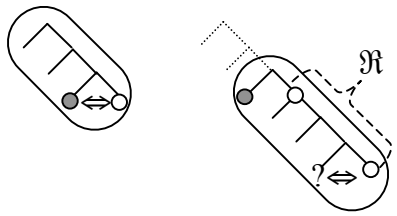
(90) a. b. c. d. e.



MATCHING RELATION ℜ IDENTIFICATION CONTRACTION

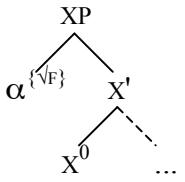
The matching of the open nodes would either result in no equivalent of "copying", so that the structure would simply reduce as pictured in (90)d/e, *or* we would have to entertain the idea that the open/unshaded and shaded nodes in (90) can instantiate a sisterhood relation which is independent of any dominance relationship, allowing us to extend the logic we introduced above regarding dominance to effect a similar "lowering". But its not clear how this latter view would work. To see what I mean by needing an independent sisterhood relation, consider (91):

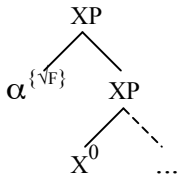
(91) a. b.



In order for the lower open node to have been introduced, as shown in (91)b, it *must already have a sister*. So its not clear how the sisterhood relation could support anything like the "lowering" operation we have been considering as a possible basis for approaching SCM phenomena. The picture in (90) above still remains a possibility that would be of interest (recall this is the same as the picture we initially offered to introduce the notion of contraction; see example (6)), but this would yield nothing like a "movement" relation, as it would not cause the shaded node to enter into any new dominance relationships in the output (or in the workspace). Below we expand on the reduced structural descriptions that were appealed to in our sketch of SCM analyses earlier on — these structures essentially deny the existence of linguistically significant "sisterhood" relationships. In addition to the technical problems for sisterhood just raised, our independent assumptions about structure will thus be seen to rule-out the very possibility of identifying contexts in this manner (only the motherhood/dominance-type relations will be available in principle).

Let us consider some possible ways of viewing the upper context of α which depend on different conceptions of phrase-internal projection distinctions, about which we have so far said nothing. Here are two familiar ones:

(92) a. 

b. 

In (92)a we have a traditional view positing three different phrase-internal projection types, the head (X^0), the intermediate (non-minimal/non-maximal) X', and the maximal XP. In (92)b we have the view that specifiers are in fact adjunction structures (in the formal sense of May 1985, 1991; Chomsky 1986, as proposed e.g., in Kayne 1994). We can note right away that the specifiers-as-adjunctions view will be difficult to render consistent with the intuition behind the matching relation ℜ as we have so far been hinting at it — that is: the general idea of understanding the regulation of the size of the active workspace in terms of recursion (repeats of like elements).

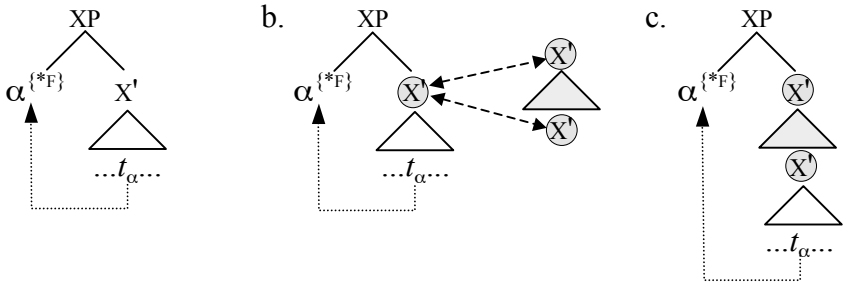
The general idea of TCG as we have been developing it is that the syntactic workspace cannot tolerate multiple tokens of a given type X, and that because of this limitation situations arise in which the workspace might *either* contract to remove one of the offending like elements from the workspace, *or* it might simply be unable to distinguish the two resulting in the sort of collapse/identification sketched informally above.

On this intuition — that recursion in structure matters for regulating the maximal expansions of the syntactic workspace — its unclear how there could be categories divided into segments of the adjunction sort. There may be technical ways of working

with such structures to implement the TCG intuition regarding a workspace limited to representing single tokens of given types, but I will not pursue this possibility here.[44]

Note that the traditional view involving intermediate-level categories in (92)a above is appealed to in TAG-theoretic derivations. Recall from above the general schema for the TAG equivalent of non-local movement relations:

(93) a. b. c.



Can we appeal directly to this idea such that TCG would simply involve the inverse of TAG-adjoining to shrink/contract structures?

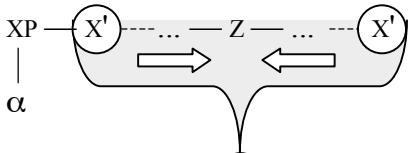
The answer, I will argue, is "no". Viewing the node-identification and contraction of structure as a kind of *anti-adjoining* will fail to generate the structures that I will argue are needed to understand the interaction between cyclic movement and certain binding-theoretic phenomena. Simply removing or splicing-out intervening structure defined by a top- and a bottom-node of the X'-type will not result in the kind of lowering that would result in α being dominated by material below its upper occurrence, though it does succeed in creating new local domains over which α will dominate. This is what I showed above in examples (90) & (91).

[44] For example, we might find technical justification for distinguishing the segments of adjunction structures based on feature values, an idea that I make use of in a different way in later discussion.

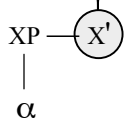
But the former (getting α to be dominated by its previously 'neighboring' initial dominance domain) is what I will argue to be required to correctly handle the binding facts (discussed below and returned to in more detail in Chapters 2 & 3).

To quickly re-illustrate the point, now in specific connection with TAG: having a anti-adjoining procedure (just reversing "direction" of standard TAG steps of derivation) could yield splicing-out of the sort in (94):

(94) a.



b.



But this sort of operation could not result in α being dominated by the intervening element Z in the output structure. Cases where this matters are illustrated with the following examples:

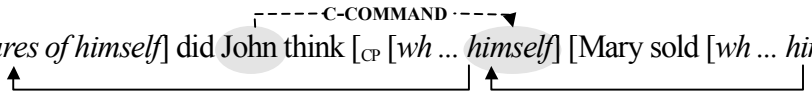
(95) a. John thought pictures of himself/*herself were on sale
b. Which pictures of himself/*herself did John think were on sale

(96) a. ??John thought Mary sold pictures of himself
b. Which pictures of himself did John think Mary sold

The *self*-form within an NP in the embedded subject position (95)a can (and *must* as the agreement mismatch shows) be bound by the matrix subject. This *self*-form can precede the matrix NP if it is within a fronted *wh*-phrase without loss of acceptability. This could be understood by connecting the analysis for (95)b to that of (95)a by positing a copy of the *wh*-phrase in the base position (or a trace that can be reconstructed into in some way).

However, note that if the phrase containing the *self*-form is in the object position, the embedded subject must bind it, and nothing higher can (96)a. But on given this observation we cannot extend the (96)a analysis to (96)b via positing a trace/copy in the object position of the embedded verb, since we've just seen that binding by the matrix subject is not possible with the *self*-form in that position. But if there is an intermediate movement, for example to the top edge of the embedded clause, nothing intervenes between the *self*-form and matrix subject thus opening the possibility of keeping the view of binding constant across these examples. That is, (96)b could be seen to involve the following partial representation as follows:

(97) [*Which pictures of himself*] did John think [$_{CP}$ [*wh* ... *himself*] [Mary sold [*wh* ... *himself*]]]



Crucially, in order for this line of analysis to work, the relevant intermediate copy/trace has to be c-commanded by the matrix subject, as pictured above. But we've just seen above that this is not what the TAG derivation provides. In our schema (94) above what we need is for the XP to somehow end up under intervening elements like Z — but this is not what we get either with TAG's adjoining or with a possible inverse operation that would otherwise be consistent with the views being developed here (e.g., contraction of X' elements).

In addition to these technical/conceptual and empirical concerns there is also the following worry, which is similar to the concern raised above for adjunction-type structures. Do we have reason to think that XP and X' are distinct such that X' would not interfere with the XP-XP relationship that we seem to need to support context-identification and the consequence "lowering" of elements? If XP and X' are distinct, and

the relevant matching/identification could involve XP, then things could work for SCM as I sketched them above. But this would require that we refer directly to the equivalent of "bar-level" specifications to get the technical details of this view of movement off the ground. Another possibility is that X' elements are (i) there in the structure, (ii) *non-distinct* from XP, but (iii) for some reason they are "invisible". I return to this issue below, but note here that I will not be pursuing this line of thinking. I will instead be adopting a view here based on a different conception of categories and structure which does not admit the possibility of an X'/XP distinction in the first place. This view allows us to sidestep a number of these technical issues and problems, and abstract away from other issues that will not be a of central interest.

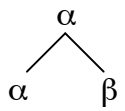
The view I will be working with is drawn from one of a few interesting recent minimalist investigations aiming to reduce the available range of distinctions in the theory of phrase-structure that analysis can appeal to. It is arguably more consistent than the salient alternatives with the general intuition about the workspace not tolerating "like elements", as we will see. I turn to these matters directly.

1.5.2. Labels & Structure

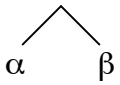
Let us consider two rather different ways of simplifying structural descriptions with respect to structure and category that have been suggested in the recent literature. First, Collins (2001) has suggested that we might head towards a theory in which label distinctions are eliminated as marks on derived structure, retaining this information as a designation only for the ultimate parts of structures (i.e., the terminal elements). So instead of the sort of object from Chomsky's (1994) Bare Phrase Structure (BPS) in (98)a,

where the underlined occurrence of the symbol 'α' is taken to be the *label* of the merge-derived complex {α, β}, we have rather (98)b, which encodes this label information only for the ultimate parts of the structure:

(98) a. {<u>α</u>,{α, β}} OR,...



b. {α, β} OR,...

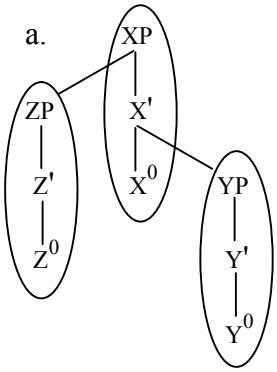


Collins approach is quite interesting, but it will not support the key idea I am aiming to develop here regarding dominance-encoding of chain-information and the node-identification procedure that I suggest as relevant for successive cyclic movement. This is so because Collins' system does away entirely with the relevant label markings on derived structure, so the system does not make available the formal means to express the general idea underlying TCG.[45]

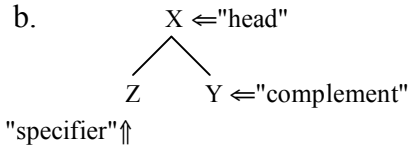
However, others have pursued somewhat similar attempts at reducing the distinctions available for principles to refer to in the pursuit of eliminating redundancies and (perhaps) thus increasing restrictiveness. For example, other such label-reducing/-eliminating kinds of moves have been suggested as follows:

[45] This may be hasty, but at present I do not see a clear way to begin articulating the TCG system I am developing here within Collins' assumptions.

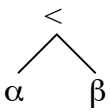
(99) a.



b.



(100)



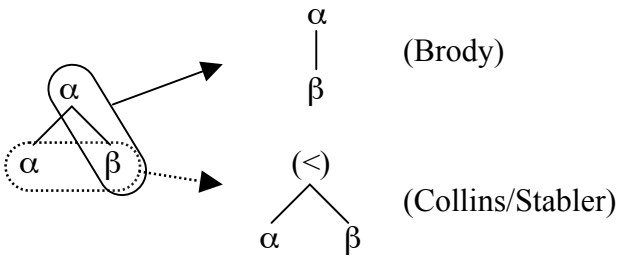
Brody (2000, 2003) suggests a reduction in the available distinctions for capturing phrase-structure generalizations, eliminating the structural distinctions in (99)a in favor of the more sparse (99)b. This is certainly a label-elimination approach, but different from what Collins pursues. Stabler (1999) suggests something along the lines of what Collins proposes with the minimal difference of including a pointer which indicates the asymmetry of projection (i.e., which dominance-line constitutes the link to the head of a given combination — as in (100)). Collins rather offers an inventory of principles which conspire to yield the results that labels are typically meant to encode (which, if correct, would eliminate the need for any such 'pointer' indicating the head of the structure). Stabler's view as far as I can see wouldn't support the system I am elaborating here either, for basically the same reason that we cannot deploy Collins' approach.

However, there is a general question about all of these approaches that is worth raising: *What is going on here*? Eliminating primitive (intrinsic/non-relational) bar-level distinctions is not a new idea in the theory of phrase structure. This was present in the work of Muysken (1983), who adopted a relational conception of these distinctions

(specified in terms of coherency conditions on the projection distribution of features like [±maximal] and [±project]), and this general relational conception is modified and adopted in Chomsky's (1994, 1995) BPS. But what is being suggested here is an elimination of the distinctions altogether.[46]

The Collins/Stabler approach differs from Brody's in which direction we understand the elimination (better: reduction?) to work if we consider a mapping/ transition from the typically assumed sort of structure to each of these proposed conceptions. That is, for the standard view of the head-complement unit in (101), we have the following two alternative conceptions, differing on what is retained in the model.

(101)



Of these two ways of thinking, Brody's approach might seem at first blush to be more radical. The Collins/Stabler approach retains the part-whole structure central to the last half century of work in generative grammar (indeed: to most if not all of the entire history thinking about language structure generally!) by maintaining the head/phrase distinction in structural terms while retaining labels only for heads. Brody clearly intends to stay within this tradition as well, though his view of basic phrasal structures, he notes, is

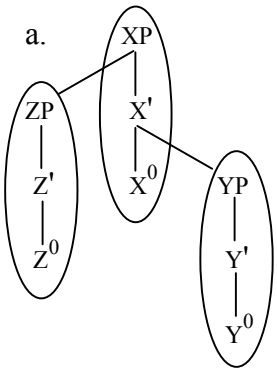
[46] Or, perhaps more accurately, relocating the conceptual/empirical burden borne by these notions onto the backs of other (hopefully independently required) ones. I refer the reader to both Collins' and Brody's discussions.

intended to eliminate as well "the apparent conflict between the long tradition of dependency theories" and "phrase structure theories of syntactic representation".[47] The issue of doubling labels in the projection relation between head and its dominating phrasal node(s) doesn't arise, as he has simply removed the distinction entirely, allowing only a single node (so, only a single label).

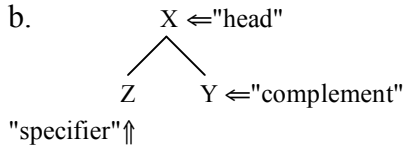
However, it seems clear that Brody's view of structure and categories can be understood to retain part-whole/constituency information via the antisymmetry of dominance relationships. Traditional heads can be understood as separate units by referring simply to a single labeled node (though see below regarding heads and PF); traditional phrasal constituents are captured via dominance as in standard approaches. The possible exception to any such straightforward mapping from standard approaches is any "junctures" involving a head with a specifier and complement. Consider (99) again, repeated here:

[47] Brody (2003:16). The "conflict" that Brody alludes to is perhaps not immediately obvious, but I think it can be unpacked as follows. It is true that classical dependency theories (e.g., Tesneire 1959) and more recent, conceptually similar approaches (e.g., Hudson's (1984) Word Grammar, among others) deploy somewhat different sorts of notations and differ at least superficially from phrase-structure/constituency based approaches in their mission statements (and there are approaches which appear to fall in both camps, Steedman's Combinatory Categorial Grammar (CCG) strikes me as one such approach). But looking at current PS-based views of structure/category its not obvious that there is any conflict. However, there *is* a difference that can be detected in the gradual historical shift from the initially deployed rewrite rules of Chomsky's early work (a pure PS-based approach) to his recent BPS. The shift has revolved almost entirely around the increasingly central role of *headship*/endocentricity. Phrase-structure rules on their own require no particular category matching between their left-and right-hand sides (e.g., X → Y Z). The recognition of generalizations stateable in terms of positing special members of local part-whole structures to play the role of determining the overall type of the local structure (headship) formed the basis of X-bar theory. Among all the notions that were subsequently introduced under this general umbrella (e.g., cross-categorial harmony, uniform bar-level limitations, etc. see Jackendoff 1977, Emonds 1985), only the key notion of headship appears to have survived in recognizable form within current thinking (see Speas 1990, Chomsky's 1994, 1995 BPS, and Chametzky 1996, 2000 for some related critical discussion). Brody's reduced structures (and, I think, Collins' as well) can be seen as attempting to remove the last barrier between the approaches, collapsing (almost) entirely the idea of formal ordering properties characterizing structure and substantive "dependency" relationships that can be understood to "live on" these dimensions. The question remains as to whether we need anything more than a single dimension. My suggestion here is that as far as narrow syntax goes we do not. This is essentially the claim that all we need is branching sequences.

(102) a.



b.



For a given sequence of head-complement relationships, dominance ordering allows us to refer to either individual nodes or to principled subsequences that respect traditional constituency (take '—' to be a dominance link in what follows, which a left-right direction on the page indicating the standard antisymmetry of this relation):

(103) A—B—C—D—E

a. D—E
b. C—D—E
c. B—C—D—E
d. A—B—C—D—E

But, for example, we might take B—C to *not* be a constituent, since there is material which both B and C dominates. It is perhaps less clear what do say about branching in this system with respect to constituency, for example (take this to be the same object as (102)b above):

(104) X — Y
|
Z

The straightforward view would say that Y (and all it dominates) is a unit, as with Z (and all it dominates), but X is not an independent unit. If it *was*, then why not X—Y excluding Z? Or X—Z excluding Y? But I just said above that we might regard each

individual node as being a separate unit in the sense of "independent head". If we are collapsing the head/phrase distinction, how is these matters resolved?

Two separate lines of discussion are relevant here. First: let me return to the discussion at the end of the previous section regarding chains as contexts and specifiers as X'-sisters versus adjunction structures, connecting it now with the possibility of adopting these reduced structural descriptions in our formulation of TCG. Second: there is the (more recent) idea that head-movement relationships might be a "PF" phenomena.

Regarding the first: there is a general idea that has been floating in the literature that non-minimal/non-maximal (intermediate-level/X') phrasal structures are invisible in some sense for the operations of the syntax. Sorting out this issue, as Chomsky (1995:382n23) observes, "depends on properties of phrases that are still unclear". For example: Kayne (1994) argues from his assumed Linear Correspondence Axiom (LCA) that all specifiers in fact realize adjunction structures; Starke (2000) argues that we dump the notion of specifier altogether, retaining only the notion of head-complement relations.

On Kayne's view we could argue that the system cannot refer to intermediate units since the equivalent element in his system would always constitute segments of category, which his approach consistently treats as essentially "one thing" (see, e.g., his definition of c-command). But, on the other hand, segments of a category on this view are labeled identically as XPs, so perhaps they *can* be referred to as independent units (certainly for the case of similar kinds of structures arising with adjuncts/modifiers we want this to be so).[48]

[48] Whether we take the modifier case to be the same as the "adjunction" case (meaning adjunction now as the C-adjunction of the sort May (1985, 1991) and Chomsky (1986) discuss) depends on whether we take these to work in exactly the same way. On "adjunct" versus "adjunction" see Chametzky (1996, 2000).

On Starke's view, which has it that the analogue of specifiers are understood to be a special case of "heads" in that they project their properties to determine (part of) the label of the dominating structure, the structure corresponding to intermediate projections in standard X-bar theory *would* be a "visible" unit since it will always be a maximal projection.[49]

Epstein & Seely (1999), on the other hand, argue that intermediate projections are real, but that they behave as "fossils" (see Chomsky 1995:382n24) having initially been maximal but losing this status when they are targeted by a merge operation on Chomsky's BPS-relational view of intra-phrasal projection. But, based on the assumption that these elements are no longer visible to the system, Epstein & Seely go on to argue that such elements cannot possibly be sisterhood contexts defining chain-links as suggested in Chomsky (1995) since the relevant elements are by hypothesis invisible — therefore, they conclude, chains cannot exist.

Both Chomsky (1999) and Lasnik (2000) point out that its not obvious that intermediate invisibility rules out the merge-context view of syntactic chains, as it seems reasonable to take the motherhood relationship to define the local structure identifying chain links (as I suggested above for independent reasons specific to my technical ambitions here).

However, one might respond to this suggestion — on behalf of Epstein & Seely — by noting that this just moves the problem around somewhat. On the motherhood view

[49] This is Starke's notion of "checking". Instead of having αP with feature {F} enter into a relation with a βP with the same feature {F}, Starke suggests that αP simply *projects* its {F} upon combination with βP. See Starke's discussion for details.

of merge contexts it is true that intermediate movements will now have visible contexts, as they will typically (always?) be dominated by XPs. But now the base position of a given chain should have an invisible element as its context, since presumably its mother will always be a non-maximal element. Though, this would depend on whether there is a specifier for the head of the base position — if so then the context will be maximal and hence visible, if not then it would be intermediate and thus invisible. Note that, as pointed out above, the specifiers-as-adjunction-structures view of Kayne and others might allow us to sidestep these technical problems if we could motivate the possibility of having segments of a category serve as appropriate contexts for the understanding of chains we've been discussing.

It is not, I think, quite clear what is really at stake here. That is, I agree with Chomsky that debate on this subject turns on presuppositions about "properties of phrases that are still unclear".

We can elaborate on this point in another general way. Given the explosion of functional categories that has attended the development of the MP, its an open question for any given element X whether an element which *appears* to be its specifier is, in fact, rather the specifier of a functional element Y that takes X as its complement. If there is such a Y, then X will be maximal and hence visible; if not, it will be intermediate and hence invisible.

The degrees of freedom that theory makes available for analysis here makes it difficult to sort out these alternatives. Note that it is not *impossible* — the present point is only that array of distinctions made available with these various degrees of freedom

simply predict more classes/groups of facts than an approach without such degrees of freedom (and are thus less restrictive).

What we might worry about even at this level of generality is what we might take as independent reasons to introduce principled/motivated constraints in the deployment of theories/models with this many alternatives (i.e., to narrow the possibilities/reduce the degrees of freedom). For approaches which adopt fine-grained functional category inventories *and* a maximal/intermediate level distinction *and* the possibility of C-adjunction (with double XP segments) things get even less clear in terms of the restrictiveness of the overall theory.

This whole set of issues ties into a discussion from earlier years, as set out helpfully in the work of Sturrman (1988), regarding projection-level types. Sturrman develops what he refers to as the Single Projection Type Hypothesis (SPTH) which divides syntactic categories into two basic types: (i) recursive and (ii) non-recursive. The latter we can take to be heads (X^0s); the former are the equivalent of maximal elements (XPs).[50]

It is exactly the concerns regarding issues of restrictiveness raised above that drives Sturrman's theoretical developments in this respect, and it is concerns of this type (as well as his aim to eliminate redundancies) that similarly drive Brody's introduction of the collapsed structures discussed above. Brody's view renders trivial, for example, the general fact that projection lines (e.g., X^0—X'—XP) can never be interrupted by some

[50] Sturrman cites early work of Emonds (1971) where the SPTH is proposed, and Emonds (1973), where the idea is rejected in favor of having two recursive types (the equivalent of modern day X' vs. XP if we take XP to be potentially recursive). Sturrman does not discuss the issues which would arise for head movement that might force the adoption of sub-X^0 structure for which one might want to posit recursive X^0s.

other intervening element of a different type since, in his reduced structures, there *are no* such internal distinctions, and therefore there is simply no room for any such interveners.[51]

The general view, however, raises questions about what *does* go in place of the distinctions typically understood to underwrite, for example, head-movement versus XP-movement (or the phrase-structure status of modifiers).[52] This brings us to our second relevant line of discussion regarding the Brody-type reduced structures and "constituency" from above: the idea of head-movement as a PF-phenomena.[53]

Chomsky (1999) (see also Boeckx & Stjepanovic 2001, Bobaljik 2001) suggest that head movement might not be part of the syntax proper, but rather is a PF-phenomena. However, note that on these views it is certainly not the case that syntactic structure simply does not matter for such operations. For example, Bobaljik's (2001) approach takes syntactic structure to yield a weak pairwise ordering which head-to-head relations are established, so saying that head-movement is a PF operation doesn't imply that it is not constrained in some manner by syntactic structure.

---

[51] But what of the SPTH of Sturrman? Do we have recursive catergories, or not? Strictly speaking, the notion of recursion refers to a function that calls itself. So we say that sentences embedded in sentences manifest recursion, and similarly with noun phrases inside of other noun phrases. But in the more recent era of separating out sequential arrays of functional and lexical types, do we ever have instances of *local* recursion in the sense of an X taking an XP complement? Work by Hoekstra (1984) suggests not, formulating what he called the Unlike Category Condition (UCC: *{X^0 XP}). See van Riemsdijk (1998) for critical discussion and an alternative formulation of the key intuition which avoids some potential problems which arise.

[52] I won't be discussing adjuncts/modification in this work.

[53] What follows regarding "head movement" superficially parts ways with Brody's discussion, who argues (following Baker 1985 and others) for a mirror-theoretic understanding of syntax/morphophonology connections. What I am about to suggest however does not strike me as incompatible with Brody's proposals (see my earlier remarks as well on pre-/post-spell-out coherency and conservation of ordering properties).

Let us now tie these two strands of discussion back into our discussion above of constituency in the Brody-type reduced structures. Consider again our abstract dominance sequence and the possible constituency groupings:

(105) A—B—C—D—E

a. D—E
b. C—D—E
c. B—C—D—E
d. A—B—C—D—E

We can now tentatively adopt the view of head-movement as a PF-phenomena by saying that the PF-relevant properties of the individual nodes (A, B, C, etc.) are PF-constituents, which are related by principles that may involve reference to syntactic structure (perhaps along the lines sketched in Bobaljik 2001) but which only actually handle the PF-relevant properties. The issues regarding constituency with respect to individual heads thus fall outside the syntactic system.

Now consider branching and constituency again with reference to these reduced structural descriptions:

(106) X — Y
|
Z

Now we are free to take the line suggested above regarding phrasal constituency in terms of traditional dominance ordering. On that view the object in (106) manifests three constituents, the entire object, Z (and whatever it dominates) and Y (and whatever it dominates).

Note however that our view of constituency can interact with *directionality* of structure building. For example, on a top-down view, the Brody-type structure in (107) would have a derivation like that in (107)a-d:

(107) A—B—C—D—E

a. A—B
b. A—B—C
c. A—B—C—D
d. A—B—C—D—E

This kind of alternative is argued for in the work of Phillips (1996, 2003) on the basis that it yields a fundamentally different (derivational) conception of constituency making available units that he argues we need for analysis.[54] It will be important to see whether the assumptions that lead to the conclusion here regarding our suggested treatment of SCM are roughly consistent with Phillips' solution to various constituency-test puzzles. If so, then the two independent lines of thinking — one regarding the dynamics of local unithood and one regarding the dynamics of reducing linked-local relations to local ones — can be seen to be pointing in (or rather, "to") the same general *direction*. (I do not address this issue here, though it seems to me that these reduced structures are consistent with what is needed to implement Phillips' analyses).[55]

However, I am not principally concerned with either of these general sets of issues (i.e., head movement or constituency *per se*). Therefore, my adoption of

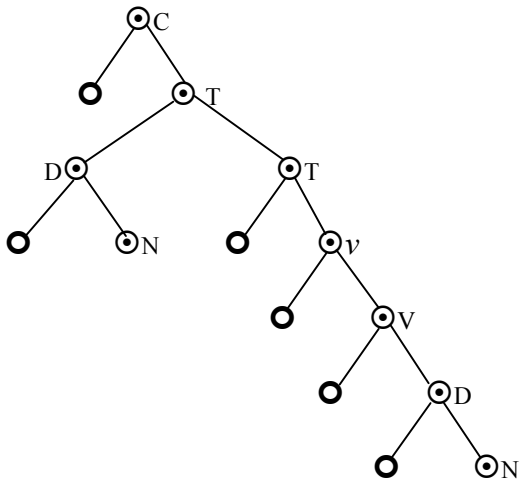
[54] Phillips gets a bit more than just what we arguably need. On his view *any* left-edge grouping is a possible constituent. In virtue of this his analyses need to appeal to other notions to avoid overgeneration (though he argues the required 'other notions' are independently motivated). See Phillips (2003) in particular for discussion.

[55] That is, what is required is to be able to refer to spec-head constituents excluding complements. As far as I can see this distinction is available in a top-down expansion of structure appealing to these reduced Brody-type structures.

assumptions regarding structure and category for the present work can best proceed by seeking out a way to concentrate on the aspects of these concepts that *are* of interest for me here. I will thus be working with the reduced structures of the type Brody proposes, though the view here will be understood to be derivational, while Brody has extensively argued in favor of a representational view (see, e.g., the papers collected in Brody 2003).

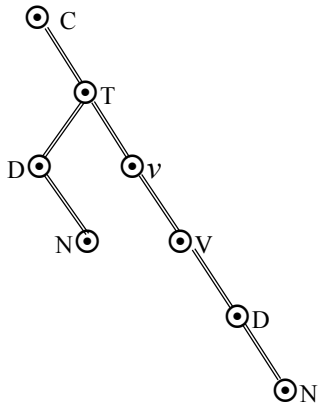
Consider the following graph of a typical transitive clause:

(108)



We can extract the Brody-type structure as follows:

(109)



It will be along these dominance spines that the all the action of the system developed here will happen. To save space in the presentation I will adopt a horizontal notation, so that (109) will look like (110) (the connecting arcs representing dominance):

(110) C—T—*v*—V—D—N
|
D—N

For everything I will be arguing here, it will be sufficient to refer to simple sequences of this kind (though we will augment the labeled nodes with more complex feature descriptions as in §1.2 above).

Note that we can take the adoption of this kind of structure as either fully embracing the Brody-type vision of structure and category labels, or we can simply understand this to be a suitable set of working assumptions which abstract away from the issues of intermediate-level categories, whether we treat head movement as "in" the syntax or not, and questions about how non-argument modifiers are integrated. That is, what this sparse representation allows us to concentrate on is the key type of information that I will be taking to be important for the TCG system — namely the nature of category sequences defining the dominance-spine of syntactic objects. Most if not all of what I will say here is consistent with this weaker view of adopting these ideas as simply a set of working assumptions. However, as noted above, this view collapsing intra-phrasal distinctions seems intuitively more compatible than some other possible approaches with the idea that the workspace cannot tolerate multiple tokenings of a given type X. And given the arguments above that we require motherhood/dominance to underwrite the context-identification view of SCM-type relationships, having a model within which this

is the *only* possibility provides an attractive convergence of independent ideas. These correspondences with our central aims, along with the ability to circumvent the numerous technical difficulties that I mentioned above in connection with some salient alternatives, will be taken as sufficient justification to proceed with these assumptions.

1.6. Chapter Summary

We now have the following ideas in place. We assume the WS/O-distinction as a basis for our TCG implementation of an MSO-system. The workspace has been suggested to be restricted in two ways:

(111) **WORKSPACE ORDER:**
The elements in the workspace manifest a weak partial order

(112) **WORKSPACE DISTINCTNESS (ANTI-RECURSION):**
The workspace does not tolerate the presence of multiple tokens of type X

For a workspace containing an X-element, I have suggested that the process of introducing any second X-element should be understood as part-and-parcel of the contraction procedure. One way that contraction can occur is in virtue of a particular response of the system when confronted with like elements — they can be identified under what I called matching relation $\Re$, which we will see in Chapter 3 requires some further elaboration.

The following strengthening of the ordering restriction on workspaces was suggested as well:

(113) **Workspace Connectedness (DOMINANCE):**
The elements in a given syntactic workspace must manifest a connected dominance order (for every *x, y* in the set, either *x* dominates *y* or *y* dominates *x*)

This effects a fairly radical partition of structures, so that the workspace always only contains essentially a "single line", indexed via feature relationships so that there is coherent maintenance of spelled-out branches of structure. Moreover, given the mechanics of node-identification, it was suggested that spelled-out structure may "re-enter" the workspace in certain principled circumstances, and then be required to "re-spell-out" (and then re-enter again, and so on). Again, I will explore this view in connection with SCM phenomena in Chapter 3.

We have adopted a reduced vision of category/structure, importing ideas from the work of Brody (2003). This view was argued above to make for a clearer, technically less complicated fit with one of the central intuitions of the TCG approach as stated above in (112). We can now note that this point-of-view can be strengthened a bit. If something like (112) is correct, then something like the Brody-type reduced structures might be in fact *required*. The alternative would be to introduce a way of distinguishing between X^0, X', and XP. But the entire *point* of Brody's proposals — and this holds of the Collins/Stabler view mentioned briefly above as well — is that these are distinctions that we can and should learn to live without, as they are redundant with other independently required concepts (we may dispute this, it is ultimately an empirical matter, but that is the claim, and it is the right one to advance on minimalist grounds). Indeed, one can take this general direction of theory-development as a natural continuation of Chomsky's (1994, 1995) BPS-project, which was aimed at (among other things) eliminating primitive bar-level distinctions.

The key ideas underlying our view of SCM-type relationships was seen to rely on a weakening of Chomsky's (1995) view of chains as sets of contexts, where "contexts"

were understood on his approach as the entire previously established structure that an element α merges with. But this view was suggested to be too strong, as it yields unique contexts for each "link" of any complex chain. The key idea here revolves on a denial of this — it is the fact that contexts are not uniquely identifiable that permits cross-domain movements of the linked-local/SCM sort.

I turn next to a more detailed empirical and theoretical discussion of SCM.

CHAPTER 2: Regarding Successive Cyclic Movement

In this chapter I discuss a number of issues regarding syntactic theory and analysis with reference to successive cyclic movement (SCM) and related phenomena. First, I canvas an array of empirical considerations that have been taken in the past to argue in favor of SCM. I then discuss the issue of what motivates the intermediate/non-target movements posited by SCM analyses (the TRIGGERING problem — see (64) above in §1.4.2) alongside the issues of how we ought to regulate phases and understand localizing evaluation for convergence (the CONVERGENCE problem).

2.1. Types of Successive Cyclicity Effects

I turn now to take stock of the sorts of considerations that have led grammarians to think that something like successive cyclic movement operations are for real. The initial motivation for positing successive-cyclic movements came from discussion and arguments in Chomsky (1973), where it was proposed that *wh*-movement ought to be viewed as clause-local, with the "edges" of clauses (a COMP node made available under S-bar) serving as escape hatches. For some time then successive cyclic movement was motivated only by theory-internal considerations arising in the proper treatment of movement locality. However, a number of other phenomena have since been brought forward that are of a different sort.

Most (if not all — see below) of these phenomena provide contribute to the body of converging evidence for the idea that movement relations are not generally "one-fell-swoop", but rather manifest a linking of local relations. This idea of successive cyclic or

linked-local relations, as we will see here, brings a remarkably diverse range of phenomena into a single abstract class.

It will be helpful in this discussion to make reference to a set of distinctions drawn from Abels (2003). He distinguishes between some logical possibilities regarding ways that movement relationships might affect or interact with intervening material, discriminating between "punctuated" and two different general types of "uniform" conceptions. Consider first this general schema:

(114) FILLER ... GAP
PATH

The possible effects that any such filler/gap relationship might have on elements along the path between them can be discussed with reference to the following tri-partition:

(115) a. α ⌊——————————⌋ β (Uniform; no effect of α—β on the path)

b. α ⌊——|——|——|——|——|——|——⌋ β (Uniform; entire path effected by α—β)

c. α ⌊...|...|...|...|...⌋ β (Punctuated; only parts of path effected by α—β)

We will see that the various phenomena that have been argued to favor successive cyclic movement analyses are not homogeneous — all of the possibilities in (115) are instantiated.

2.1.1. *Wh*-Copying

What strikes me as one of the most intuitively convincing types of evidence is the following: in certain languages we actually *see* overt copies of moved elements. The

following data are drawn from an interesting summary and theoretical discussion in Felser (2004) illustrating this phenomena for *wh*-movement in a number of languages:[56]

(116) a. ***Wen*** glaubst Du ***wen*** sie getroffen hat? *German*
who think you who she met has
'Who do you think she has met?'

b. ***Wêr*** tinke jo ***wêr***-t Jan wennet? *Frisian*
where think you where that-CL J. resides
'Where do you think that John lives?'

c. ***Waarvoor*** dink julle ***waarvoor*** werk ons? *Afrikaans*
wherefore think you wherefore work we
'What do you think we are working for?'

d. ***Kas*** o Demìri mislenola ***kas*** i Arìfa dikhla? *Romani*
whom Demir think whom A. saw
'Who does Demir think Arifa saw?'

This phenomena is *punctuated* in Abels' sense — this kind of copying is only available at clausal boundaries (i.e., CPs). Below we will see other types of evidence that suggests the possibility that *wh*-movement might generally be "more successive cyclic" than this, implicating the edges of VP as well (specifically *v*P). So we will want to ask why this copying phenomena does not show up anywhere but clause edges.

Interestingly, we also see this kind of copying phenomena in the L1-acquisition of English, where the target grammar does not ever permit this kind of copying. Consider the following case of so-called 'medial-*wh*' (De Villiers *et al.* 1990; McDaniel *et al*. 1995; Thornton 1990):

[56] Felser (2004) draws the following examples from the following sources: (116)b is from Hiemstra (1986: 99); (116)c from Du Plessis (1977:725); and (116)d is adapted by Felser from data in McDaniel (1989:569n.5).

(117) a. **What** do you think **what** Mini put on _ ?
b. **Who** do you think **who**'s in the box?

The existence of this kind of copying in early child English raises interesting challenges for the Subset Principle (Berwick 1985), as this apparently lies within a superset of the grammar of standard English. It would seem then that learners would require negative evidence to abandon such *wh*-copying. But regardless of how this is sorted out (perhaps in terms of an *indirect* sort of negative evidence), the existence of such cases points strongly towards the reality of something like SCM.

2.1.2. Q-Stranding

Other cases show an effect that is intuitively related to the copying phenomena illustrated above, where it is alleged that we can see *part* of a moved expression "stranded" in positions which it has by hypothesis moved through. Facts of this kind include the patterns of so-called quantifier stranding reported in a dialect of Irish English by McCloskey (2000) — specifically in *West Ulster English*.[57] Consider:

(118) a. What all did you get *t* for Christmas? *Standard English*
b. Who all did you meet *t* when you were in Derry?
c. Where all did they to *t* for their holidays?

(119) a. What did you get *t* for Christmas? *Standard English*
b. Who did you meet *t* when you were in Derry?
c. Where did they go *t* for their holidays?

(120) a. ***What*** did you get ***all*** for Christmas? *West Ulster English*
b. ***Who*** did you meet ***all*** when you were in Derry?
c. ***Where*** did they to ***all*** for their holidays?

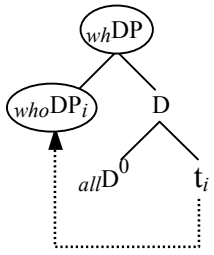
[57] McCloskey refers readers to the work of Henry (1995). Apparently the Q-float-type phenomena under discussion varies, like the copying phenomena discussed above, by dialect, and like "Standard" English is not present/possible in *Belfast English*. See McCloskey's paper for further discussion of dialect differences in this regard.

McCloskey notes that in Standard English the cases in (118) versus (119) differ in whether they require "that the answer is a plurality [...] insisting on an exhaustive ((118)), rather than a partial, listing of the members of the answer set". The interesting cases from West Ulster English, in (120), are claimed by McCloskey pattern in these interpretative properties with the examples in (118), and not those in (119). He notes that this phenomena is not exclusively tied to matrix clause interrogatives, but appears in embedded environments as well:

(121) a. I don't remember **what all** I said *West Ulster English*
b. I don't remember **what** I said **all**

McCloskey develops a stranding-type analysis for these phenomena, whereby *wh*-elements move successive cyclically and may abandon the associated element *all* at places along the movement path. Following earlier proposals (Postal 1974, Koopman 1999), McCloskey assumes a structure like (122) for the *wh*-element plus the quantificational element *all*, analogous to ideas that have been put forward in analyses of similar phenomena involving NP-movement (see below on stranding and raising-to-subject):

(122)



This structure, McCloskey argues, allows the possibility of either of the two circled nodes to undergo movement (i.e., he makes the not unreasonable assumption that both the specifier and the dominating DP node both bear the *wh*-properties as marked above). If

this is correct, then in principle every position to which the *wh*-moves could be a position where *all* is stranded.[58]

Important then for the notion of successive cyclic movement are the following cases in (123) which illustrate that *all* can be stranded in an intermediate position:

(123) a. **Where** do you think **all** they'll want to visit *t*?
b. **Who** did Frank tell you **all** that they were after *t*?
c. **What** do they claim **all** (that) we did *t*?

This kind of stranding phenomena is sometimes referred to as "floating" in virtue of earlier approaches to these matters which involved transformational operations that literally moved (floated) such elements from position to position.[59]

In addition to the logical possibility of true "floating" as a way to possibly analyze such cases, there is also an extensive literature treating phenomena of this type in terms of analyses that base-generate *all* in the various positions where it occurs, thus suggesting the possibility that the statement of the laws governing distribution of elements of this type might be independent of the putative path of successive cyclic movement operations.[60]

We can push this point a bit with a consideration of similar phenomena as it arises in cases of A-movement (these data from standard English):

---

[58] McCloskey acknowledges that this analysis relies on the possibility of left-branch extraction — in order for the *wh*-element to strand the quantificational element *all*. Though he notes as well that such an operation falls within the bounds of known cross-linguistic variation.

[59] Analyses positing literal transformational floating were offered, for example, in Kayne (1978). See Bobaljik (2001) for a thorough review of the issues surrounding these elements and what they may (or may not) be able to tell us about the nature of syntactic structures and the properties of movement operations.

[60] Some of these base-generation analyses posit that elements like *all* have the special property that they can only be generated in positions where they can enter into a relation with a movement trace. On such views the distribution of *all* is not, in fact, independent of movement — rather the difference is in terms of whether that element ever was "in" any positions other than the one in which it surfaces.

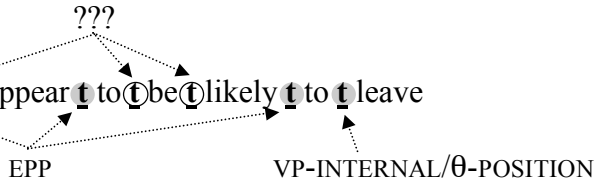
(124) a. The men **all** seemed to appear to be likely to leave
b. The men seemed **all** to appear to be likely to leave
c. The men seemed to **all** appear to be likely to leave
d. The men seemed to appear **all** to be likely to leave
e. The men seemed to appear to **all** be likely to leave
f. The men seemed to appear to be **all** likely to leave
g. The men seemed to appear to be likely **all** to leave
h. The men seemed to appear to be likely to **all** leave

The possible stranding sites appear to be a bit more prolific here than is typically assumed.[61] Although, as with McCloskey's A'-movement cases, there is apparently dialect variation here as well.[62] If the distribution of these elements marks the path of movement, then this suggests that we have the following intermediate positions:

(125) The men seemed **t** to **t** appear **t** to **t** be **t** likely **t** to **t** leave

A subset of these are motivated under fairly standard assumptions about the existence of a base/θ position internal to the most embedded VP and the existence of EPP-features marking the "subject" positions of the relevant infinitivals (marked below). But three others are less straightforward (marked with "???" in (126)):

(126) The men seemed **t** to **t** appear **t** to **t** be **t** likely **t** to **t** leave



The distribution of *all* in these cases might then be suggesting that movement is *very successive cyclic*. In contrast to the *wh*-movement cases of stranding discussed by McCloskey, which appear punctuated, this A-movement case appears to manifest a

[61] Accounts that insist that movement targets the edge of *every intervening* XP would do fine in accounting for this pattern, but then its not clear why we shouldn't see more prolific stranding in A'-movement then.

[62] Judgments on a-h vary, though to my ear they are all equally acceptable. (Norbert Hornstein informs me that he finds b, d, and g less acceptable than the rest). See Hornstein (2000) for a possible explanation.

uniform "all positions affected" relationship between the surface position of the subject NP and its base/θ-position. (That is, if we have reason to think movement really does target every single projection on the path).[63]

However, as mentioned above, there exists analyses which posit instead base-generation. While the case is by no means closed, its not crystal clear that these kinds of facts actually bear on the question of successive cyclic movement (see Bobaljik 2002).

2.1.3. Agreement on the Path

Assuming agreement relationships are typically local (not trivial; see below), the distribution of ϕ-properties might tell us something about the path of movement. However, given the possibility of an operation of the AGREE sort proposed in Chomsky (1999), where the relation between two agreeing elements can be potentially non-local, such phenomena might *not* tell us anything about the path of movement. Nonetheless, consider the following.

Kayne (1989) offers cases like the following illustrating participal agreement on the path that the relevant local A-movements would usually be assumed to take:

(127) Les filles **sont** apparu**es** avoir été reporté**es** disparu**es**
*the girls are.***3RDPL** *appeared.***3RDPL** *have been reported.***3RDPL** *disappeared.***3RDPL**
'The girls appeared to have been said to have disappeared'

However, if these relations can be licensed without movement, then its not clear what we should make of these agreement facts. On Chomsky's view such series of embedded non-finite clauses manifest no internal *phase* divisions, so it is not implausible that agreement

[63] The view I will end up endorsing in Chapter 3 regarding this particular case will be inconclusive.

could be understood to occur long-distance. It would then be an open question regarding the presence/absence of "EPP-features" that would determine whether the movement of the NP (*les filles*) from its base position would have to involve all, some, or none of the intermediate nodes.

But such a non-movement (long-distance "agree") approach is not as plausible for similar local agreement phenomena that occur in a variety of languages in the domain of A'-movement.[64] For example, consider the complementizer agreement phenomena in Irish (McCloskey 1979, 2000):

(128) Credim **gu-r** inis sé bréag
I-believe **go-PAST** *tell he lie*
"I believe that he told a lie"

(129) an t-ainm **a** hinnseadh dúinn **a** bhí ar an aít
the name ***aL*** *was-told to-us* ***aL*** *was on the place*
"the name that we were told was on the place"

Finite clauses in Irish manifest a difference between finite clauses that do versus do not contain an A'-movement trace. In the former we see the **bold** particle in (128); in the latter we see the particle **aL** (129). This phenomena manifests at every clause edge, strongly suggesting something like successive cyclic movement is at work.

Chamorro (Chung 1982, 1994) has been argued to show similar kinds of intermediate agreement effects (though see fn65 below). Consider (from Chung 1994:1):

(130) H***um***ällum si Maria [na ha-pänak si Juan i pätgun]
AGR-*assume Mary* COMP AGR-*spank Juan the child*
"Maria assumes that Juan spanked the child"

[64] Its not plausible that non-local agreement is at work here since these agreement relationships would have to cross phase-boundaries (on Chomsky's view) and clausal boundaries in general on anyone's account, *if* there is no local movements involved.

(131) Hayi h***in***alomña si Maria [*t* p***um***änak *t* i pätgun]
who? WH.*assume* *Maria* WH.*spank* *the child*
"Who does Maria assume spanked the child?"

Chung notes that, "in simple *wh*-constructions [...] the presence of a moved *wh*-phrase is signaled morphologically on some head in the extended projection of [+V]" and that in "long-distance *wh*-constructions, the special morphology shows up on every such head along the path" (glossed as WH in (129) above). Extraction across multiple boundaries thus shows the agreement effect at every clausal edge:

(132) Hafa s***in***angani-n Juan as Dolores [*t* ni m***in***alago'ña [*t* pära un-taitai *t*]]?
WHAT? WH[OBJ2].*tell Juan* OBL *Dolores* COMP WH[OBL].*want*-AGR FUT WH[OBJ].AGR-*read*
"What did Juan tell Dolores that he wants you to read?"

Similar kinds of local agreement phenomena have been documented in a number of other languages.[65] Setting aside some interesting complications facts such as these provide fairly strong support for SCM.[66]

Both the Irish and Chamorro cases manifest a punctuated effect on the movement path. Moreover, it is not possible to "skip" intermediate positions such that the agreement effects would show up both above and below a position in which the effect would be absent. Where it occurs, it occurs all the way down the structure from the fronted element to the extraction site.

---

[65] In, for example: Kikuyu (Clements 1984, Sabel 2000), Moore (Haïk 1990), Palauan (Georgopolous 1985), Passamaquody (Bruening 2001), Malay/Bahasa Indonesia (Cole & Hermon 2000, Saddy 1991)

[66] The "complications" include (i) the local agreement along the movement path evident in Chamorro is not analyzable as agreement between the *wh*-element and the local head of CP and is not strictly speaking agreement with the moving XP, but rather with distinguishable properties of the *trace* elements (see Chung 1994:7-11) and (ii) it turns out that there are cases for which such successive cyclic agreement is *optional*. Chung argues that these cases manifest a D-linked versus non-D-linked distinction (see Pesetsky (1987), and see Cinque (1991) for a view subsuming D-linking under referentiality) with the D-linked cases manifesting the optionality thus suggesting only optional successive cyclic movement.

2.1.4. Some Binding Theoretic Effects

Consider the binding possibilities of the *self*-form in (133)a:

(133) a. Which pictures of himself did John know Bill wanted?
b. [which...himself] did John know [which...himself] Bill wanted [which...himself]

Cases such as (133)a are ambiguous — the *self*-form can be interpreted as anteceded by either *John* or *Bill*. Such cases are not totally straightforward — there are a number of confounding factors that must be controlled for (see §3.2.2 in Chapter 3 on "logophors").

Here we will simply note that such ambiguities have been tied to successive cyclicity in A'-movement via the idea that the local movements create local contexts for the licensing of the *self*-elements.

Similarly, we see from the following pair of examples that similar phenomena appear to manifest in both A'- and A-movement. Consider (from Barss 2001):

(134) a. The women$_1$ asked [which pictures of themselves$_{1/2/3}$] the men$_2$ had said that the children$_3$ had brought t^{WH} to the school fair

b. The women$_1$ consider [old pictures of themselves$_{1/2/3}$] to have struck the men$_2$ as [appearing to the children$_3$ [t^{NP} to be amusing]

These cases, on a successive cyclic movement view, would be related to structures involving local copies (or traces which could be "reconstructed into") as follows:

(135) a. The women$_1$ asked [*wh*...selves$_{1/2/3}$] the men$_2$ had said [*wh*...selves$_{1/2/3}$] that the children$_3$ had brought [*wh*...selves$_{1/2/3}$] to the school fair

b. The women$_1$ consider [...selves$_{1/2/3}$] to have struck the men$_2$ as [[...selves$_{1/2/3}$] [appearing to the children$_3$ [[...selves$_{1/2/3}$] to be amusing...]

Both A'-movement ((135)a) and A-movement ((135)b) appear to manifest the same sort of "expansion" effect, as Barss notes, in terms of the available antecedents for the relevant *self*-forms. The full set of available antecedents is interpretable with a fairly tight view of binding in mind if we adopt a successive cyclic view for both types of movement as sketched above.

Other interesting cases involving binding *im*possibilities as they appear to arise in A-movement. First, consider the following cases in (136)a&b, and what seem to be a binding-theoretic violation in (136)d and (ii) the absence of ambiguity in (136)f, as contrasted with (136)c:

(136) a. $John_1$ seemed to $himself_1$ to appear to Mary to be getting fat
b. $John_1$ seemed to appear to $himself_1$ to be getting fat
c. $John_1$ seemed to Mary to appear to $himself_1$ to be getting fat
d. *Mary seemed to $John_1$ to appear to $himself_1$ to be getting fat
e. It seemed to $John_1$ to appear to $himself_1$ that he was getting fat
f. $John_1$ seemed to $Bill_2$ to appear to $himself_{1/*2}$ to be getting fat

First, note there is no problem with *John* as the antecedent for the reflexive *himself* in (136)a. Similarly, we can increase the distance and have the reflexive occupying the experiencer-PP of the second raising verb, and the reflexive binding is still possible, as in (136)b. So the first point is to take note of a three-way disjunction of possibilities: (i) either binding domains can span across levels of embedding *or* (ii) the antecedent must somehow be "in" the lower infinitival clause in addition to participating in its overt position *or* (iii) the reflexive must somehow be "in" the superordinate matrix clause in addition to participating the relations of its overt position.

All of (i)-(iii) have been advocated at one point or another, so its worth considering in terms of what else we say here which of these we options we can remain consistent with.

Now consider (136)c/d. The first experiencer-PP (*to Mary*) does not serve to block the antecedence relation between *John* and *himself* in (136)c, though somehow the relation *is* blocked where there is no overt intervening element, despite the fact that antecedence between the two experiencer-PPs is otherwise perfectly legitimate.[67]

This whole array of facts falls out nicely if we assume that the following movement operations have transpired:

(137) a. John$_1$ seemed to himself$_1$ (t^{John}) to appear to Mary (t^{John}) to be getting fat
b. John$_1$ seemed (t^{John}) to appear to himself$_1$ (t^{John}) to be getting fat
c. John$_1$ seemed to Mary (t^{John}) to appear to himself$_1$ (t^{John}) to be getting fat
d. *Mary seemed to John$_1$ (t^{Mary}) to appear to himself$_1$ (t^{Mary}) to be getting fat
e. It seemed to John$_1$ to appear to himself$_1$ that he was getting fat

The case in (136)/(137)a can be straightforwardly understood with a clause-local conception of binding domains, as can (136)/(137)b&c. In the former the overt position of the subject *John* licenses the reflexive; in the latter two it is the trace/copy of *John* in the subject position of the infinitival *to appear* which provides the local licensing.

The interesting case of blocking/intervention now arises in (136)/(137)d, which we understand to be out for essentially the same reason that **Mary appeared to himself to be getting fat* is out; that is, there is a mandatory local antecedent for the reflexive, and it

[67] Some speakers do not find the binding of the *self*-form in these cases to be acceptable. These speakers appear to prefer the a-case(s) with a simple pronoun over the b-case(s) with a *self*-form in (i):

(i) a. It seemed to John$_1$ (to tend) (to be likely) to appear to him$_1$ that he was getting fat
b. It seemed to John$_1$ (to tend) (to be likely) to appear to himself$_1$ that he was getting fat

See Chapter 3, §3.2.2 for some discussion.

disagrees in gender specification. So (136)/(137)d is out via a minimality type effect since the trace/copy of *Mary* constitutes a closer potential antecedent than *John* does. And we know from (136)/(137)e that *John* being embedded in a PP structure has nothing to do with the impossibility of binding in (136)/(137)d, since such relations are independently fine. There are some interesting wrinkles here that require addressing (in particular the status of these *self*-elements in relation to the anaphor/logophor distinction — recall I noted above that the binding between experiencer-PPs does not appear to be clause-local). I will return to these matters in the discussion and analyses in Chapter 3.

2.1.5. Interaction with Variable Binding

Consider also the following (see Bošković 2002; Lebeaux 1991; Nunes 1995):

(138) a. *[His$_1$ mother's$_2$ bread] seems to her$_2$ _ to be known by every man$_1$ to be _ the best there is

b. [His$_1$ mother's$_2$ bread] seems to every man$_1$ _ to be known by her$_2$ to be _ the best there is

This case, like the A'-movement case I will discuss in a moment, shows the necessary availability of intermediate position reconstruction. We understand the bracketed phrase to have moved through the positions marked via the underscores. In order to license the bound variable reading for the a-case, the bracketed structure must be in the scope of *every man*. But this puts the element in the c-command domain of the pronoun *her*, which induces obviation (Condition C effect). What the b-case shows, however, is that it is possible to have the bracketed phrase reconstruct to the intermediate position, where it is within the scope of *every man* but above the pronoun *her*. And, as Bošković points out, the ill-formed a-case with the indicated co-indexing is fully acceptable on the bound

variable reading so long as we have disjoint reference between *her* and *his mother*. The combination of these observations suggests that intermediate reconstruction is *possible*, but not necessary. It moreover suggests that the output structure handled by the interpretative systems must be coherent in the sense that the moved phrasal complex must be interpreted as "in" one or the other positions, but not *both* (as this would cause conflicts that would presumably correspond to unacceptability). This case, like the case involving the *self* elements discussed above, constitute to my mind quite strong evidence for something like cyclic movement.

Notice as well that the same sort of example can be constructed in the context of A'-movement, but with a twist which suggests that A'-movement is in fact even more local than just COMP-to-COMP. Consider (from Fox 1998):

(139) a. √ [Which of the papers that he_1 gave $Mary_2$] did every $student_1$ ask her_2 to read carefully?
b. * [Which of the papers that he_1 gave $Mary_2$] did she_2 ask every $student_1$ to revise?

The pronominal element *he* in (139)a must occur in a position *below* the quantifier *every* in *every student* in order for the bound variable reading to obtain the standard assumption that c-command relations determine scope possibilities relevant to such variable binding. But, in order for the construction to avoid a Condition C violation between *her* and *Mary*, the complex *wh*-expression must appear *above* the surface position of *Mary*. This means that the *wh*-expression in (139)a must reconstruct in the √-marked position and not the *-marked one, as in (140)a:

(140) a. √ [Which of the papers that he$_1$ gave Mary$_2$] did every student$_1$ [$_{vP}$ √ [ask her$_2$ to read * carefully?

b. * [Which of the papers that he$_1$ gave Mary$_2$] did she$_2$ [$_{vP}$ * [ask every student$_1$ to revise * ?

But when we switch the positions of the *every*-phrase and the relevant pronoun, as in (139)b, we see (in (140)b) that there is no position which could permit the variable binding of the pronoun (*he*) within the *wh*-expression without having *Mary* c-commanded by *she* thus resulting in a Condition C violation (this is indicated above by *-marking the relevant possible positions).

This kind of case suggests that not only is successive cyclic movement real, but that it involves more than just the typically assumed movements of the COMP-to-COMP sort. Rather, movement must proceed as follows:

(141) [$_{CP}$ [*WH*] [$_{C'}$ C^0 [.... [$_{vP}$ [*WH*] [$_{vP}$ ... [$_{CP}$ [*WH*] [$_{C'}$ C^0 [... [*WH*] ...

So the question about what features might be involved in motivating successive cyclic A'-movements appears to require properties that can be present not only in intermediate CP-positions, but also at the edges of *v*Ps.[68] Some other phenomena bear on this possibility as well, some of which I will discuss in Chapter 3.

Its worth noting here that this kind of phenomena can be shown to extend to boundaries in structure where there is arguably no *v*P, as in passives (from Legate 2000):[69]

[68] As noted earlier, *wh*-copying phenomena (as well as McCloskey's stranding) appears to never manifest at positions *other than* the relevant C-positions. Assuming this is true, whatever the motivations for these two types of cyclic movement (to CP and to *v*P), we are need of a story which explains why there should be such differences.

[69] In the first examples, Legate indicates the intended interpretation to be that Mary keeps being introduced to her own dates at parties; in the second, there is a charity auction at which dates with bachelors are sold. The argument

(142) a. √ [At which of the parties that he$_1$ invited Mary$_2$ to] was every man$_1$ [$_{v\text{P}}$ _ √ [introduced to her$_2$ *?

b. * [At which of the parties that he$_1$ invited Mary$_2$ to] was she$_2$ [$_{v\text{P}}$ _ * [introduced to every man$_1$ *?

(143) a. √ [At which charity event that he$_1$ brought Mary$_2$ to] was every man$_1$ [$_{v\text{P}}$ _ √ [sold to her$_2$ *?

b. * [At which charity event that he$_1$ brought Mary$_2$ to] was he$_2$ [$_{v\text{P}}$ _ * [sold to every woman$_1$ *?

Such cases are problematic for the view that phases are just *v*P and CP. I will discuss these matters a bit in Chapter 3.

2.1.6. Inversion Effects

In Spanish we see the following ordering alternation:

(144) a. **Contestó** la pregunta **Juan**
answered the question Juan
'Juan answered the question'

b. **Juan contestó** la pregunta
Juan answered the question
'Juan answered the question'

In certain cases of *wh*-movement, this inversion ordering is *obligatory*:

(145) a. ¿Qué **querían esos dos**?
what wanted those two
'What did those two want?'

b. *¿Qué **esos dos querían**?
what those two wanted?

---

is extended with an examination of unaccusatives as well. Legate also considers other tests for phase-hood and concludes that they all point towards a wider inventoy of phases than just *v*P and CP.

Of relevance to the existence of successive cyclic movement are cases such as the following, where we see this inversion in every local domain that we would think the *wh*-element has passed through:

(146) ¿ Qué pensaba **Juan** [que le había dicho **Pedro** [que había publicado **la revista**]]]
what thought Juan that him had told Peter that had published the journal
'What did John think that Peter had told him that the journal had published?'

Effects similar to these exist in French — so-called *stylistic inversion* — and were initially documented and argued to support successive cyclic movement analyses in Kayne & Pollock (1978). However, the relevant phenomena in French appears to be generally optional, thus it is difficult to say whether the relevant cyclic movements themselves are optional or not. The Spanish cases brought forward in Torrego's work however were argued to be an *obligatory* phenomena.

However, these data have become in subsequent years somewhat controversial as there appears to be dialect differences regarding which types of *wh*-element must trigger such effects. These differences may be systematic however. Baković (1995) compiles some of these differences as have emerged in the literature and supplements them with an extensive survey eliciting speaker judgments. The following differences emerge in Baković's study:

(147) a. No inversion with any *wh*-phrases (Suñer 1994)
b. Inversion with argument *wh*-phrases only (Torrego 1984; Suñer 1994)
c. Inversion with all but reason *wh*-phrases (*por qué*/"why") (Goodall 1991a,b)
d. Inversion with all *wh*-phrases in matrix clauses; all but reason *wh*-phrases in subordinate clauses (Baković's survey)
e. Inversion with all but reason *wh*-phrases in matrix clauses; only argument *wh*-phrases in subordinate clauses (Baković's survey)
f. Inversion with argument *wh*-phrases in matrix clauses; no inversion in subordinate clauses (Baković's survey)

These findings have not, to my knowledge, been corroborated by any independent investigation. However, the pattern suggestively converges with optionality of agreement in Chamorro, which Chung (1994) shows to hold in cases of D-linking.

Consider along similar lines the phenomena in English of matrix subject-auxiliary inversion:

(148) a. Who _ liked John?
b. Who did John like _ ?

(149) a. Who did Mary think _ liked John?
b. *Who did Mary think did John like _ ?

In many standard dialects of English, SAI is an exclusively matrix phenomena, as the ill-formed case in (149)b shows. However, there exist dialects in which such aux-inversions happen in embedded contexts as well, and they share some of the interesting properties of other cases that suggest cyclic movement. Consider the following cases from *Belfast English*:[70]

(150) a. Who did John hope [would he see _]?
b. What did Mary claim [did they steal _]?
c. I wonder what did John think would he get _ ?
d. Who did John say [did Mary claim [had John feared [would Bill attack _]]]]?

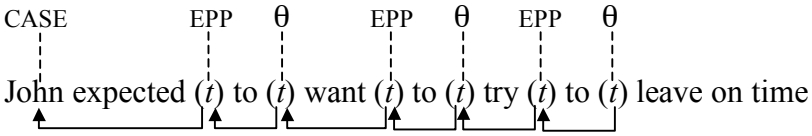
This last case appears to show the local/SCM-type effects of *wh*-movement on Subj/Aux inversion in ways analogous to the Subj/V inversion of the Spanish type.

[70] These are taken from Pesetsky & Torrego (2001), who themselves cite the extensive work of Henry (1995) on this dialect.

2.1.7. Control as Raising?

On at least some views of control phenomena sequences of control predicates must manifest successive cyclic movement. Hornstein (2000) and Manzini & Roussou (2000) develop analyses within the MP that attempt to assimilate control and raising. For example, on Hornstein's story (151)a is derived by operations allowing movement of the nominal expression into (and then out of) superordinate θ-positions, finally landing in a Case position ((151)b ignores the matrix θ-position for *expect* and marks positions as either θ or "EPP"):

(151) a. John expected to want to try to leave on time



Such a movement (raising) analysis of control phenomena has been the subject of some recent controversy.[71] In the TCG approach developed here we will encounter theory-internal reasons prohibiting any straightforward identification of control and raising, though I will suggest that the TCG mechanics bear on the relation between the two in an interesting way. In particular, a way is discussed in connecting (as does Hornstein) control with reflexivization, but in a way that remains distinct from raising (though both will involve node-identification/contraction).

[71] See in particular Culicover & Jackendoff (2003), Landau (2004), and Boeckx & Hornstein (2003).

2.2. Some Thinking on Successive Cyclicity

Consider again the case of *wh*-movement with which we began our discussion of SCM (repeated here as the b-cases in (152)-(154)) to which we now add for consideration instances of NP-movement argued to be similarly successive cyclic: so-called Raising-to-Subject (RtS; the a-cases in (152)-(154)):

(152) a. John seems to be likely to appear to like carrots
b. What did Dave think that Mary believed that John liked?

(153) a. John [seems [to be likely [to appear [_ to like carrots]]]]
b. What [did Dave think [that Mary believed [that John liked _]]]?

(154) a. John [seems [_ to be likely [_ to appear [_ to like carrots]]]]
b. What [did Dave think [_ [that Mary believed [_ [that John liked _]]]]]?

I argue that although there is ample evidence that the picture of these dependencies as decomposed in (154)a/b is largely correct, current theory does not appear to have reached anything like a consensus as to what *explains* the fact that the system works in this way.

The crux of the problem that these phenomena present for theory and analysis centers on the fact that it seems that we cannot approach these linked-local dependencies with a general reduction to the typical simple local cases in mind:[72]

(155) a. John appears [_ to like carrots]]]]
b. What [did John like _]]]]]?

[72] Strictly, the "base" position of the movement illustrated in (155)a would be a derived position, connected to the underlying VP-internal position. My focus at the moment is on the properties of the "target" position, so this example does just fine in this respect. As it happens, I will latter on suggest that the relationship depicted by the *link* in (155)a is not, in fact, a CHAIN, whereas the relation in (155)b is. What *will* be a CHAIN in the area of A-movement will be the connection between the Case/ϕ position and the *v*P θ-position (e.g., [$_{TP}$ John$_i$ [$_{T'}$ T^0 [$_{vP}$ t_i [$_{v'}$ *v*^{0} [...]]]]]. Local passive structures will similarly constitute CHAINS (e.g., *John was kicked* _).

Any simple reduction of the linked-local A- and A'-movements above to local cases like these is not possible because the properties which we might take to drive the establishment of these local relations (Case/ϕ-properties for the (155)a and *wh*-properties for (155)b) are arguably not present in the intermediate positions for the more complicated structures involving embedding.

In fact, when such positions with the relevant properties *are* available, movement relationships which would then have to cross such positions to relate to a more distant landing site are impossible, as *wh*-island and superraising cases show.

(156) a. Dave wondered [what John likes _]

b. *What did Dave wonder [whether John likes _]

(157) a. It seems John was told [_ to arrive on time]

b. *John seems it was told [_ to arrive on time]

The complement of *wonder* in (156) selects for a [+*wh*] complementizer which can be a local/direct landing site for *wh*-movement. But when this property is present, it cannot be moved across, as (156)b shows. Similar restrictions hold for raising, prohibiting an NP from moving past a Case/ϕ position ((157)b).

It is also impossible for an element which has satisfied the relevant properties be moved further to satisfy them, as it were, "again". Consider what are sometimes called "freezing" facts such as the following:[73]

[73] There exist, however, some curious apparent exceptions to these generalizations. Consider, for example, so-called 'copy-raising' constructions (Rogers 1967):

(i) John seems [like/as if [he likes carrots]]

These share with the raising-to-subject constructions the property of having a non-thematic subject (as Potsdam & Runner (2003) show in some detail). But this suggests that the overt matrix subject *John* somehow enters into a movement or chain relationship with the embedded subject pro-form *he* in (i). I will not be discussing these

(158) *John seems that _ liked carrots (vs. *It seems that John liked carrots*)

(159) *Who do you wonder _ Bill liked _ ? (vs. *You wonder* [*who Bill liked* _])

So now how do we deal with such intermediate movement cases in the general framework of feature-checking and Last Resort? The idea of Last Resort has been a central notion in minimalist research: syntactic operations are not free, but must be *driven* or *caused*.[74] Such driving causal forces are subsumed under the notion of feature-checking. So what causes the intermediate movements in this sense?

2.2.1. M-Features

Standard minimalist thinking views movement relationships as governed by a condition of LAST RESORT which generally disallows syntactic operations unless they are *required* to license/check a property or feature that would otherwise cause the derivation to crash (see Chomsky 1995:280 for one technical version of this idea). This constitutes a rejection of the notion of Move-α developed within Government and Binding (GB) approaches in the 1980s and early 1990s. On this earlier GB view, operations were regarded as essentially free, but with restrictions imposed on either derivations or on levels of representation to rule out impossibilities.

On the GB-view then, the work of stating principles governing movement was thus not typically handled within the statement of the operations themselves. Lasnik &

---

constructions here, though the general approach to them in the present framework should be clear — these cannot be of the same embedded clausal type as raising predicates.

[74] Like many ideas in current work, feature-driven operations have a longish history, present in very early work in the form of so-called *triggering* morphemes that were tied to various specific kinds of transformations (Klima 1968). In current MP work, the inventory of "operations" is now quite general, so there are not "construction specific" transformations that could be specifically triggered (or, at least, that is the aim for theory construction). Rather, what is triggered in current theory some response out of a small set of general options (e.g., 'move', 'delete', etc.).

Saito (1992), for example, suggest a general formulation "Affect-α" which simply says "Do anything (move, insert, delete) anything". Constraints on levels of representation (like the ECP or Subjacency) were then understood to inspect the broad range of possibilities that arise from the general Affect-α formulation, and reject those outputs which did not comply with the conditions on the various levels (e.g., S-Structure, LF).

One way of looking at this would be to say that GB-type systems looked at grammatical constraints governing well-formedness as restricting *outputs* of otherwise unrestricted operations. The MP, in contrast, can be viewed as imposing conditions on the *inputs* to such operations — that is, in the statement of laws governing what can count as a legitimate structural description over which an operation can apply. Operations can be, and in fact are, still viewed as general in the MP (e.g., merge, move, insert, delete; as opposed to, e.g., "passivize"), but the leading idea for theory construction is one of economy — nothing can happen unless it is forced. Viewed in this way, the differences between GB and more recent work in the MP take on a bit more subtlety. We can see the subtlety of the GB/MP difference as it emerges in current work's appeal to what I will call "M-features".

Much current work which attempts to formulate analyses of successive movement consistent with Last Resort does so at the cost of advancing closer towards genuine explanation. In effect, what much recent work does to motivate these movements in a way consistent with Last Resort is to adopt a brute force solution; that is, to posit a

feature or property of the positions constituting the intermediate links in these composed dependencies. Call this a Merge/Move-feature or M-feature for short.[75]

This is potentially saying something quite exciting, perhaps despite appearances. We could understand this claim about the forces driving movement relationships as having hit upon a genuine explanation. Thus, movement is *primitive*: a deep, central, irreducible fact of the matter regarding the workings of the faculty of human language (FL). In minimalist terms, it is neither to be motivated with reference to the nature of the interface between grammar and other cognitive systems (natural interactions) nor in virtue of the internal coherence of the system based on virtually conceptually necessary assumptions about its workings. Movement is, rather, a *basic property* of the narrow syntactic component of FL.

On the other hand, the M-feature view could also be taken to be saying something perhaps more depressing. That is: we simply do not have a clue why there might be displacement properties in human language syntax. The best we can do at the moment is catalog the facts, and hope that something turns up which might lead us towards a better understanding. Or, slightly less depressing: we may have *lots* of ideas about why there ought to be such displacement properties, but no terribly strong reasons for thinking any one of them versus any other is *right*.[76]

---

[75] In certain cases this feature is suggested to be satisfiable by Merge alone, as in cases of expletive elements. I return to this later on.

[76] My sense is that we currently face something more like the latter situation in theoretical linguistics. The only way out is of course to keep pursuing the options. Note that the "depressing" picture needn't be taken to be all that depressing. Sometimes we go through stages in the development of ideas where really the best we can do is to put like things into the same box, and keep looking for good generalizations that have further predictive power. So while I will be writing here with a skeptical tone about what I'm calling "M-features" (e.g., EPP-/P-features), its important to realize that Chomsky's recent generalization of these properties is a sensible move in this respect in that it broadens the range of phenomena claimed to go "in the same box", whatever their ultimate explanation. My view, developed here, is that M-features actually pick out a heterogeneous set. Local and successive cyclic "movement" are not the same phenomena.

M-features have taken on an increasingly important role in recent developments in minimalist syntactic theory. Chomsky (1999) suggests that the M-feature conception of movement be generalized beyond just the kinds of intermediate/non-target movements taken to be involved in successive cyclicity. He proposes alternative mechanisms to handle the licensing relations that in earlier manifestations of minimalist syntax had been understood to be the movement-driving ones (like Case/ϕ or *wh*-features). Fukui (1986) refers to such properties as Kase features; I will refer to them here as Core Licensing Properties (CLPs).

So, these relationships between CLPs in Chomsky's recent work now fall under his PROBE/GOAL relations and the notions of "matching" and AGREE. And, having relocated the job of handling these CLPs in this potentially non-local (or, rather, *less* local) way, he suggests that what drives actual displacement is rather M-features — for *all* movement. This means abandoning M-features in connection with current approaches comes with the burden of rethinking movement generally, not just the mysterious successive cyclic ones, but the local cases as well.

A quick illustration with the A-movement example from above can serve to illustrate this recent generalization of the role of M-features, as well as the basic idea behind the AGREE operation. Whereas minimalist accounts previously understood all but the last of the sub-steps of movement to be driven by M-features (as in (160)a), now all such steps are so motivated, as in (160)b', following the independent Case/agreement (κ/ϕ) licensing which happens "long distance" via the operation AGREE. The agreeing item, dubbed the probe, scans its c-command domain for a matching element or goal

(e.g., *John* in (160)b in virtue of its matching κ/ϕ-features). Once AGREE takes place, it is the presence of M-features which drive the various sub-movements in (160)b':

(160) a. $\Downarrow^{\surd[\kappa/\phi]}$ $\Downarrow^{\surd[M]}$ $\Downarrow^{\surd[M]}$ $\Downarrow^{\surd[M]}$ $\Downarrow^{[\theta]}$
John [seems [_ to be likely [_ to appear [_ to [_ like carrots]]]]]

b. **PROBE⇒⇒⇒⇒⇒⇒⇒⇒⇒⇒⇒⇒⇒⇒⇒⇒⇒⇒⇒GOAL**
_ [$^{[\kappa/\phi]}$seems [_ to be likely [_ to appear [_ to [(*John*$^{[\kappa/\phi]}$ like carrots]]]]]
⇐AGREE⇐⇐⇐AGREE⇐⇐⇐AGREE⇐⇐⇐AGREE

b'. $\Downarrow^{\surd[M]}$ $\Downarrow^{\surd[M]}$ $\Downarrow^{\surd[M]}$ $\Downarrow^{\surd[M]}$ $\Downarrow^{[\theta]}$
John [seems [_ to be likely [_ to appear [_ to [_ like carrots]]]]]

In (160)a note that there are three types of movement chains that we can differentiate in terms of what licenses their upper and lower 'links'.[77] There are: (i) chains with θ-tails and M-feature heads, (ii) chains with both head and tail characterized/licensed by M-features, and (iii) licensed (final landing site) heads with M-feature tails. Nowhere in this scenario is there a chain that resembles the basic local situation, with both the head and tail in substantive licensing configurations (e.g., a chain CH = ⟨DP$^{\kappa/\phi}$, DP$^{\theta}$⟩). (However, note that on at least one technical view discussed in Chomsky (1995), discussed above, there *would* be a chain connecting the base and final target positions).[78]

In (160)b/b' however, we see only two types of chain relation, one with a θ-tail and an M-feature licensed head, and the rest with M-feature heads and tails. Again, in the latter scenario κ/ϕ relations are taken to be an independent precondition for the M-feature

[77] I will here follow the common terminology of referring to the "top" of a given chain as its HEAD, and the "bottom" of a given chain as its TAIL. Take CHAIN for now to be a descriptive term of convenience, without intending commitment to a technical notion of CHAIN as opposed to MOVEMENT or a binding-type relation, or what-have-you. Later on I will commit to a specific view in this regard.

[78] Chomsky himself appears to have lost interest in the potential differences between various technical ways of handling movement (see in particular his closing remarks in Chomsky 1999 on this topic where both "chains" and "multiple merge" are labeled "terminological conveniences" with yet another set of terminological distinctions used to pick out the "more stern 'official' theory").

driven movements — there must be a derivationally prior PROBE/GOAL relationship. M-features have thus, in Chomsky's recent work at least, become synonymous with "movement" itself.

I find the general line problematic. The presence/absence of M-features is not principled. They are stipulated as obligatorily present wherever evidence suggests movement is not optional, and optionally present wherever evidence suggests movement is not obligatory. We might be tempted to view this — the M-feature Generalization — as a kind of reintroduction of the Government & Binding (GB-)era conception of MOVE-α. What is the status of Last Resort when we can say that it necessarily applies where features are present that require checking, but have the presence/absence of the driving properties be themselves unregulated?

But this move is not as clear as the GB-era conception of Move-α given the lack of supporting machinery of the GB-type that has by-and-large been stripped away under MP assumptions. In GB we understood applications of the general "move anything anywhere" rule as being restricted by other formal and substantive requirements, typically (though not always) stated as conditions on levels of representation so that movements could not result in violations (of Case theory, Binding, the ECP, etc.).[79] So its not true that optional M-features are reintroducing Move-α in this sense — we can't move anything anywhere. Rather, we can move when (i) the appropriate AGREE relationship holds, and (ii) when there is an M-feature.

[79] I say "not always" because accounts typically varied as to whether conditions were built-in to rule-applications or stated in terms of output filters (e.g., we can take movement to be restricted to occurring between elements in a c-command relation, or we can take movement to be "free" with conditions on CHAINS, or the like).

What governs the distribution of M-features is something of a mystery. This is, perhaps, appropriate, in the following sense. It amounts to the assertion that the existence of displacement phenomena in human language is an unsolved open set of problems, perhaps an "imperfection" in Chomsky's (1995, 1998, 1999) sense. Why *is there* displacement? What is the content of the overt versus covert movement distinction? Were we fooling ourselves in GB-era architectures or early minimalist approaches by thinking that constraints on levels of representation (in GB) or checking of core licensing properties (e.g., "morphology driven movement") in the MP was really offering us an *explanation* for the relevant phenomena? The early implementations of the Last Resort logic needed to appeal to "strong" versus "weak" flavors of core licensing properties — doesn't this distinction do what M-features are doing now (i.e., code the overt-covert distinction)?

Are we to expect a development of the concept of M-features? Should we expect to find some independent motivation for their existence? It could be, but at the moment our understanding of M-features does not appear to really extend beyond this inter-motivation — where there is movement there is an M-feature, and vice-versa. All we really know is that they are *not* features of the sort that we *used to* take as the properties driving movement (e.g., *wh*/Q, Case, ϕ, etc.).

It has been suggested for the case of successive cyclic *wh*-movement that the relevant M-features may be of a quasi-interrogative nature, but it is quite unclear why these properties should, for any *local* syntactic reason, become associated with the outer edges of embedded declaratives to drive the relevant intermediate movements, as in our

example above.[80] Alternatively, it might be suggested that these features come along for free whenever there "will be" the relevant core licensing properties present in some later derivational stage. But this means that the presence/absence of these properties requires reference to arbitrarily large stretches of syntactic computation to establish their local legitimacy. Current worries in the minimalist literature about whether this or that view of derivations requires "look-ahead" or not ought to be focused now on the existence of M-features and with how their presence is locally justified.

Other suggestions have explored the idea that such M-features are focus related, but this appears to run into similar problems regarding how to motivate them for exactly where they are needed to drive successive movement and not elsewhere (see Felser 2004). And, in general, these features are required to be present at the edges of otherwise rather different kinds of categorial domains (at least C and *v*, and perhaps D as well) in order to drive the relevant movements "out of" the domains Chomsky calls *phases*. The question of why these properties should be associated with exactly the domains that are otherwise suggested to be phases has no principled answer that I am aware of. Why should M-features line-up with phase-edges?

For A-movement, the postulation of M-features takes the form of what amounts to a return to "square one" regarding the Extended Projection Principle (EPP) introduced in Chomsky (1982) (see Bošković 2003 for some discussion of this point), essentially stipulating the need for filled specifiers for the relevant intermediate/non-target head

[80] See Lasnik & Uriagereka (forthcoming) for some discussion of this point and the notion of "lookahead" in adding what I'm calling M-features.

elements, and thus artificially creating positions that then cannot be passed over without violating minimality of movement restrictions.[81]

The salient alternative approach to the locality of such movements is to introduce an analogously artificial restriction on movement, stipulating that A-movement must go from INFL to INFL (T-to-T) and A'-movement must go from COMP-to-COMP (C-to-C). I think that this is correct as a description, as will become clear in our development of TCG below. But the important question that needs answering is, "why?". Stipulating T-to-T or C-to-C movement appears to be Bošković's (2002) conclusion regarding what I'm here calling M-feature-driven movement, namely: M-features do not exist, but movement must go through these local positions anyway (see Grohmann 2003 for an extension of this stipulation). Why? Because that's the way movement works — its local. But this is not, to my mind, solving the problem.

To be fair, Bošković does address half of the problem and this is important (see our analyses of raising and the "EPP" in Chapter 3). What he shows is that where M-features seem to be redundant with core licensing properties, we can do just fine handling facts without the postulation of M-features. However, this still leaves us with the intermediate movement cases as motivations for M-features, at least if we stay within the general borders of approaches consistent with Last Resort.

To sum up, I take current theory to be in the following bind:

[81] Frank (2002) has suggested a sort of "doubling" of EPP-feature types, so that there are both A-movement-relevant EPP-features and A'-movement (*wh*-)EPP-features — that is, two types of M-features (he posits also a doubling of Case properties, one relevant for A-movement and one for A'-movement — his so-called *wh*-Case).

(161) a. Empirically, intermediate (successive cyclic) movements seem to be real

b. If LAST RESORT governs syntactic operations without exception, there must be some feature(s) involved in these intermediate movements,...

...BUT,...

c. ...the relevant motivating properties can't be in the set of core licensing properties, since these (e.g. Case, WH) are typically satisfiable only once

d. So, either:

(i) M-features exist, and are what is responsible for *at least* intermediate movements,

OR,

(ii) M-features do not exist, and something else is responsible for intermediate movements (so either Last Resort is not completely general, or something overrides it, or it can somehow be restated so as to provide non-feature-driven movement)

Chomsky's recent move in the face of this bind is (161)d(i). However, Chomsky (1999:33) notes that what I am calling M-features are simply "selectional" features (linguistic properties) that are "uninterpreted" and which moreover constitute "an apparent imperfection, which we hope to show is not real by appeal to design specifications". This is the general route the present work aims to take: to show that M-features as such are dispensable, though the general property of the syntactic component they have been (spuriously) introduced to describe is certainly real.

In my view, the earlier stages of the MP were closer to being on the right track in the following sense: movement *is* about what I called Core Licensing Properties (CLPs). Where things have gone astray is in the specific kinds of efforts made render intermediate/non-target movements in evidence in cyclicity phenomena consistent with Last Resort. The central direction of early versions of minimalist inquiry regarding CLPs

strikes me as being essentially correct and I think that this is the direction inquiry should continue to go in the characterization of *local* relations and dependencies. The missing piece then, with such a story about local relationships in place, is a way to understand how to truly reduce the non-local cases to the local ones.

2.2.2. Cyclicity without M-features?

The larger context of the minimalist program includes the general idea that syntactic operations are governed by Last Resort. Technically this manifests as the idea that movement must effect some (local) licensing of properties that would be otherwise ill-formed (i.e., without the movement operation to create a local context for their satisfaction/licensing). A local instance of *wh*-movement happens in order to check/license the *wh*-properties specifying the interrogative properties of the matrix clause (as in *Who did Bill hit*?). But the situation involving embedding, illustrated above, requires that the intermediate C-element *not* be of the sort for which WH/Q-properties are present. So something else must drive these intermediate movements.

Work in the MP has taken one of two general approaches, either (i) some inventory of special features are introduced to drive the intermediate/non-target movements (i.e., M-features as discussed above) or (ii) some attempt is made to derive the effects of local movements *without* supposing that there is some special property there in the structure. I turn now to (ii).

This latter ((ii)-type) strategy has manifested in a number of different ways. One route of thinking along these lines simply denies SCM and proposes non-local movement is "direct" (one-fell-swoop) and thus attempts to account for local phenomena exhibited

along the movement path by other means. Zwart (1993) for example supposes that in long-distance *wh*-movement some dummy items are first inserted in all the intermediate CP-related positions, and that the *wh*-element itself then moves directly to the target position (thus obeying his view of movement economy: "Fewest Steps").

However this does not appear to be much of an advance over the M-feature view, but rather a different conception. The problem under this approach is then: what drives the insertion of intermediate elements that then allow the moving *wh*-element to pass over those positions? Radford (2001) argues for a convergence-based understanding of phases which is wed to a one-fell-swoop view of movement as well (see Felser 2004:548 for some critical discussion). Epstein & Seely (1999), as mentioned briefly above, argue on the basis of technical issues regarding chain definitions that movement must be similarly a one-step process.

It would take me too far afield to consider all the alternative possibilities involved in contrasting stories positing SCM with those of the one-fell-swoop variety. I will therefore narrow my focus in what follows to examining approaches that in one way or another have appealed to linked-local movements but which have attempted to dispense with M-features. However, it is worth pointing out that this general class of approaches requires *something like* Zwart's solution, as discussed above. Somehow the locality of SCM-*type* effects need to be accounted for, and its unclear how one-fell-swoop stories can do this without evoking strongly non-local principles. For example, we noted above regarding Chung's (1983, 1994) facts from Chamorro that the relevant agreements that manifest are tied to each and every relevant local domain on the movement path. Its unclear how to do this without postulating linked local relations of *some kind*. And note

as well that completely disallowing linked-local movements requires a special view for understanding the binding-related connectivity effects discussed above (§??). Moreover, as we will see in Chapter 3, there appear to be situations in which it looks like SCM is maybe not a necessary feature of the relationship (i.e., situations where the local effects do not show up uniformly). We saw some instances of this in the form of "dialect variability" in Irish English *Q*-stranding vs. standard English, variability in the possibility of S-V inversion effects in Spanish, variation in matrix/embedded SAI across different dialects in English, and so on. I will suggest in the next chapter a general schema for analysis which relies on the possibility of SCM enforced in the series of workspaces defining the derivation versus a rather different sort of relationship (unselective binding of the sort discussed in Pesetsky 1987 to handle so-called D-linking) that may hold over *output structures*. If the general idea is on track, then we will actually need *both* SCM and some other sort of potentially non-local (or less-local) kind of relationship in order to account for the presence/absence of the sorts of effects canvassed above. So, with an exception or two in the discussion below, I will from here on drop discussion of one-fell-swoop accounts.[82]

Another instance of the (ii)-type approach (attempting to live without M-features) comes in the form of accounts claiming that movement is supercyclic. Boeckx (2001), following ideas introduced in Takahashi (1994), suggests that intermediate movements are motivated by a direct condition on the "distance" a given movement relation can span,

[82] This is not to say that no such account is *possible*, just that there are reasons for suspicion of one-fell-swoop which do not arise for the SCM view. I take the general reserach agenda of attempting to reduce all apparently non-local phenomena to local domains seriously, which is why the interest in SCM-type accounts. I refer the reader to the references cited in the text for discussions of one-fell-swoop views.

such that movements must involve the edge of every intervening phrase between the base and final/target landing site positions. This view predicts, in Abel's (2003) terminology mentioned above, uniform effects on the movement path, and so clearly would require further assumptions to capture the various phenomena that appear more selective in their manifestation (e.g., *wh*-copying, the binding/connectivity facts).

Heck & Müller (2000) propose that Last Resort is a violable condition, and posit a system of violable constraint interaction that conspires to drive local movements without direct feature-checking involved at the non-target sites (see their notion of "Phase Balance"). However, as pointed out by Felser (2004), their approach relies on undesirable look-ahead logic (i.e., cast in the framework of Optimality Theory, their approach requires reference not just to phase-sized numerations but rather to arbitrarily large ones).

Castillo & Uriagereka (2000) offer an approach that is interesting for the present investigation in virtue of the relation of their proposals to the TAG-theoretic sort of analyses discussed earlier. I discuss them here together which allows us to make some points about the "movements" assumed in TAG-analysis within elementary trees and their need for something like EPP-requirements.

Castillo & Uriagereka (C&U) suggest there are initial movements that are local, and which involve the ultimate target position housing core licensing properties. Like TAG approaches, they suggests that such an initially formed dependency can then be subsequently "stretched" by adding intervening elements.

However two aspects of their approach are unlike analyses offered in the TAG framework. First, C&U suggest that individual merge operations do the job of "stretching" the initial local dependency by adding intervening elements one-at-a-time, as

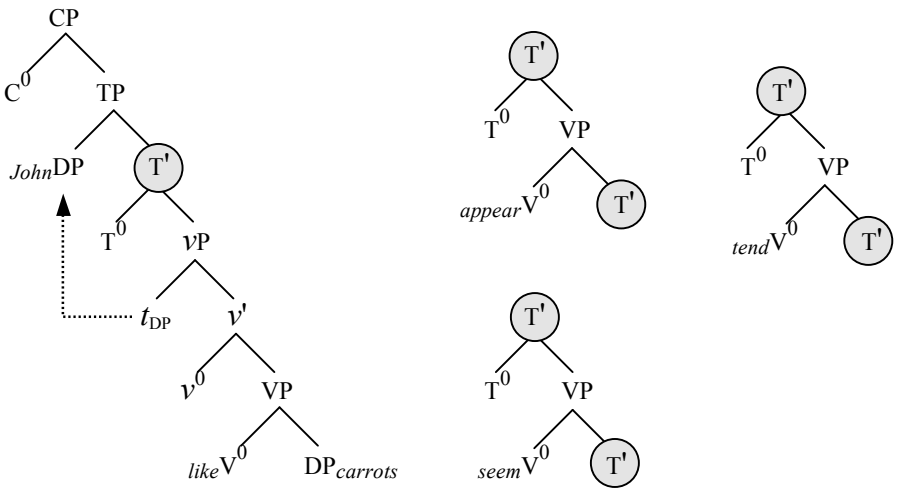
opposed to the *en bloc* insertion (adjoining/splicing-in) of auxiliary structures as in TAG. C&U take this one-at-a-time addition of intervening elements to be a merge-based generalization of so-called "tucking-in" that Richards (1997) argued to be relevant for movement operations (i.e., allowing non-root mergers, which is what C&U need to effect the incremental stretching of local dependencies with step-wise additions of single elements).

A second difference between C&U's approach and standard TAG analyses is that they take the local movement relationship to be one that connects the moving item with its actual "target" landing-site. Its worth taking a moment to show why this *isn't* how things are typically handled in TAG-theoretic elementary trees. We can illustrate this with what corresponds to cyclic A-movement in TAG, though the point carries over directly to A'-movement cases as we will see. Consider:

(162) a. John seems to tend to appear to like carrots

The TAG derivation for this example *might* be (but as we will see, typically *is* not) understood to evoke the following elementary structure (left below) and the corresponding auxiliary trees (right-hand side below):

(163)



ELEMENTARY TREE **AUXILIARY TREES**

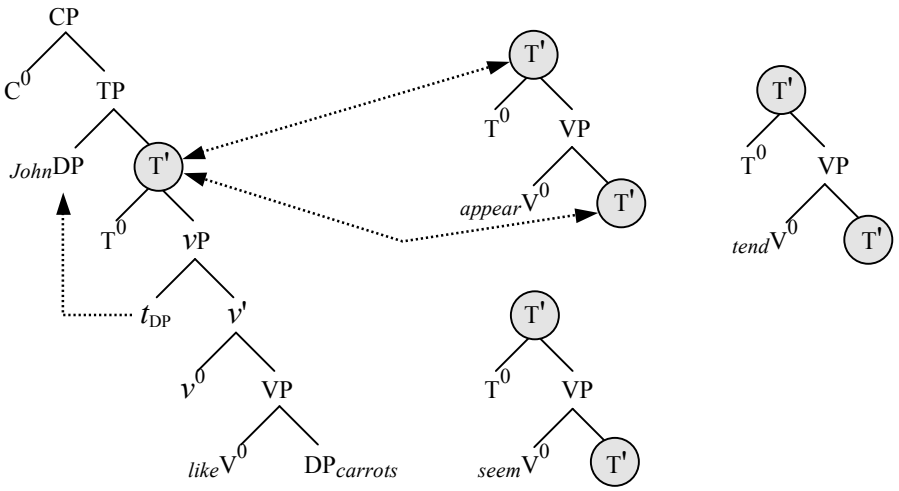
Assume that T in the elementary structure above is *finite* (i.e., that this structure converges as an independently well-formed clause). This corresponds to Castillo & Uriagereka's idea for the *wh*-movement case above.

This particular sort of division of basic tree structures turns out to be *problematic* in TAG, but stepping through this derivation will allow us to see more clearly in comparison how at least one version of current TAG analysis of such phenomena is taken to work. As discussed earlier in the last chapter, auxiliary structures have matching root and foot nodes, which are T' for raising predicates (on the right-hand side above). This reference to X'-level categories that are not dominated by a corresponding maximal node (XP) may however be inessential to the account.[83]

[83] This is an important issue connecting with the position one takes regarding primitive "bar-level" distinctions. In a BPS-style account, for example, it is unclear how one would motivate this particular property of root and foot nodes in auxiliary tree structures without abandoning the central idea of relativistic phrase-level status.Frank (2002:??n?) suggests that it may be a straightfoward task to render this approach consistent with a purely relational conception of bar-level distinctions. But this is not clear to me. He also correctly notes that having primitive C'-nodes is not at all inconsistent with Muysken's (1982) approach which worked with feature sets (e.g., ±max, ±project). Muysken's initial view is relational in that projection is understood to be a coherency condition on the values and structural distribution of such features in domination sequences (e.g., a X{+max,-proj} cannot coherently enter into an immediate dominance relation with another X{+max,-proj} node (etc.)). Thus a C'-node with no dominating CP would simply be a {-max, +proj} element. It does seem unreasonable to think that verbs might differ, in this feature-based view, as to whether they select a C' or a CP conceived in Muysken's original feature-based terms. However, Chomsky's

In any event, these root and foot nodes can be understood match/correspond to the T'-node in the elementary tree depicted on the left. The movement of the nominal element (*John*) is understood to take place from the base/thematic position (here: specifier of *v*) to the derived matrix position directly, consistent with the general idea in play within TAG approaches that all dependencies should be localized to basic tree structures that form the input to TAG adjoining and substitution. In this case, the relevant operation is *adjoining*, which works as follows for the combination of the elementary structure and the raising auxiliary headed by *appear*:

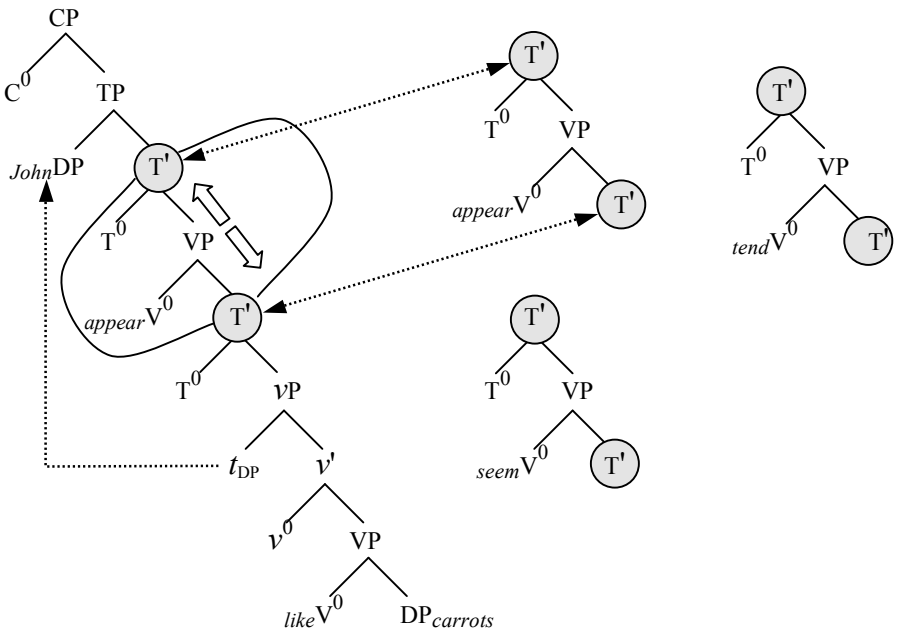
(164)



ELEMENTARY TREE **AUXILIARY TREES**

conception in the BPS approach is different in this regard. On that view projection level is *purely* relational — there can be no C' without there being a CP.

(165)



ELEMENTARY TREE **AUXILIARY TREES**

Adjoining generally effects a kind of splitting of an atomic node (here: T') in the elementary tree, splicing the auxiliary structure into its position as depicted above. The same operation could then be repeated to adjoin (splice-in) the other two raising predicates. Note that the addition of the auxiliary structures serves to "stretch" the movement dependency created in the initial elementary tree — again, there are no intermediate traces of movement in the TAG architecture.[84]

This specific derivation has to my knowledge never been advocated for RtS phenomena within TAG, and its easy to see why. The element which constitutes the matrix T-head (T^0) in the elementary tree is *no longer* the matrix instance of this category

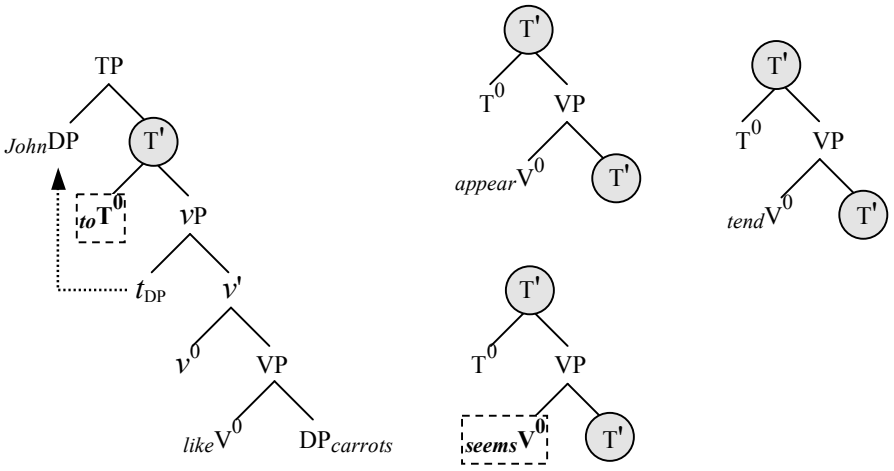
[84] The absence of intermediate traces is touted as a virtue in these approaches, though I actually think this raises some puzzles for TAG analysis; see §1.5.1, (94)-(97) above, and §3.2 & §3.3 below. However, see Frank & Kroch (1995) for some relevant discussion relying on an alternative way to think of some of the relevant connectivity facts.

in the output. Rather, it is necessarily stranded in the lower clause, with the T^0-element in the *auxiliary tree* becoming the highest instance of this head.

If finiteness is encoded on these T-elements directly (i.e., if T comes as either {+fin} or {-fin} — not a trivial assumption), then this sort of derivation strands the T{+fin} element in the lower clause. Also, if we are taking the agreement and Case properties of the T-element in the elementary tree to be enter into some kind of checking/licensing relationship with the "raised" nominal, then similarly the relevant agreement properties should be stranded in the downstairs clause as auxiliary structures as adjoined.

For reasons of these sorts, researchers working within the TAG framework have developed rather different analyses than the one sketched above. For concreteness I concentrate here on the proposals of Frank (2002). Frank's approach to Raising-to-Subject (RtS) assumes a rather different inventory of structures constituting the input to the TAG derivation:

(166)



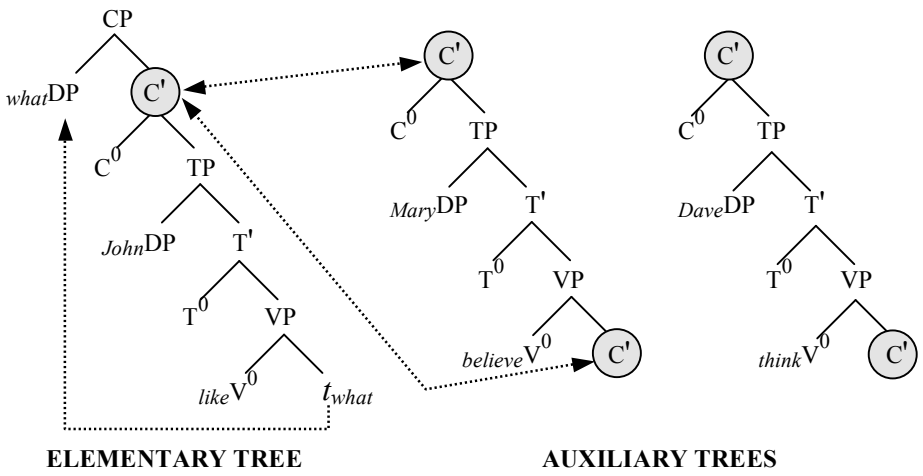
The important difference is in the nature of the elementary tree with respect to *finiteness*, and the nature of one of the auxiliary trees (for this example, the auxiliary structure headed by *seems*). The relevant elements above which differ from the problematic derivation given in (165) above are boxed. Frank assumes the auxiliary tree that will become the matrix verb is itself marked as {+finite}, while the top of the elementary tree containing the local NP-movement is has the properties that it manifests in the output, namely that it is {-finite}. This property is thus assumed to be a property of the relevant T-nodes.

Setting aside for now how the adjoining operations are constrained so that the one finite raising auxiliary ends up "at the top", we can see now that the trouble that arises in the derivation given above in (164)/(165) doesn't arise for (166) since the elementary tree is itself taken to be non-finite (as it appears in the target derived structure). The finite/matrix structure is introduced by an auxiliary structure. This requires that there be some force which is responsible for the initial A-movement within the elementary structure. Frank (2002) formulates a version of the EPP to get this result (the view is interesting but requires too much discussion to get into here — see Frank (2002) — the point here is simply that TAG requires something other than Core Licensing Properties to drive the initial movement).

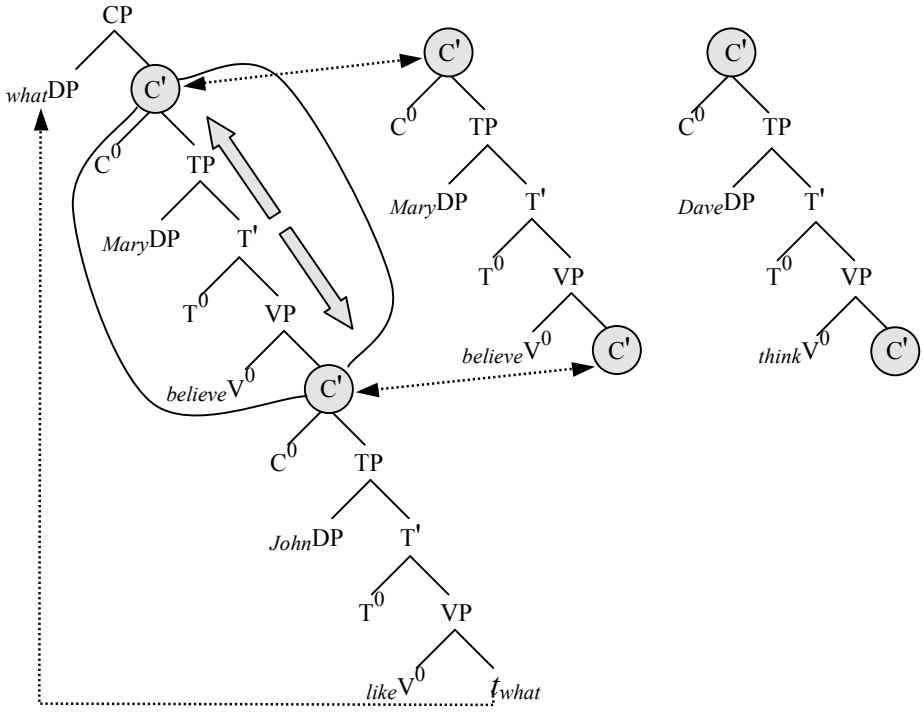
Now observe the somewhat subtler though similar situation that arises in *wh*-movement. A TAG derivation for the *wh*-movement case above would go as follows. We assume the elementary and auxiliary trees underlying the TAG derivation of (167)a as in (167)b, and the first of two adjoining operations works as pictured in (167)c:

(167) a. What did Dave think Mary believed John liked?

b.



c.



Following the adjoining step pictured in (167)c, the second auxiliary tree would be similarly spliced-in (adjoined) yielding the full output structure (or, again, perhaps the two auxiliaries first combined and then adjoined as one). As adjoined material is spliced-in in this manner, we see again that the movement dependency formed over the input elementary tree (in (167)b) is systematically "stretched". The addition of the second

auxiliary tree would further stretch this elementary-tree-defined relationship yielding in the output the superficial appearance of a "long-distance" syntactic dependency.[85]

The present point however is that the C-head that on MP (and other) views would be taken to house the *wh*-properties driving the movement to the ultimate target position is the C-head in the elementary structure above. But it is *not* this head which ends up as the "matrix" C. That head is provided by auxiliary tree. So, as with the raising-derivation given above, there must be some *other* property in the elementary tree which motivates the initial (and only) "movement". TAG derivations of this sort thus need *something like* the M-feature (EPP-type) requirements.

These matters need not distract us any further. Three points emerge from this digression about TAG versus the C&U proposal which do concern us however.

First: the view which Castillo & Uriagereka adopt involves an initial CLP-driven movement operation, and this is not how the comparable TAG derivations are usually taken to work. Rather, conventions on the adjoining operation in TAG are evoked such that *wh*-licensing is built into the "splitting" of nodes that accompanies the splicing-in of auxiliary structure. Issues arise in this TAG view however concerning (e.g.) *do*-support and auxiliary inversion generally, which I will not get into here (see Frank 2002, Rogers 200? for some discussion and some solutions within the TAG approach).

Second: our sketch of how the TAG derivations *do* work has brought to the surface the fact that the elementary-tree-internal movements in NP-raising and *wh*-movement need to be driven by an EPP-type requirement. In later discussion (Chapter 4)

[85] I'm assuming here that the auxiliary structures are spliced-in one at a time. Alternatively they could be joined together first and then spliced-in in a single instance of adjoining. See Frank (2002) for discussion.

we will see that it may be possible to import some of the assumptions about structure and category (of the Brody-type discussed at the end of Chapter 1) into the TAG framework to do away with EPP-type requirements on the movements within elementary trees. Sketching this possibility will be important to our Chapter 4 discussion of the differences between the TCG approach developed here and TAG. So C&U's idea of an initial CLP-motivated movement may turn out to be of interest in discussions of EPP-type requirements across the TAG and MP-type frameworks.

Third: C&U's discussion turns out to be of interest in the context of the TCG approach developed here for another reason. In particular, they make use of a notion of *distinctness* in discussing *wh*-islands that is helpful for us to consider. Take the *wh*-island violation in (168):

(168) ??What did John wonder whether Mary thought Dave liked _ ?

On C&U's view this begins with the following local movement:

(169) $[_{CP}$ what $[_{C'}$ C^0-[+WH] $[_{TP}$ Dave liked $t_{WH}]$

They then suggest that as the items are added-in between a moved *wh*-element and the structure containing its base position. In *wh*-island cases, as C&U point out, the derivation will inevitably reach a point where a "like" C-element will have to be introduced and projected below the initially moved *wh*-element and its associated CP structure. Such situations, they suggest, result in pathology (the derivation crashes or perhaps cancels at this point, or is perhaps problematic at the interfaces for this reason).

I have already mentioned the role that distinctness will play in enabling SCM in the present account. Later on I will consider a story similar in spirit to C&U's as an account of *wh*-island and other related phenomena.

Consider now another reasonable line of thinking which begins with an observation about one way we might simply deny SCM and implement the "one-fell-swoop" logic. We might be tempted to say that intermediate movements are *impossible* for exactly the reason that they cannot be final landing sites for elements (as noted above — see (155)-(159)). Locality of movement would then be understood in terms of closest possible landing site.

This would be in place of the somewhat odd counterfactual view of "closest" which is often appealed to, under which the "closest" landing site is simply one that belongs to the right more general super-category of positions which *could have otherwise* hosted the licensing features to make the position an actually licit target/final landing site. That is, non-finite T is of the type that could have, for example, hosted Case and ϕ-properties, if it were only of the right sub-type of the general type T; embedded declarative C is of the general type that could have hosted *wh*-properties, and so on.

But this general outlook about categories being of a more general type which could have housed the relevant core licensing properties raises another possibility. We could maintain the idea that intermediate positions do not properly license elements that move to them, but still somehow make use of the observation that such intermediate landing sites are in fact "of the general type" that could otherwise serve as target sites, if only they housed the relevant features specifying the appropriate sub-type. The idea

would be to separate-out landing-sites for movement from the motivating forces of movement.

We thus might alternatively pursue a kind of "false-advertising" view that could support the SCM view after all. Suppose subordinate elements can see up the c-command path and can detect the general (super-)category of an element as a possible suitable target of movement, and this *possibility* is enough to locally license movement to that element's domain. The fact that this would turn out to *not* license the position as an "ultimate" landing site simply presents the situation which we want to obtain, namely that the element is then in a higher position from which it may seek a target in the next higher domain and it is still "active" in terms of having its properties not yet satisfied (licensed/checked).

Note that this perspective would make movement contingent on the needs of the moving element, thus falling under some version of GREED, in contrast to approaches which either partition-out the responsibility for movement triggering between the source/target positions and the moved element, or else makes the driving force of movement the upper element's responsibility (e.g., so-called ATTRACT based view). A fully attract-based approach couldn't support the false advertising view of cyclicity, obviously, since the idea is that the relevant properties that could actually license movement are "not there". And even if do not develop an attract-based view, the basic idea seems pretty stipulative. Why should only the category, and not its features (or their *absence*) be detectable in the local context? Maybe however this view could be serviced as a landing site theory, with some other motivation given to drive the movement operation itself.

There is another way of looking at things that relates to remarks made earlier in this discussion (specifically when introducing Chomsky's DbP view in the context of our WS/O-distinction; see §1.4). That is, we might maintain that it is the local ill-formedness of a to-be-displaced item in the context of its base position that is sufficient to trigger movement. Thus, local contexts with elements in their base position (with no features checked) could be understood to somehow *expel* their uninterpretable contents.

SCM would then be driven or not depending on how the system assesses the well-formedness of sub-stretches of derivation — localizing inspection of the derivation for *convergence* would establish sub-domains which would have to have certain elements displace in order to be well-formed. This localized non-feature-driven view of movement as LAST RESORT might also be coupled with the "false-advertising" view suggested above (so there still would perhaps have to be a category of the right general type to house the expelled element).

In particular, having some notion of localized convergence evaluation (output "size" restrictions) seems promising, since we have a theory involving the idea that an item with uninterpretable properties might crash a derivation if these properties are not checked/licensed. This can perhaps be exploited to drive local movements. So, one interesting possibility for SCM is to consider shrinking the domains over which we evaluate well-formedness, so that an element could in principle *violate* Last Resort in order to avoid rendering its local environment ill-formed. And this is what the MSO-type vision of Chomsky's DbP does (even if the specific categories he has in mind as phase-inducing aren't quite right — see below — the *general* idea has the right form). We can regard this intuitively as a fit with the general idea to the view of uninterpretable features

as "viral"; that is, local contexts are forced to expel the "sick" elements in order to not remain "infected".[86]

Also, we have the idea that expelled items might have to move to a category of the right general type, which again we take to mean a category which is of the type that *could in principle* house the right sort of core licensing properties. Expelled items, needing to license their properties, go to the closest place where it looks like this could happen. That this *does not* actually happen with intermediate movements simply means that the relevant moving element remain a source of "infection" for the next highest domain, and must be expelled again to the next super-ordinate domain, and so on. (Presto! SCM!). Lasnik & Uriagereka (forthcoming) propose of view of this sort which relies on the idea of "expelling" elements "up the tree" as each locally evaluated structure is forced to come to terms, as it were, with a contained element housing viral properties.

Let us sum up these two ideas as follows:

(170) EXPULSION OF VIRAL ELEMENTS: elements with uninterpretable features must move when necessary, this means either (i) to check a feature, or (ii) to avoid crashing a derivation when a local structure is evaluated[87]

(171) MOVING INTO CATEGORIALLY LIKELY CHECKING CONFIGURATION: elements know where to look to get their uninterpretable features checked/satisfied (but the features aren't always there)

What drives SCM then is the presupposition that there are sequences of categories between a base and a target position for a "to-be-moved" element, and that each category in this sequence just happens to have the following property: ***it is an element with a***

[86] On the notion of features as "viral", see Uriagereka (1998) and a discussion in Lasnik (1999b).

[87] This is more-or-less a vision of movement economy enshrined in Lasnik's (1999a) notion of Enlightened Self Interest (versus standard Greed).

subset of the properties of the target landing site. I will take this to be central in the analyses of the next chapter.

2.3. Chapter Summary & Further Critical Remarks

We have encountered some potentially interesting preliminary ideas that might support SCM without the postulation of intermediate M-features. Recall from above we identified the following two general issues (borrowing from the formulation in Felser 2004) pertaining to MSO-type phases of derivation and SCM:

(172) Assuming SCM exists,...

a. What motivates it? Properties of the moving element? Properties of the intermediate target? Both? Neither? (i.e. something *else*)? (**TRIGGERING**?)

b. What are the mechanics of movement like such that unlicensed features {*F} do not remain to cause convergence problems in spelled-out domains (either on the copy, or on the remerge view)? (**CONVERGENCE**?)

The discussion in the previous section has focused on some alternative accounts addressing the TRIGGERING problem, and we have so far not said much about the CONVERGENCE issue as stated above, though as we just discussed at least one type of solution to the TRIGGERING issues relies on assumptions regarding local convergence evaluation.

First, note that our canvassing of some of the empirical terrain evoked in discussions of SCM raised the following additional problem, related to both (172)a and (172)b — call this the DISTRIBUTION problem:

(172) c. What are the range of potential non-target landing sites for SCM? (**DISTRIBUTION**?)

This dovetails with the distinctions drawn from Abels (2003) regarding uniform presence/absence of effects along the movement paths versus the possibility of "punctuated" effects. Note that on a purely direct-feature-driven view of movement (e.g., of the sort postulating M-features), the triggering and distribution problems are the same. But this is not necessarily so on other views, as we will see in a moment. Specifically, we encountered the following ideas with respect to TRIGGERING:

(173) a. M-features exist, and these motivate intermediate movements (consistent with Last Resort)

b. Intermediate movements occur (consistent with a version of Last Resort) in order to avoid crashing the derivation when local sub-structures are evaluated

Chomsky's (1999) position seems to couple (173)a and (173)b. But if (173)b can be shown to suffice, then one might think that we can dispense with M-features (a desirable outcome, as noted, on both Chomsky's view[88] and following the general argumentation above in §2.2.1).

But its not clear that this localized vision of last resort actually does suffice. That is, on a phase-based view, the DISTRIBUTION question is supposed to get an answer in terms of whatever we motivate as the relevant phase-inducing heads, coupled with localized convergence evaluation. If non-feature-driven movement can happen on the view of Last Resort sketched above (to avoid a local crash), then we might expect these movements to be tied to the "edges" (external domain) of phase-inducers.

[88] As noted above, Chomsky regards these properties as an apparent imperfection that we hope on minimalist grounds to show is "not real".

However I will now argue that even on these views we may still need something like M-features (or, as I suggest below, a different view of the categories present in verbal extended projection series).

Recall on Chomsky's view that the complement domain of phase-inducing heads spells-out *when the next such phase-inducer is introduced*. But on the assumption that C and *v* are the phase-inducers, this means that the complement domain of, for example, *v*, will not spell-out until the next phase-head (C). But then why should there be movement to the edge of *v*P, and not any other position between the phase-heads?

If we can show that movement is the edge of *v*P, and not higher, this suggests that the way to have Last Resort consistent movements would be to go with Chomsky's M-feature view. And if *this* is correct we also need to have movement target positions below the root. This may in fact be independently required, but the present point is that on Chomsky's view of phases the localized convergence motivation doesn't work to drive elements to the edge of *v*P. That is, it would not be strictly the *last* resort to move to such edges, since it is only after introducing further material that the spell-out of the complement domain will be required. A strict last resort view would be to have the movement only justified at the point in derivation where the problem arises, and as we have just seen this is *not* when a phase-inducing head is introduced.

To see the point, consider the following derivation (deploying a version of the reduced structures from Chapter One for convenience):

(174) a. D_{wh}
b. V — D_{wh}
c. *v* — V — D_{wh}

In (174) we have a derivation with *wh*-element as the object of V up to the point where *v* — a phase-inducer — is introduced into the derivation. At this point, there is no immediate threat that the *wh*-element will be stranded in this position. So unless we assume that the operation driving movement has some look-ahead capability, there is no reason under the localized Last Resort logic that would drive movement to the edge of *v*P.[89] The derivation continues, adding the subject-D to get the external-θ role from *v* (174)e, adding T (174)f, and moving D to T (174)g:

(174) d. *v* — V — D_{wh}

e. *v* — V — D_{wh}
|
D

f. T — *v* — V — D_{wh}
|
D

g. T — *v* — V — D_{wh}
| |
D D
▲ ⋯⋯

h. C — T — *v* — V — D_{wh}
| |
D D
▲ ⋯⋯

[89] One might object to this, pointing out that the 'lookahead' required would still be fairly local (only up to the next phase-head. But on Chomsky's view, if C and *v* are the only relevant phases this could be an unbounded stretch of structure (e.g., as he argues for successive cyclic raising environments, which are assumed not to introduce phase-distinctions). For the SCM in raising, as mentioned, Chomsky posits a EPP-features to drive the local movements, but there if there are no phase heads in such sequences then these motivations are decoupled. The point of the argument I am running in the text is that there is reason to think even where we do have phase-divisions that we would need M-features anyway.

The last step pictured here is the addition of the next alleged phase-inducer (C). Do we need spell-out of the complement domain to happen prior to this step? Yes, if there is to be movement to some position *below* C, otherwise its not obvious that the *wh*-element couldn't first move directly to C and have the spell-out operation apply afterwards. So this means that the local movement out of the complement domain of *v* must happen prior to the addition of C. The natural point in derivation to apply the Last Resort logic would be between steps (174)g and (174)h.

But then, *where* does the *wh*-element move? And why? It seems that moving to adjoin to the T-domain is just as plausible as moving to the edge of *v*P. (Though note movement to T would obey extension/root-merge, and movement to *v* would not — assuming this is not issue, both possibilities remain). However, the kinds of facts discussed in §2.1.5 strongly suggest that if there is very cyclic movement (within clauses) then it must involve a position *below* TP (see, e.g., (139)-(140)).

Note as well that we cannot apply the logic described above of moving to a "categorially likely checking configuration", since its completely unclear that *v* ever hosts the relevant CLPs.

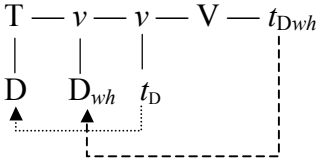
The conclusion is that if we need movement to the edge of *v*P we need an M-feature as Chomsky proposes. Or, we need to configure the general outlook such that derivations "know" that there is an upcoming second phase-inducing head. Chomsky has suggested that derivations may begin from limited inventories of elements — sub-arrays — which are understood to be selected from the lexicon in such a way that they can contain only one phase-head. This could potentially solve the problem if we can somehow restrict sub-arrays to only containing information that will end up *below* phase-

inducing heads — that is, material constituting the complement domain. Then we might motivate movement to the phase-edge of *v* (in the example above) in terms of a condition keyed to the introduction of a *second* sub-array containing another phase-inducing element. This would be conceptually similar to both Uriagereka's idea of having "derived terminals" — essentially treating phase-sized arrays as atomic element with respect to later stages of derivation. It would also be quite close then to a "mini-TAG" view, which posited phase-sized elementary tree domains.

But its unclear that this isn't simply a statement of the problem. Why couldn't the initial phase include everything up to, but not including the C-element? At issue here is how we partition structures into phases/spell-out domains. Having them restricted to containing only a single phase-head doesn't obviously constrain what *else* can be in an initial sub-array, so they may include material above *v* or not. If so, then the problem of distinguishing phase-edges from structure above such heads (but below the *next* phase head) still arises.

This situation is general. Suppose that the *wh*-element has somehow reached the *v*P edge (the double-*v*'s here can be taken to be an adjunction structure, e.g., *v*P — *v*P, or a maximal and intermediate level *v*P — *v*', it doesn't matter for present purposes).

(175)



The same situation described above arises here. The next step introduces a C element. The step after that presumably a selecting V, perhaps with a local object, and only after this we might encounter another *v* which by assumption forces the complement domain

of C to spell-out. So, again, we either need to the movement to not be strictly speaking consistent with last resort, or we need an M-feature associated with C.

We have reached then two main conclusions. First, the expelling of viral elements story, and any other like-approach seeking to drive cyclic movements out of independently evaluated domains based on convergence needs, needs some additional assumptions to get things to work. In particular, it seems that movement to the edges of phases cannot be motivated by a strict localizing of Last Resort. To the extent that we render the idea with respect to triggering a local/non-target movement consistent with a localized vision of economy (e.g., see Collins 1997), its not obvious how to discriminate between the phase edge and any other category that may be present below the next phase-inducer. Therefore, it seems that something like M-features are required *if* we have reason to think that movement is to *v*P and not to higher elements below the next phase-inducing head.

Second, its not clear how the idea of moving to a "categorially likely checking configuration" could drive *wh*-movement (or NP-movement) to anything but intermediate C (or T for NP-movement), at least not on standard assumptions about the structure of verbal extended projections (e.g., assuming roughly C-T-*v*-V).[90]

We saw in addition in our discussion in this Chapter that TAG approaches (at least on the general approach advocated in Frank 2002) also require reference to an EPP-

[90] However, as we will see in the next chapter, it has been argued (see Butler 2004, Belletti 200?, Pesetsky & Torrego 2004) that there may indeed be such elements below *v* but above V. If this is correct, then we could salvage the combination of expelling viral elements and moving to a likely checking configuration (this requires a version of the "split-VP" hypothesis (Koizumi 1993, 1995, Lasnik 1999a, Johnson 1991, Runner 1995). So, if verbal extended projections involve internal "mini-clauses" — C-T-*v*-C-T-V — or perhaps some equivalent, then we may have a road-in to a uniform story about cyclicity stateable in these terms.

type of property to motivate initial movements within elementary trees. Problems were seen to arise for the view of adjoining auxiliary structures at X'-nodes for SCM in of the *wh* and NP varieties if it was assumed that the initial movement involves what we have called Core Licensing Properties (CLPs; i.e. if the initial movement is the ultimately licensed one). Instead, EPP-properties are understood to drive local movements, and the licensing of the sort involving CLPs is regulated by cross-tree feature-checking built-in to the adjoining mechanism.

In the next chapter, I turn to the task of further developing the assumptions and mechanics of TCG and the WS/O-distinction on which it is based. The main argument is that TCG makes available an account of SCM that dispenses with the need to appeal to M-features. We will provide some fairly coarse-grained examinations of some of the SCM phenomena sketched in this chapter, providing enough of a story to show how the mechanics could be deployed in more detailed investigations in each domain.

CHAPTER 3: TCG Analysis

The structure of this chapter is as follows. We here deploy and further develop the TCG-approach with reference to some general sets of facts, concentrating mostly on points bearing on the architecture under discussion, rather than on analyses pursued to any great depth. We begin with a discussion of local relations which brings up some problems that went undiscussed in Chapter 1. This leads then to two discussions regarding SCM — one pertaining to A-movement, and one to A'-movement (concentrating on *wh*-relations).

We then speculate on a general extension of the logic deployed to SCM cases to other cases within local structures, opening the suggestion to pursue so-called 'split-VP' or "stacking"-style analyses to a perhaps extreme point of dividing individual thematic elements into their own little "mini-clauses", forming the clause internal equivalent of traditional-clause divisions. This is then shown to be relevant for cases left out by the traditional view of the clause assumed in the SCM discussion — that is, cases suggesting that at least *wh*-movement involves relations beyond clause edges, including as well some internal phrases. Then we stop, and move to concluding remarks, a discussion of some further architectural issues and open questions/problems.

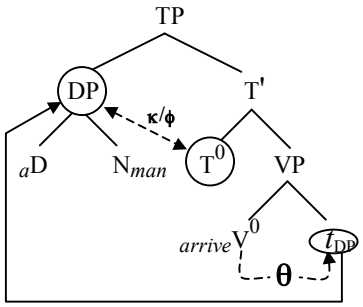
3.1. Local Relations: Part One

Consider the following simple case with an unaccusative verb in (176):

(176) A man arrived

A relatively standard view of the structure of (176) would look as follows (ignoring the C-domain):

(177)



Deploying the reduced structures and derivations discussed in Chapter One we would have (ignoring the internal assembly of the nominal):[91]

(178) a. $T^{\kappa:n}_{\phi:\varnothing}$ b. $T^{\kappa:n}_{\phi:\varnothing}$ — $D^{\kappa:\varnothing}_{\phi:f}$ c. $T^{\kappa:n}_{\phi:\varnothing}$ — $D^{\kappa:\varnothing}_{\phi:f}$ d. $T^{\kappa:n\Rightarrow\varnothing}_{\phi:\varnothing\Rightarrow\phi:f}$ — $D^{\kappa:\varnothing\Rightarrow\kappa:n}_{\phi:f}$ e. $T_{\phi:f}$ — $D^{\kappa:n}_{\phi:f}$

f. $T_{\phi:f}$ — V^{θ} — $D^{\kappa:n}_{\phi:f}$ g. $T_{\phi:f}$ — $V^{\theta[\phi:f]}$ — $D^{\kappa:n}_{\phi:f}$

The D and T elements associate and swap κ and ϕ as outlined earlier (§1.2), pictured in step (178)c. This process results in the deletion of κ on T — the outcome is a "discharge" of the κ-property of T, and a ϕ-relation between T and D. The addition of the thematic (v) element (*arrive*) leaves us with the structure in (178)f. And, as suggested earlier, the θ-property of V takes ϕ:f as its value, completing the A-relation circuit (closing off the open position of the thematic predicate). In this way D is indirectly connected (via the κ/ϕ feature-complex) to the internal role.

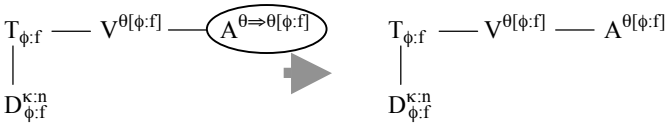
I mentioned in Chapter One as well that we will be viewing θ-variables as inherently undistinguished open slots, so that κ/ϕ properties are actually required to

[91] I assume that D-elements typically come without ϕ-specifications (i.e., with ϕ:∅), and that this property is valued in virtue of associating with N, which comes with ϕ:f (valued ϕ).

individuate them. We can expand on this point by considering a minimal addition to (176):

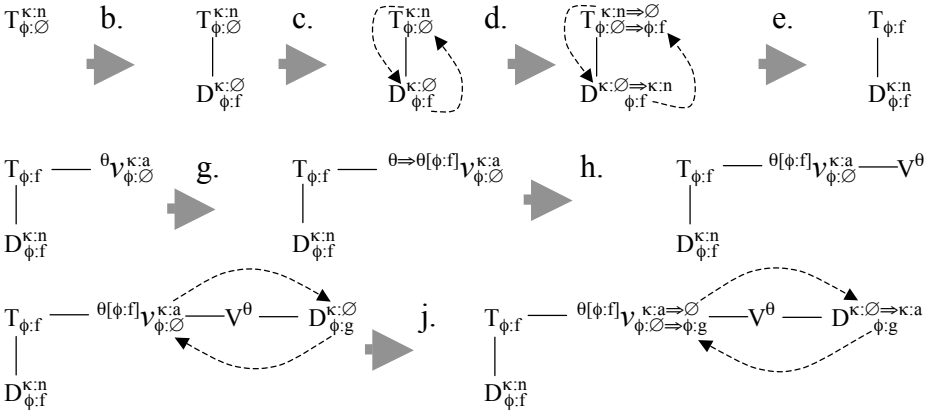
(179) A man arrived sad

Following the basic structure of proposals of Williams (1983, 1994) and others, I will take this to be a situation where non-distinct θ results in an identification in virtue of the ϕ-property of the unaccusative-θ comes to be borne by the adjoined adjectival, as follows:

(180) 

We will now see that on the rather loose view sketched in Chapter One regarding possible configurations for feature-relationships, even fairly small increases in complexity appears to present us with problems. Consider a simple transitive:

(181) He likes her

(182) a. 

Assuming that *v* introduces both the "external"-θ and accusative (κ:a), I will mark this element as "$^{\theta}v^{\kappa:a}_{\phi:\varnothing}$" to indicate that the θ-role is upwardly directed, and the κ/ϕ properties

downwards, as in (182)f. However, this talk of "upward/downward" raises the issue of whether there can be configurations where the θ-role dominates its argument in this approach. So far we have sketched a view under which A-relations, including the external-θ assignment, are mediated by κ/ϕ properties. So such mediation is suggested to be possible; but is it necessary?

Consider the steps beyond the introduction of *v* in (182)h-j. In (182)i it is indicated that the κ/ϕ-value swap happens independent of the θ-role taking its value. Is this the right way to think about this? What is at stake?

For one thing, since we have stated the mechanisms of feature valuation in a way that allows for probes to dominate goals or *vice-versa*, its not clear what prevents, e.g., the *v*-introduced θ-role from associating with its own κ/ϕ properties. Or, for that matter, having the nominative-T's ϕ-properties value the internal (ν-introduced) θ. It would seem, in order words, that we need to introduce some asymmetries (e.g., have *v*-θ only look "upwards"; or ν-θ not able to look past the *v*-introduced κ/ϕ-properties). In other words, we need some notion of locality here so that we don't get *he likes her* meaning *she likes him*, and other impossibilities.

Recall that we made a small fuss earlier in this discussion (Chapter One) about the possibility of redundancies between statements of locality built-in to rules or imposed on their outputs (e.g., minimality-type restrictions) and approaches which offered some characterization of domains, leaving the operations otherwise unrestricted. Here we are presented with a situation for which it isn't at all obvious that our conception of workspace restrictions (of the distinctness and ordering sorts) could be relevant. What recursive structure or repeated elements are there in such local domains for the

workspace to resist via the contraction mechanics? We need a story about the possible/impossible feature-connections in local domains. Importantly, if we have to introduce local notions regarding relative or absolute "structural distance" for operations to apply, it opens the question as to whether it would be best to treat everything that way (since it is required for the most local cases).[92]

I will suggest below, following some other developments, that the right position here may be to bite the bullet, and explore the possibility that there is a bit more structure in these local domains than meets the eye. The strategy here will be to work backwards from the more complicated cases (in particular linked-local relations of the SCM type) to the (superficially) simpler ones. It is in the domains where we see SCM type effects that our suggestions regarding workspace distinctness find their plausibility. The idea will then be to see whether we can find motivation for extending the ideas to seek out possible divisions within *local* structures that allow us to carry-over the central ideas about workspace-demarcated domains. So let us develop the analyses in more detail in these domains, and return to the local considerations.

3.2. Linked-Local Relations I: Raising to Subject (RtS)

Consider again the following standard case of cyclic A-movement in raising-to-subject (RtS):

(183) John [seems [_ to tend [_ to appear [_ to like carrots]]]]

[92] See Frank (2002) for some similar discussions regarding the possible need for locality of movement restrictions *within* elementary trees, and the redundancies such views would pose for the TAG architecture; basically the same issues arise there as here as one would expect.

First, the subject position of raising verbs is standardly thought to be athematic, as the presence of expletive elements and idiom chunks has traditionally been taken to suggest:

(184) a. There seems/is-likely/appears to be a man here
b. The shit seems/is-likely/appears to have hit the fan

I will assume then, as in our Chapter One sketch, that these predicates do not involve a small-*v*. I also adopt here the assumption that these infinitival complements are T's, and not C's.

3.2.1. Some Raising Impossibilities & Expletive/Associate Relations

Consider raising from a finite clause (185) and the ill-formedness of "superraising" in (186) where the subject moves past a position where it *could have* landed (were the position not otherwise filled):

(185) *John$_1$ seems [$_{CP}$ that _ is here]

(186) a. It seems John was told [_ to arrive on time]
b. *John seems it was told [_ to arrive on time]

The properties of these cases might be taken to follow from our notion of workspace distinctness, repeated here:

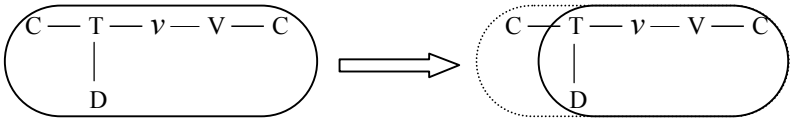
(187) **WORKSPACE DISTINCTNESS (ANTI-RECURSION):**
The workspace does not tolerate the presence of multiple tokens of type X

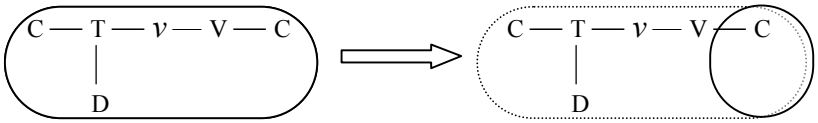
We provided in our earlier sketch of SCM a general story in terms of (187) plus the requirement on workspace ordering that highlighted one kind of response that the system might make to potential distinctness violations. The outcome was a process of node-

identification plus a contraction of the active workspace (to avoid ordering conflicts). Let us now consider a different situation:

(188) John believes [that the earth is flat]

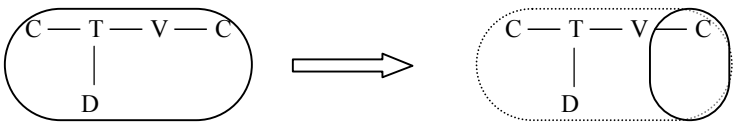
Here we will have, upon encountering the edge of the embedded clause (remember: derivations go top-down!) a potential distinctness violation between matrix and embedded C. At least two different responses to this situation are possible:

(189) 

(190) 

The response in (189) would be a conservative one in which the workspace would simply shift ("move-on") as we sketched in our introductory discussion. This would obey the restriction on distinctness as the higher instance of C would simply be abandoned to the output. The alternative is a more radical shift, essentially beginning an entirely new domain, as in (190).

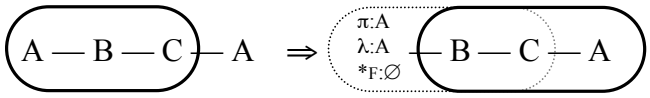
The more radical response would provide us with an explanation for the impossibility of (185) and (186)b. In both cases we would have derivations which expanded top-down as follows in (191), up the boundary of the embedded C-domains. If the more radical contraction occurs, then the subject would be stranded outside the workspace without having had its associated κ/ϕ properties connected with θ:

(191) 

Thus both raising from a finite clause and raising past a possible landing site would be subsumed under the same explanation. The result is a non-θ-marked subject in both cases:

(192) a. *[$_{CP}$ John seems [$_{CP}$ that ...
b. *[$_{CP}$ John seems [$_{CP}$ C^0 [$_{TP}$ it ...

This would thus be a different instance of sort of schema offered in Chapter 1, repeated here, regarding an unlicensed property being abandoned from the workspace to the output structure:

(193) 

However, it is important to note that on the view of κ/ϕ-θ connections being explored here, the violation is perhaps best viewed in these broader terms — a faulty A-relation — rather than just saying that an NP has not received a role. That is, there will be more than one way that an A-relation can be faulty — but the general story about feature relations on the dominance path will be seen to connect them all into the more general class.

That is, while it might be true that the violations above on par with (194)

(194) *John seems

we clearly do not want to say the same for the otherwise superficially similar violation in (since *there* is athematic):

(195) *There seems that a man is in the room

Given the top-down nature of structure expansion in the present system, the fairly strong intuition of the ill-formedness of (195) at the following point in a left-to-right pass can perhaps be taken to be relevant in this connection:

(196) *There seems that...
(vs. √ *It seems that* ...)

At the point where the finite embedded clause is encountered, the judgments are fairly uniform across (185), (186)b, and (196). But (196) involves athematic *there*, suggesting that the detection of anomaly at the clause border in (185) and (186)b, which has a rather similar profile, ought not be conceived of in exclusively θ-theoretic terms. Consider also:

(197) a. *What seemed that Mary liked _ ?
b. *What seemed that Mary liked pizza?

Now, both (197)a&b are clearly out, but they seem to have the same profile despite being rather different sorts of violations on standard views. In particular, (197)a would be presumably be a Case-theoretic violation, since *what* occupies two Case positions, whereas (197)b would involve full satisfaction of κ/ϕ, but the *wh*-element would receive no θ-role. Recall our schema for a local A-relation (e.g., a κ-marked "subject" indexed to the external/*v*-introduced θ by the ϕ-properties):

(198) $C—T^{\phi:f}—v^{\theta[\phi:f]}$
$\quad\quad |$
$\quad D^{\phi:f}_{\kappa:n}$

What the ill-formed examples involving non-expletives discussed above have in common is the following stage of derivation (just prior to contraction), where V is a raising predicate (an athematic element compared to (198)):

(199) $\text{C—T}^{\phi:f}\text{—V—C}$

$$\begin{array}{l} \text{C—T}^{\phi:f}\text{—V—C} \\ \quad\;\; | \\ \quad\; \text{D}^{\phi:f}_{\kappa:n} \end{array}$$

If (C, C) satisfy the matching relation, and result in a contraction, then it is true that $\text{D}^{\phi:f}_{\kappa:n}$ in (199) will not be θ-associated, but that does not explain the otherwise very similar feel to the violation involving *there* in (195), which on most views of these elements is *athematic* (though see Moro 1997 for a view in which *there*, if not "thematic" is at least viewed as an abstract sort of predicate).

Bošković (1997, 2002) discusses what he calls (following a suggestion of Howard Lasnik) the Inverse Case Filter. The traditional Case filter was stated in terms like the following:

(200) ***Extended Case Filter***:
*[$_{NP}$ α] if α has no Case and α contains a phonetic matrix or is a variable
(Chomsky 1981:175)

Case-theoretic violations on this view are pinned on a failure to meet a requirement of NPs. However, it is not unreasonable to suggest that Case-theoretic violations might be (or might *also* be) a matter of Case-assigner's needing to "discharge" their κ-property. This is the idea of the "Inverse" Case Filter. Bošković discusses examples such as the following:

(201) a. * is likely John will leave
b. *John believes to seem Peter is ill

Both of these kinds of violations have been discussed in terms of the "EPP", currently implemented in feature-licensing terms within minimalism. However Bošković points out that the cases in (201) can be explained in terms of failure to discharge/assign Case

(nominative in (201)a and 'exceptional' accusative in (201)b). Can we make use of this general idea in handling the crucial facts regarding expletive-*there* in a way that connects them to what otherwise (on standard views) appear to be θ-theoretic violations?

We could assume (with Chomsky and others) that *there* is minimally specified for agreement, but cannot check case. Then we would understand the ill-formed *there...that* case above as a matter of the Inverse Case Filter. However, another view is possible, which I believe can play a role in determining the distribution of *there*-type expletives (at least in English).

Suppose that *matching* of a feature, any feature, is sufficient to license combination, so that X—Y can be established as a dominance link if they bear the same feature F. Suppose that expletive *there*, unlike regular "thematic" nominal expressions, bears just unvalued κ, and no ϕ-specification whatsoever. Then we would have the following:

(202) $C\text{—}T^{\phi:\varnothing}_{\kappa:n}\text{—}V\text{—}C$
$\quad\quad\;\;|$
$\quad {}_{there}D_{\kappa:\varnothing}$

Now two options present themselves — either κ-properties can enter into licensing/valuation independently of ϕ, or (as Chomsky 1998, 1999) suggests, κ-licensing is parasitic in some sense on ϕ-relationships. Suppose that κ-valuation/licensing can happen independently. Then we have:

(203) $C\text{—}T^{\phi:\varnothing}_{\kappa:n\Rightarrow\varnothing}\text{—}V\text{—}C \quad\Longrightarrow\quad C\text{—}T^{\phi:\varnothing}\text{—}V\text{—}C$
$\quad\quad\;\;|\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad|$
$\quad {}_{there}D_{\kappa:\varnothing\Rightarrow\kappa:n}\quad\quad\quad\quad {}_{there}D_{\kappa:n}$

This view could help us with the distribution of expletive-*there* as follows. Suppose in the well-formed local A-relation, contra to what we have suggested so far, that thematic elements (*v*/V) come to the derivation with an unvalued ϕ-feature as their argument. I suggested earlier that we might in general view κ/ϕ-properties as individuating the variable positions of thematic predicates, and two ways of thinking about this were offered: (i) the open positions are undifferentiated "slots" and (ii) the one open position is indistinguishable from another because they all "start" with a general default value. Suppose then that they start as θ[ϕ:∅]. An expletive element in the subject position of a transitive verb in English then will encounter the following stage of derivation if the sketch in (203) is correct:

(204) $C—T^{\phi:\varnothing}—v^{\theta[\phi:\varnothing]}$

$\quad\quad\quad\quad |$

$\quad\quad {}_{there}D_{\kappa:n}$

(e.g., **there hit* ...)

Assuming that subsumption (see (31), §1.2) is required to hold for valuation, ϕ:∅-ϕ:∅ will make θ-discharge as we have envisioned it impossible. The prediction is that expletive-constructions must require some other way to get T-ϕ valued. How could that happen? It must be, according to this view, that T-ϕ relates directly to a subordinate nominal element, something that is independently valued for ϕ. And that nominal must be in a configuration that is somehow appropriately thematic, but without involving κ-assignment. Moreover, there cannot be any intervening θ[ϕ:∅], as this could be seen to block the relevant relationship between T-ϕ and some lower nominal "associate".

This picture seems roughly correct. Expletives generally appear in κ-positions that are not immediately/locally associated with θ, as in raising, copular constructions, and unaccusatives. They can appear as well in ECM environments in English, so long as the condition on there being no intervening θ-element is met, e.g.,:

(205) a. I believe there to be a man in the room
b. I believe there to have arrived a man
c. I believe there to appear to be a man in the room
d. *I believe there to have a man left
e. *I believe there to be an idiot (vs. *I believe John to be an idiot*)

Let us consider a counterpart to (205) to take look at how these relations are established — the idea then is that *there* will serve to license matrix κ, but not ϕ. The result is that ϕ must be valued in another way, and it *cannot* connect directly with a θ-element (because the ϕ-argument of θ will be ϕ:∅ as well, and so the subsumption condition on valuation will not be met). Consider then the classic case:

(206) There seems to be a man in the room

We begin then with a local structure as discussed above:

(207) a. $C—T^{\phi:\varnothing}—V$ b. $C—T^{\phi:\varnothing}—V—T$

$\qquad |$

$\quad {}_{there}D_{\kappa:n}$ $\qquad {}_{there}D_{\kappa:n}$

The defective T complement to the raising predicate V is introduced, resulting in contraction, leaving us with the following workspace (ignoring the output structure here):

(208) a. $C—T^{\phi:\varnothing}$ b. $C—T^{\phi:\varnothing}—V_{be}—D^{\phi:\varnothing}_{\kappa:\varnothing}$

$\qquad |$

$\quad {}_{there}D_{\kappa:n}$ $\qquad {}_{there}D_{\kappa:n}$

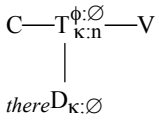
c. $C—T^{\phi:\varnothing}—V_{be}—D^{\phi:\varnothing}_{\kappa:\varnothing}—N^{\phi:f}_{\kappa:\varnothing}$ d. $C—T^{\phi:\varnothing\Rightarrow\phi:f}—V_{be}—D^{\phi:\varnothing\Rightarrow\phi:f}_{\kappa:\varnothing}—N^{\phi:f}_{\kappa:\varnothing}$ (← ϕ-agree)

$\qquad |$

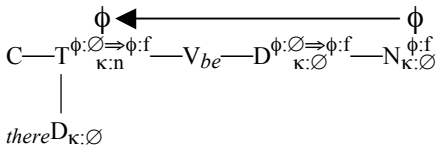
$\quad {}_{there}D_{\kappa:n}$ $\qquad {}_{there}D_{\kappa:n}$

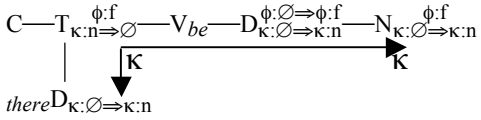
In this situation we could understand the nominal element introduced (*man* in *a man*) to value the ϕ-properties that are unvalued prior to (208)d, but then what of the κ-properties of the nominal associate? Given our adoption of a single dominance order, *there* bears no direct relationship to the associate.

We can solve this problem by taking the other option suggested above (following Chomsky) and suggest that κ and ϕ valuation are linked. We keep the idea that *there* has a lone unvalued κ-property. This will explain why local κ/ϕ-valuation cannot happen. This will leave us with an initial stage of derivation more like this:

(209) 

Now a different question arises, namely: when a θ-element is introduced why can't the valued κ-property of T "fill-in" the value of θ as we suggested for *wh*-questions? I will return to this in a moment. First, consider how this will implement a familiar transmission type story regarding expletive-associate relationships. If T and expletive-*there* are able to relate via matching (Tκ-Dκ), but with valuation of the properties impossible because *there* bears no ϕ-property, then the final stage of derivation above in (208)d would look as follows in (208)d', with κ-valuation happening then parasitically on successful ϕ-agreement as in (208)e:

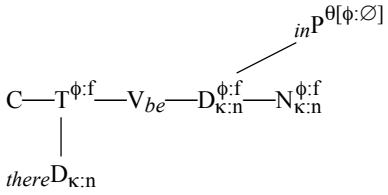
(208) d'. 

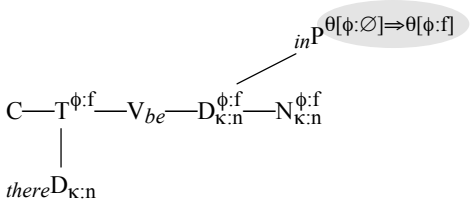
e. 

Thus, T-κ values both the expletive and the associate, deleting as in the regular case. We have then a mechanism whereby expletive-*there* is a κ-marked element, but so is the associate. We thus agree with the assessment of Lasnik (1995) that the expletive checks the matrix Case property, but do not posit an additional *partitive* case (e.g., as in both Lasnik 1995, and Belletti 1988) that is responsible for licensing the associate. Rather, the associate will be able to enter, in virtue of its ϕ-properties, into 'caseless' types of predication, as in small clause structures for example:[93]

(210) a. I consider [John a genius]
b. I consider [John intelligent]
c. ...[[a man] [in the room]]

For the present point regarding expletive/associate relationships, I will assume that the structure of the relationship between [a man] and [in the room] comes out as follows for the continuation of our (208) derivation in (208)f:

(208) f. 

g. 

[93] I mean by 'caseless' here just that case-assignment is not part-and-parcel of the local relation in the sense that Case must be assigned from outside the basic predication.

The D-element is κ-marked, and so it may enter into a relation with an element which does not, itself carry a potentially interfering κ-property. The assumption then is that there is a class of θ-like relationships which are "direct" in one sense — they involve ϕ-ϕ connections (valuation) that is not accompanies by κ-valuation — but "indirect" in another sense, in particular any nominal element entering into such relationships will have to have found its κ-properties licensed in some independent κ/ϕ complex. I will not, in the present work, get into the issues regarding definiteness effects and the like, though the suggestion here would be that direct ϕ-ϕ relationships that are reliant on some other instance of κ-assignment are at least one kind of DE environment. The present outline seems compatible with one or another version of Deising-style mapping, but I won't pursue this here (e.g., see Deising (1992); and see Hornstein (1995) for some relevant discussion, and see Safir (1987) for what strikes me as a related conception involving "transmission" and the notion of "unbalanced" chains).

So, let us now consider our ill-formed cases together from above:

(211) *John$_1$ seems [$_{CP}$ that _ is here]

(212) a. It seems John was told [_ to arrive on time]

b. *John seems it was told [_ to arrive on time]

(213) *There seems that a man is in the room

These would thus correspond to the following two scenarios, where both (211) & (212) manifest a reasonable A-relation as in (214), but one which does not connect with θ, thus

leaving the matrix subject *John* unintegrated. In the case of (213) the subject-related T element (along the the expletive) must be spelled-out with unlicensed properties:

(214) $\mathrm{C}—\mathrm{T}^{\phi:f}—\mathrm{V}—\mathrm{C}$
$\qquad |$
$\quad {}_{John}\mathrm{D}^{\phi:f}_{\kappa:n}$

(215) $\mathrm{C}—\mathrm{T}^{\phi:\varnothing}_{\kappa:n}—\mathrm{V}—\mathrm{C}$
$\qquad |$
$\quad {}_{there}\mathrm{D}_{\kappa:\varnothing}$

In (215), *there* can enter the structure in virtue of κ-matching, but given the assumption that κ-valuation is parasitic on ϕ-relationships, nothing further can happen at this point. Assuming that properties which are unvalued are illegible at the interface, (215) will crash as the material intervening between C—...—C is spliced-out.

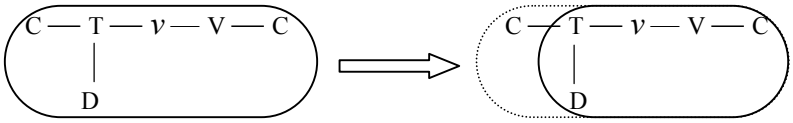
So we diagnose two kinds of A-relation deficiency, and have in hand a reasonable story that looks to be able to contribute to understanding expletive-there's distribution, based on a particular implementation of the "transmission" logic (Safir 1982, Chomsky 1986). Also we have implemented a view of κ/ϕ, with suggestive connections to GB-era views. Consider again the Extended Case Filter mentioned above:

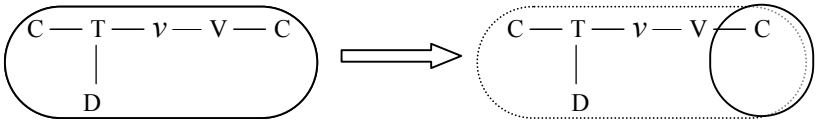
(216) ***Extended Case Filter***:
*$[_{NP}\ \alpha]$ if α has no Case and α contains a phonetic matrix or is a variable
(Chomsky 1981:175)

On the view here this is only part of a more general conception of A-relations (and A'-relations as we'll see below). Why should Case be hooked up with the notion of "variables"? The idea here is that κ-properties *name* variables, thereby distinguishing them, and that the interconnections of κ/ϕ-properties are what mediates the connections to between thematic predicates and nominal expressions, whether operator-like/

quantificational or not (as with ordinary NPs). Vermeulen (1995) (see also Visser & Vermeulen (1996)) point out that we can in general distinguish three things regarding variables, (i) the "variable itself", (ii) the *name* of the variable, and (iii) the value of the variable (what it ranges over). The "variable itself", I am suggesting is the open position in θ-elements for the participants that these elements relate to eventualities. ϕ-properties create a path to a κ-feature, which either is bound from above (as in WH-cases) or is assigned to an overt nominal.

Before moving on let us return to an issue we left dangling above. Recall the possibility of having the less radical workspace contraction (189) raises again the issue of locality as potentially independent of these dynamic domains, as discussed above in connection with the derivations for simple transitive structures. I repeat the possible workspace contractions here for convenience:

(217) 

(218) 

Its not clear without imposing some other locality mechanism why the conservative response to a distinctness violation (217) wouldn't then simply allow a some featural-relation with the embedded T-element.[94]

[94] Note that the kind of node-identification suggested to underlie SCM presumably couldn't apply between finite T's, since these would not manifest the subsumption relationship that we have posited as being the relevant condition under which node identification takes place.

Note that the conservative response to a potential distinctness violation would always end up with there being (globally) more instances of contraction/spell-out than the radical response. Assuming that transderivational comparisons of more-vs-less numbers of contractions is an undesirable property to have in a minimalist approach, note that there is a local way to ensure that the global number of contractions will in fact be minimized. This will always hold if local contractions are maximal — that is, if the system responds in the "radical" manner depicted above (which yields the A-movement locality facts discussed).

Note also that this issue of the difference between the radical vs. conservative response to distinctness violations does not arise when the relevant context nodes (e.g., C or T) are identified as in A- or A'-type SCM (since the relevant nodes are identified there is no way to remove "just one of them", leaving the other in the workspace — so the only possible response following identification is the radical contraction in order to keep the ordering properties coherent, as sketched in Chapter One).

However, if we adopt a local economy view for the non-movement case (maximize contractions) to get an account of the A-movement locality facts above, we would then have two separate motivations for what otherwise seem to be rather similar sorts of processes, differing only with respect to whether the relevant nodes are identified or not prior to contraction. My suspicion is that there may be a way to derive these contractions in a unified way, but I have not yet found a satisfactory formulation to this effect. I will leave the matter open here, assuming the following:

(219) **Workspace Economy**: Contraction/Spell-Out is locally maximal

Note that this is actually consistent with a "least effort" line of thinking, perhaps despite appearances. The idea would be that maintaining distinctions in the workspace is what takes effort, so whenever this burden can be eased (by spelling-out) it is *maximally* eased.

3.2.2. Binding Interventions & SCM

Consider now some of the SCM-type effects regarding binding, in particular the cases in (136) repeated here as (220):

(220) a. John$_1$ seemed to himself$_1$ to appear to Mary to be getting fat
b. John$_1$ seemed to appear to himself$_1$ to be getting fat
c. John$_1$ seemed to Mary to appear to himself$_1$ to be getting fat
d. *Mary seemed to John$_1$ to appear to himself$_1$ to be getting fat
e. It seemed to John$_1$ to appear to himself$_1$ that he was getting fat
f. John$_1$ seemed to Bill$_2$ to appear to himself$_{1/*2}$ to be getting fat

We noted the following two key points about these cases in our Chapter 2 discussion. First, the impossibility of (220)d was suggested to be traceable to an intervention effect on a cyclic raising story, where we would understand *Mary* to occupy the embedded non-finite subject position of *to appear*, thus constituting a closer possible binder for the *self*-form in the experiencer-PP. But this creates a ϕ-conflict, and so it is out.

The second point was to note that, regarding (220)e, it appears that *John* is a suitable binder for the *self*-form, despite the apparent lack of a c-command relationship. We can now expand on these observations as follows.

Recall we noted as well in our earlier discussion that (220)a and (220)b suggest that either the binding domain for the *self*-form includes more than one clause (perhaps specified in terms of the presence of an subject or suitably "subject-like" element, as in some approaches to binding), or some relation must be involved to bring the antecedent-

dependent pair into a more local relationship. The two salient possibilities for this latter sort of solution are the kind of T-to-T-domain SCM of the subject *John* in these examples, or perhaps an LF-type movement of the *self*-form.

Note that a strict clause-mate condition on these relationships is implausible if we assume that the experiencer-PPs are not implicated in any movement between domains (i.e., assuming their positions are fixed where they surface). Consider the following additions to the examples in (220) of one more intervening raising predicate (*tend*, in (221)) or two more (*tend* and *be likely*, (222)):

(221) a. John$_1$ seemed to himself$_1$ **to tend** to appear to Mary to be getting fat
b. John$_1$ seemed **to tend** to appear to himself$_1$ to be getting fat
c. John$_1$ seemed to Mary **to tend** to appear to himself$_1$ to be getting fat
d. *Mary seemed to John$_1$ **to tend** to appear to himself$_1$ to be getting fat
e. It seemed to John$_1$ **to tend** to appear to himself$_1$ that he was getting fat

(222) a. John$_1$ seemed to himself$_1$ **to tend to be likely** to appear to Mary to be getting fat
b. John$_1$ seemed **to tend to be likely** to appear to himself$_1$ to be getting fat
c. John$_1$ seemed to Mary **to tend to be likely** to appear to himself$_1$ to be getting fat
d. *Mary seemed to John$_1$ **to tend to be likely** to appear to himself$_1$ to be getting fat
e. It seemed to John$_1$ **to tend to be likely** to appear to himself$_1$ that he was getting fat

Importantly, the judgments remain the same as in (220). The crucial cases are those involving binding between two elements situated within these experiencer-PPs — that is, (221)e and (222)e. What these (e)-cases show is that binding is independently possible between the nominals in these PPs, and that the phenomena is not sensitive to the boundaries introduced by intervening embedded non-finite clauses. If the position of these PPs is "fixed", then binding of these *self*-forms cannot be required to happen within a single clause. Similarly, if movement from these positions is generally not possible, the idea of having the *self*-form move into a more local relation with its antecedent is

implausible as well. This leaves the possibility of defining the binding domains in terms of something like the local presence of a "subject" (so *if* there is no local A-movements we could still have arbitrarily large binding domains in this sense).

The question then is why on such a view of binding domains would we have the contrast between the d-/e-cases in (220)-(222)? It would seem that *John* can be a local binder — that's what the e-cases show.

Moreover, as the following show, many kinds of dependencies that are typically understood to require a c-command relationship appear to be licit between two such PP structures, providing further strength to the assertion that there is no complicating factor of structure in the d-cases in (220)-(222), and that it thus stands as a piece of evidence that appears to demand that something like successive-cyclic A-movement occurs. Consider (from Castillo, Drury, & Grohmann 1999:95):

(223) a. It seems to every boy_1 to appear to his_1 mother that the earth is flat
b. It seems to no man to appear to any woman that the earth is flat
c. *It seems to him_1 to appear to $John_1$ that the earth is flat
d. It seems to his_1 mother to appear to $John_1$ that the earth is flat
e. ?It seems to $John_1$ to appear to $himself_1$ that the earth is flat
f. It seems to $John_1$ to appear to him_1 that the earth is flat
g. It seems to $them_1$ to appear to $each\ other_1$ that the earth is flat

Thus, variable binding of a pronoun by a quantifier ((223)), negative polarity licensing ((223)b), and Condition C violations ((223)c) as well as their expected absence in (223)d all point to the generalization that these experiencer elements can bind (etc.) out of their PPs.

Curiously, there is an unexpected absence of strong complementarity between Conditions A and B, as the acceptable judgment for (223)f shows in comparison to

(223)e. That is, (223)f is not degraded with the indicated coreference as is the following case in (224)a with respect to the well-formed (224)b:

(224) a. * John$_1$ is believed to seem to him$_1$ to be a genius
b. √ John$_1$ is believed to seem to himself$_1$ to be a genius

Some speaker in fact detect a slight advantage in acceptability for the pronoun case versus the *self*-form in case of binding between elements in experiencer-PPs, finding (223)f slightly better than (223)e. But the rest of judgments are stable, including the possibility of reciprocal binding as in (223)g.

The presence of the strong contrast in (224)a/b and its absence in (223)e/f led Castillo, Drury, & Grohmann (1999) to suggest that *self*-form in these cases is actually a *logophor* (Reinhart & Reuland 1993, Sells 1987), and that, if true, this fact would undermine the argument for successive-cyclicity of A-movement based on the intervention effect in (223)d (due initially to Danny Fox, as pointed out by David Pesetsky (p.c.)). The argument is essentially this: since we do not know what governs the distribution of logophors, it is unclear that we need to posit an intermediate copy/trace of A-movement to explain (220)d, repeated here:

(220) d. *Mary seemed to John$_1$ to appear to himself$_1$ to be getting fat

What should account for speaker judgments regarding (220)d should thus be some to-be-specified story about logophoricity.

But this argument from Castillo *et al.* does not go through — the case for successive A-movement made by such observations, I think, still stands. While it is true that the non-complementary distribution of pronouns and reflexives in these experiencer-

PP environments suggests that something like logophoricity is in play here (e.g., as in "picture-NPs", see below), this observation says nothing about what still appears to be an intervention-type effect in the contrast between (220)c and (220)d above. That is, *whatever* logophoricity ultimately amounts to (connected to various matters concerning "point-of-view", "psychological state", and similar notions; see below),[95] it still seems to be sensitive in some manner to structural factors in the determination of *im*possible antecedence relationships.[96]

The interest of logophors stems from their extra distributional freedoms in comparison to ordinary *self*-anaphors. For example, in comparison with the complementarity of the straightforward local cases of pronouns/reflexives (as in (225)a vs. b), we find the lack of a strong contrast for the a/b cases in (226) & (227) surprising, as we do with the antecedence between the PP-embedded experiencer elements in (223)e vs. (223)f (repeated here as (228)a vs. (228)b):

(225) a. * John$_1$ liked him$_1$
b. √ John$_1$ liked himself$_1$

(226) a. √ John$_1$ liked the pictures of him$_1$
b. √ John$_1$ liked the pictures of himself$_1$

(227) a. √ John$_1$ thought pictures of him$_1$ were on display
b. √ John$_1$ thought pictures of himself$_1$ were on display

[95] Note Boeckx (2001) contains an interesting discussion drawing on Rooryck (19??) who argues for a view of certain raising predicates (e.g., *seem*, *appear*) connecting to verbs of comparison, and thus indirectly to concerns relating to "point-of-view". For discussions on the notion of logophoricity, see Clements (1975), Sells (1987), Reinhart & Reuland (1993), Williams (1994), Reuland & Everaert (2001)). See below for some further discussion of logophoricity and why it probably isn't in play in the present case.

[96] Castillo et al. do note various "structural" factors that seem to be involved in restricting logophor interpretation, suggesting a preference hierarchy for c-commanding vs. m-commanding vs. merely "previously established in the discourse" elements as potential antecedents, but shy away from the problematic conclusions that I reach here regarding the argument these cases still present for cyclic A-movement.

(228) a. ?It seems to John$_1$ to appear to himself$_1$ that the earth is flat
b. It seems to John$_1$ to appear to him$_1$ that the earth is flat

But all this is to notice an extra dimension to the distribution of *self*-elements — something about these contexts allows something additional possibilities where induction from the basis of the strictly local cases suggests it ought to be out.

Some speakers, as mentioned above, find there to be a slight advantage in acceptability for the pronoun in (228)b as compared to the *self*-form in (228)a. However, this difference rather like the contrast between (229)b&c where there is usually a slight favoring of the pronoun over the *self*-form.

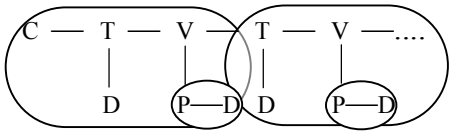
(229) a. *Mary sold the pictures of himself
b. ??John$_1$ thought Mary sold the pictures of himself$_1$
c. John$_1$ thought Mary sold the pictures of him$_1$

But the crucial Fox-cases have the unacceptability status more in-line with (229)a. The conclusion is that negative restrictions on these *self*-elements are clearly in force, whatever their extra distributional freedoms. An explanation of this fact is straightforward on the SCM view of raising.

The conclusion can be avoided only if a story about the distribution of logophoric elements could be produced which would rule out by other independently motivated means the cases that can otherwise be handled as a straightforward intervention fact under the cyclic movement analysis.

But what is the domain for the binding theory on our restricted workspace view? It seems clear that binding of these *self*-elements is *not* something which occurs within the boundaries of the workspace as we have set things out here. Setting aside the picture-NP situation (I won't be discussing the status of recursion in NPs here), however we view

the integration of the experiencer-PPs with respect to raising predicates the mechanics of contraction will always set them off into separate workspaces. Consider:

(230) 

Assuming that these PPs are somehow V-associated canonically, T-T identification and contraction would result in the structures being in separate domains.

I do not have anything to say about how it is that elements (e.g. in (230)) can relate to each other from within their containing PPs, either in the binding of the *self*-form (228)a or in any of the other relations that seem to be possible between these positions (223). Whatever factors underlie the possibility of such relationships, however, what seems clear is that the possibility that the SCM analysis makes available — of positing an intervener — directly explains the sharp anomaly of (220)d.

It is also clear that the general patterning of the availability of these *self*-forms include positions that we will certainly want to say are in distinct phases of derivation (separate workspaces), like the binding of these forms by a matrix element when they are in embedded subject positions as in (227)b.

Recall from our discussion in Chapter One the suggestion that we think of the WS/O-distinction as essentially drawing the *interface* line, such that output would be conceived as a syntactic structure populated by only PF and LF relevant properties. On the local view of domains being entertained here, the distribution of logophors *must* be a matter of relationships established over the output structure. Thus I tentatively suggest here that we regard the logophoric *self*-forms under discussion as distinct from local

reflexives in these terms. But note that our view of the output involves a general maintenance of the ordering properties created by construction in the workspace, so we still have "structural" distinctions that can be made over the output. Let us assume then that the connection between a logophor and an antecedent element is captured over the output structure with a Higginbotham's (1985) linking mechanism, though we will take this linking to be contingent on matching ϕ-properties of the elements.

This will be understood to be different from the matching and valuation that we have so far discussed. I will in fact suggest below that local reflexivization is a process involving workspace-local valuation of ϕ; logophors, however, will be understood to be independently ϕ-specified, and either they match up with their independently ϕ-specified antecedents, or not, under linking. That is:

(231) a. **Local Reflexives**: ϕ:f—θ ... ϕ:∅ — θ
MATCHING/VALUATION

b. **Logophoric *-self***: ϕ:f—θ ... ϕ:f—θ
LINKING

This requires that we view ϕ-properties as "there" in the output structure, and not just the narrow syntax (workspace). But we have been presupposing this in the general outlook on these properties anyway in terms of their assumed role in mediating θ-discharge. The linking relation, following Higginbotham, runs from a referentially dependent element to a referential one. Higginbotham's view assumed that such links get created in two ways, as a reflex of movement, and independent of movement. (e.g., a variable bound by a quantifier would enter into such a linking relationship, though no one, as far as I am

aware, has ever suggested that quantifier/pronoun relationships are of the movement sort).

So what are the structural conditions on logophoric linking? One general answer that has been offered in the literature is that, essentially, *there are no such conditions*. This was the basis for Castillo *et al*.'s (1999) rejection of the alleged binding intervention case in (220)d as an argument for SCM in raising. Rather, conditions on logophoricity are understood to rely on things like the following (see Sells 1987:445, and Williams 1994:86):

(232) Logophors connect with a *logophoric center*, which is an NP that must be a "thinker", "perceiver" meeting one of a-c:

a. The referent of the NP is "the source of the report"
b. The referent of the NP is "the person with respect to whose consciousness the report is made"
c. The referent of the NP is "the person from whose point of view the report is made"

Note that no mention of structure is made. In fact, *self*-forms of this kind can appear without any structurally present antecedent at all:

(233) As for myself, Paris is great this time of year
(i.e., "as for me/my-point-of-view, (I think),...X")

Let us consider (220)c-e again, to be sure that these notions regarding logophoricity might not, after all, be put to work in explaining the central cases that I am taking to provide evidence for SCM in raising:

(220) c. $John_1$ seemed to Mary to appear to $himself_1$ to be getting fat
d. *Mary seemed to $John_1$ to appear to $himself_1$ to be getting fat
e. It seemed to $John_1$ to appear to $himself_1$ that he was getting fat

It is not at all clear that these notions make the right predictions. In (220)d, for example, the *seeming* and *appearing* are both to *John*. So it would seem that that if there is a candidate logophoric center for (220)d, it is *John* and not *Mary* (i.e., *Mary* is the one "seeming"/"appearing" to be such-and-such, not the one that such-and-such is "seeming/appearing-*to*). The point-of-view criteria should pick out the experiencer, not the matrix subject, as the logophoric center.

But this then suggests that there are, after all, some structural conditions or other on these elements, in the sense suggested above — logophors are indeed keyed to extra-syntactic factors that influence where they may find there antecedents (including implicit arguments, or merely presupposed entities in the discourse), but in the right local environments with a referential NP, their connection to that NP appears to be mandatory. If this is right, then we really do have a good argument for SCM in raising. I will return to these issue below, as they bear on the issue of similar arguments as they arise in *wh*-movement.

What we have seen then is that (i) binding domains in terms of an accessible subject or the like don't seem to work properly for explaining this particular range of facts, (ii) movement of the *self*-form is also somewhat implausible, since the elements within dative-experiencers typically cannot undergo movement, and (iii) a possible explanation in terms of logophoricity doesn't seem to be able to make the right distinctions either.

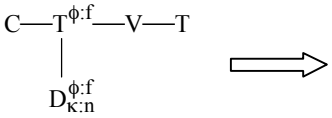
In addition, although we didn't make a big fuss about it above, the following cases involving reciprocals don't obviously fall into the class of possible logophors, but nonetheless show all the same effects as the *self*-forms:

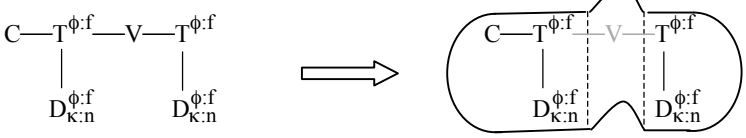
(234) a. The boys$_1$ seemed to Mary to appear to each other$_1$ to be getting fat
d. *Mary seemed to the boys$_1$ to appear to each other$_1$ to be getting fat
e. It seemed to boys$_1$ to appear to each other$_1$ that they were getting fat

So in absence of some other story to explain these facts, we need SCM. The question now is *what motivates movement to the intermediate positions*?

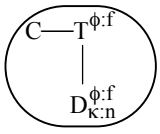
I argued in Chapter 2 that we should be suspicious of an "M-feature" solution (e.g., so-called intermediate EPP-features postulated at the edges of embedded non-finite clauses). Other approaches that can handle these facts do so by brute force stipulation that A-movement moves T-to-T (e.g., Bošković 2002, Grohmann 2003).

However we have in our development of the SCM mechanics in TCG an account which relies on independently required notions of (i) formal ordering properties, (ii) and a system of types. With these ingredients we stated our conditions on our workspace, and these can be understood to drive the intermediate movements without appeal to M-features. Recall from Chapter One the schema in (29) for SCM in raising environments (I repeat the relevant portion of derivation here:

(29) e. 

(29) f. 

And recall that this contracted workspace on the right-hand side of (29)f is really just:

(29) f'. 

So without M-features we are able to derive the binding patterns above. However, it also seems that views which posit movement to the edge of every phrase would do just as well with these facts (e.g., see Takahashi 1994, and a recent revival of Takahashi's view in Boeckx 2003). Such super-cyclic views would also seem to do quite well regarding the distribution of "floated" *all* seen in English RtS:

(235) a. The men **all** seemed to appear to be likely to leave
b. The men seemed **all** to appear to be likely to leave
c. The men seemed to **all** appear to be likely to leave
d. The men seemed to appear **all** to be likely to leave
e. The men seemed to appear to **all** be likely to leave
f. The men seemed to appear to be **all** likely to leave
g. The men seemed to appear to be likely **all** to leave
h. The men seemed to appear to be likely to **all** leave

Moreover, these facts would perhaps be puzzling on the TCG view offered here, since SCM (the "lowering" effected by node-identification) is predicted to only involve the equivalent of Spec-TP positions. Therefore (235)c, e, f, and h are all problematic.

I am not going to pursue these matters here. I don't fully understand at present the wider array of facts — a thorough recent review of the relevant theoretical and empirical issues surrounding such "floated" elements (Bobaljik 2002) urges a kind of caution that time and space limitations do not allow me to respect here. I will note only that there is reason to doubt a "stranding" analysis in general and that at present it seems to me that base-generation analyses have the best empirical coverage, so its not entirely clear that these facts bear directly on SCM.

Consider a few examples that bring up the kind of problems that arise (the following are drawn from Bobaljik 1995). Note that the stranding analysis seems to

presuppose that elements like *all* make a well-formed unit with the DP that can strand them. But this isn't general, consider:

(236) a. Some of the students might *all* have left
b. *All (of) some of the students might have left

Although (236)a seems fine, it cannot surface with *all* as a unit, as ill-formedness of the b-case shows. Another classic case that has been brought up as a challenge to the stranding analysis that are relevant to our A-movement discussion are unaccusatives and passives:

(237) a. *The men have arrived all
b. The men have all arrived

(238) a. * The men were kissed all
b. The men were all kissed

It quite unclear why these positions should be out on the standard A-movement plus stranding idea.

Again, I will leave these issues to the side, but note that the matter is an *important* one however — should it turn out that the stranding-style analysis is independently demanded, this would be inconsistent with the *general* intuition underlying TCG. In any case, I will leave the matter open here for further future investigation, noting that these general types of facts could constitute a crucial set of cases that could strongly call into question the basic ideas proposed here.

3.2.3. Variable Binding/Condition C Interaction

Consider another case discussed in Chapter 2 (see the discussion there for references), showing an interaction between variable binding by a quantificational element and

Condition C. This will lead us into our discussion of *wh*-movement below, as well as raising some general issues that I will leave open here.

(239) a. *[His$_1$ mother's$_2$ bread] seems to her$_2$ _ to be known by every man$_1$ to be _ the best there is

b. [His$_1$ mother's$_2$ bread] seems to every man$_1$ _ to be known by her$_2$ to be _ the best there is

In the a-case, in order for *his* in the subject to be bound by *every man*, it must be interpreted in the more embedded position — but there it gives rise to a Condition C effect, so the reading on the provided coindexing for the a-case is impossible. However, if we switch-around the QP and the pronoun, as in the b-case, the bound-reading becomes possible, but this only makes sense of the interpreted position is below *every man* but above *her*.

The SCM type of analysis that our framework makes available can account for this pattern as well, though note that it requires that the "A-moved" expression [*his mother's bread*] must reside for *interpretation* in a non-thematic position.

As we noted in Chapter 2 (pointed out in Bošković 2002) the a-case on the bound variable reading is perfectly acceptable so long as we have disjoint reference between *her* and *his mother*. The combination of these observations suggests that intermediate reconstruction is *possible*, but not necessary. It moreover suggests that the output structure handled by the interpretative systems must be coherent (see Bobaljik 2002, Hornstein 2000 on this sense of LF coherence) in the sense that the moved phrasal complex must be interpreted as "in" one or the other positions, but not *both* (as this would cause conflicts that would presumably correspond to unacceptability). However, it seems

that how to understand this isn't entirely straightforward on the view we have been entertaining — nor is it straightforward on standard views. The question is why/how/when it should be possible to have a nominal in such raising environments be forced to be "interpreted" in an intermediate position.

Note that both variable binding and Condition C effects cannot, in general, be workspace-mediated relations on the TCG view. Like the cases of logophors discussed above, these are potentially long-distance relationships (*actual* "long-distance", not superficially so as in linked-local/SCM cases). However, unlike the logophor case, there appears to be structural factors involved — something in the vicinity of c-command is required, whereas this isn't an absolute condition on logophors.

We can now bring up the "re-spell-out" mechanism discussed in Chapter One in connection with a concrete case, in particular the combination of our anti-recursion with workspace connectedness (the latter is repeated here) (see §1.4.1):

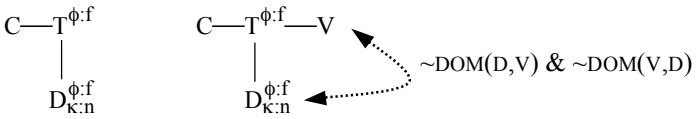
(240) **Workspace Connectedness (DOMINANCE):**
The elements in a given syntactic workspace must manifest a connected dominance order (for every *x, y* in the set, either *x* dominates *y* or *y* dominates *x*)

Insisting that the workspace always maintain a fully connected dominance order yields the need to spell-out (void from the workspace) *every* time branching occurs.

For the raising case, this means that the subject, which in our top-down view begins the derivation in its putative surface position, must associate to T and then spell-out when V is introduced. However, in virtue of (i) the feature-connection between the subject and matrix-T, and (ii) the introduction and identification of the embedded

(defective) non-finite T in multiple raising constructions, the subject can, and in fact must "re-enter" the workspace. Consider:

(241)

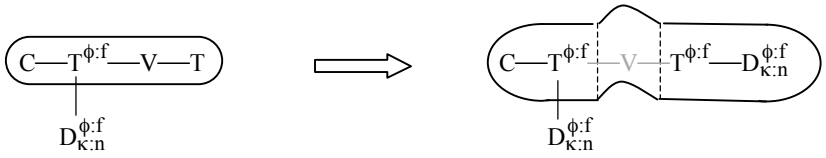


(242)

(241) & (242) show the workspace both prior to and following the addition of V. By hypothesis the D-T relation has occurred, but the addition of V violates connectedness, since D and V enter into no ordering relationship. So D spells-out (the workspace contracts to maintain a connected order). The connection between D and T is understood to be maintained in virtue of their featural (κ/ϕ) relationship.

Next, when defective T is introduced, we have the following:

(243)



In (243) I have for convenience collapsed some steps of derivation. On the left we have the introduction of the embedded non-finite-T. On our anti-recursion assumptions, given that T subsumes T$^{\phi:f}$, the node are identified. Keeping with ordering consistency, this requires that the intervening V be spliced out of the workspace, and the T-T identification effects the reintroduction of the matrix subject (right-hand side of (243) — I include this reintroduction on the horizontal line simply for presentational purposes, left-right and top-down on the page both represent the single dominance relation).

The SCM-raising facts we have canvassed above demand that the entire LF-content of this D-element be in this lowered position. Two sets of questions arise.[97] First, is the LF-content in both the matrix and this new embedded position? Or does the λ-material have to "collapse" to a single position? Second, why are such lowered elements never *re*-pronounced in either intermediate or base positions?

Taking the second question first, the right generalization appears to be that these D-elements are spelled-out in the contexts in which they are initially κ-valued. This accords with the general intuition of their being a "PF"-function to such properties, but note that our story regarding κ-transmission given above for expletive/associate relationships then runs into some trouble. For example, the idea there was that *there* in raising constructions can relate to the structure by κ-matching even though valuation does not occur. In virtue of T-T contractions in raising, the *expletive* element was suggested to be lowered along with T, according to our general story about SCM. In fact, it is difficult to see how we could mangage to avoid lowering the expletive given how we have treated regular nominal expressions in such contexts. However, if PF-spell-out is contingent on the context in which κ-properties are valued, then we expect one of two incorrect results for the raising constructions, either: (i) expletives should appear in every intermediate position in raising (244)a, or (ii) they should only appear in the lowest such intermediate position (244)b:

[97] Actually at least *three* sets of questions arise. The third pertains to the structure of the matrix subject for this example (e.g., [his mother's bread]). Recursion in the nominal domain is not something I have discussed at all here, but presumably this will involve two head elements coding possession relating to the nominal and pronominal. I am abstracting away from this important issue to concentrate on how information flows in these derivations along our equivalent of verbal extended projection sequences.

(244) a. *There seems there to be likely there to appear there to be a man in the room
b. *∅ seems ∅ to be likely ∅ to appear **there** to be a man in the room

The problem lies in the way we have conceived of node-identification. T-T contractions result in the lower "defective" instances of T taking on the matrix κ/ϕ values. I will postpone a sketch of the solution as we will need to say something similar in the domain of WH-movement in our discussion in the next section.

Regarding the first issue raised above: what the TCG mechanics provide, I am arguing, is a natural way to understand why there ought to be intermediate-type effects of the SCM sort. I have argued that it is an attractive platform for studying these phenomena, and sketched some preliminary analyses in terms of one possible implementation. But the general account does not tell us everything, further development is required to understand the issues that arise in interpretation in cases like the one above (and others, see below).

The principles that govern reconstruction/connetivity type effects in A-movement are not well-understood. Some have denied they exist entirely (e.g., see Lasnik 1999, Chomsky 1995), while others have countered that such effects do sometimes show up (Boeckx 2003, Wurmbrand & Bobaljik 1999) and that evidence to the contrary simply points to a lack of full understanding of the differences between the inventory of potentially movable elements, and does not bear on the general idea of SCM.

Some controversy exists over, for example, the status of examples of the following sort (this discussion draws on Wurmbrand & Bobaljik's 1999 presentation, the example is due initially to the work of May 1977):

(245) Some politician is likely *t* to address John's constituency

The claim about this case is that it manifests a scope ambiguity, with "some politicians" taking scope from either the overt/matrix position or the embedded position from which on standard views it is taken to "raise" from. The ambiguity is thus with respect to the predicate "is likely", and in particular whether the existential introduced in the subject nominal is under or over this predicate scope-wise. For example:

(246) **∃ >> likely** = "there is some politician who is likely to make the address"
likely >> ∃ = "it is likely that there is some politician who will make the address"

The ambiguity is clear, and a "copy"-type story, which we have motivated a version of here, can in principle account for this in terms of "interpreting" the nominal in either the upper or lower position (or taking "ambiguity" here to mean that somehow both positions are occupied, so that we may flip back in forth mentally between the two).

Lasnik (1998) has argued, however, that this ambiguity can be explained without reference to scopal distinctions, but rather in terms of specificity. Consider:

(247) Some politician addressed John's constituency

This has a specific and a non-specific reading, where we may or may not (respectively) have a certain politician in mind. And this distinction corresponds to the ambiguity present in the raising case above. Lasnik's point is that scope ambiguities are Q-Q interactions (e.g., of the *everybody loves somebody* sort), and its not obvious that there are such relationships at play in the raising case. But nonetheless it is possible to think about specificity differences in cases like the one in (247) where there is no issues regarding high/low positions from which to interpret an element (though perhaps the

issue is best understood in terms of *v*/VP internal/external, that could be involved in (247)'s ambiguity and the raising one above).

Bobaljik & Wurmbrand (1999) have responded a bit to this line of argumentation (as well to some other challenges raised by Chomsky regarding the A-traces/copies), but I won't go into this further here. Relevant here is their general conclusion, which is just that while there is reason to doubt that even if SCM in A-relations is totally general, it doesn't always necessarily lead to reconstruction/connectivity effects, but that there are nonetheless some cases where it seems that such analyses are required to understand the cases where such effect *do* manifest. Here (above) I have concentrated on one main type of case involving interference effects in binding of *self*-forms, but there are other cases which bear on these matters that will require further attention, and which will be required to help sort out further details for the TCG-style analysis I am offering.

3.3. Linked Local Relations II: *Wh*-Movement

We can pick up the thread from the last section regarding variable-binding and obviation interactions by posing the following question: if we keep the raising construction in (239) the same in all other respects, but change the subject element housing the relevant NP and pronoun to a *wh*-phrase, do we see the same pattern as we saw for the A-movement case?

3.3.1. κ-Identification & Local Movement

Consider:

(248) a. *[Which of his$_1$ mother's$_2$ pies] seems to her$_2$ _ to be known by every man$_1$ to be _ the best there is

b. [Which of his$_1$ mother's$_2$ pies] seems to every man$_1$ _ to be known by her$_2$ to be _ the best there is

On standard bottom-up derivational views such a *wh*-element would begin in the base/θ-position, and A-move just like the NP in RtS, but the "last" movement would be to the C-domain to licensing the *wh*-properties. This case manifests exactly the same pattern as the ordinary raising case examined in the previous section. In particular, binding of the pronoun *his* by *every man* is possible without there having to be obviation between *her* and *mother*. This raises some questions on our view (though not on standard approaches). We have suggested A'-movement to be a dominance-encoded feature licensing relationship, so the beginning of the derivation for either of the above cases would look as follows:

(249)
$$\begin{array}{l} C^{\phi:\varnothing}_{WH}\text{—}T^{\phi:\varnothing}_{\kappa:n}\text{—}V\text{—}T \\ \;| \\ \;|\phi:f \\ D^{WH}_{\kappa:\varnothing} \end{array}$$

The problem is that we have understood so far the relevant relationships to go as follows:

(250)
$$\begin{array}{l} C^{\phi:\varnothing\Rightarrow\phi:f}_{WH}\text{—}T^{\phi:\varnothing\Rightarrow\phi:f}_{\kappa:n}\text{—}V\text{—}T \\ \;| \\ \;|\phi:f \\ D^{WH\Rightarrow\varnothing}_{\kappa:\varnothing} \end{array}$$

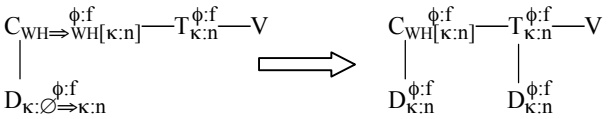
And then what effects the "raising" ("lowering") in such constructions are T-T identifications. But this does provide a mechanism for the content of the matrix *wh*-element to be lowered to the embedded edges of the non-finite complements, as these have by assumption been understood to not involve a C-layer. But the binding/scope interactions above seem to insist that this is what is required.

On standard views this is unproblematic: the *wh*-element begins in a θ-position, and is raised from A-to-A position (involving the edges of the non-finite complements), finally landing in the matrix κ/ϕ position (matrix T), and then A'-moving to C. Thus on a copy view within such standard assumptions we do not have a problem understanding both the raising case offered above nor the *wh*-movement variant, since the latter kind of relationship *includes* the former.

We are now in a position to further specify the "variable" role that we are suggesting for κ-properties. Note that the traditional GB-era view that we discussed above viewed κ-marked traces as "syntactic" variables, which on some implementations were understood to map to semantic ones. This is more-or-less what we have been presupposing in our discussion so far. However, it is possible given the current structure of our account to entertain a different claim: κ-properties are *literally* syntactic variables, in the following sense.

We have so far entertained the idea of ϕ-features indexing the open positions of thematic predicates. The idea here is that these elements are the syntactic side of relations to semantic variables (i.e., the open positions of θ). The path of ϕ-agreeing nodes in the structure was suggested to "lead to" a κ-position in regular A-relationships, and that "κ-marked nominals" are connected in this way to θ. Suppose that the T-D relation resulting in κ-valued on D and deleted from T is a process, as we have been suggesting, that we might call ARGUMENT IDENTIFICATION. In the case of an overt nominal, say a subject, this mark signifies the element that is connected to θ via the sequence of ϕ-properties on the path.

Now consider the situation above. If κ serves to identify arguments, we might entertain the following possibility similar to the node-identification discussed earlier:

(251) 

In short, it looks like we need something like local movement after all. So far the only things in our implementation that really resembled movement was the edge-to-edge lowering effected by context/node-identification. However, the suggestion here is that this is tied to the special role of Case as a syntactic variable. In virtue of the WH-feature valuation by the local κ-property, the *wh*-element will come to be κ-marked. The suggestion is then that in virtue of this identification, the C-related element comes to be dominanted by T — κ-properties thus mark the entire unit, and where there is locally co-valued κ, there is essentially the same sort of effect that we see with categorial node identification. That this doesn't happen with ϕ-properties is thus a constitutive difference between these feature types. ϕ-properties define a local unithood (a stretch of co-valued nodes in a dominance sequence — a "chain"), and κ marks arguments that are then related by this chain to θ.

There may be other technical ways to implement a solution to this issue, but I will assume this for the rest of this work. So, to sum: κ-valuation marks arguments, and every occurrence of κ on the path is understood to dominate the κ-marked element. Note that D ends up in a derived relation with T, so that A vs. A'-relationships involving D-T are distinguished by the absence/presence of κ on T (respectively).

With these mechanical assumptions the regular T-T node identification procedure discussed above for the raising cases will now function to lower (the copy of) the *wh*-element to each non-finite complement edge, thus yielding a structure accounting for the possibility of the variable-binding/condition-C interactions for (248).

3.3.2. Core Cases of A'-SCM (& Some Technical Problems Addressed)

The general structure of the TCG account of SCM carries over to the core cases of *wh*-movement more-or-less straightforwardly. However we are now in a position to re-raise and discuss some possible answers to some technical issues left open earlier. In particular, we considered the possibility in Chapter One of having a "Make-OP" style operation built into our WH-feature licensing mechanism — basically C-WH deletes D-WH, leaving this property only on C. This suggests then being able to treat the "residue" of such a deletion on D as a (potentially complex) variable like element.

However, the mechanics of node-identification and contraction suggest that we should understand this WH-property as copied to all the lower C-nodes in our SCM analyses, in virtue of the identification which makes the lowering of the actual D*wh* (now a "residual" structure interpreted as a variable). We contrasted these two schemas in §1.4.1 ((54) & (55), repeated here):

(252) $C_{WH[\kappa:\varnothing]}$—….—$C_{WH[\kappa:\varnothing]}$—….—$C_{WH[\kappa:\varnothing]}$—….—
$D_{\kappa:\varnothing}$ $D_{\kappa:\varnothing}$ $D_{\kappa:\varnothing}$

(253) $C_{WH[\kappa:\varnothing]}$—….—C—….—C—….—
$D_{\kappa:\varnothing}$ $D_{\kappa:\varnothing}$ $D_{\kappa:\varnothing}$

We noted that it is really the latter, and not the former that we want, though as we have stated things the former is the one that the TCG derivations seem to produce.

Let us consider then a somewhat subtler formulation of the process of node-identification. What we want is for the process to yield an identification that will justify the "re-entering" into the workspace of the dominated, "to-be-moved" element. However, we want to this to proceed *without* a copying of all of the information associated with the upper element, *but* we want the upper-element properties to remain "visible" in the workspace, so that the lowering can result in a local structure where licensing occurs.

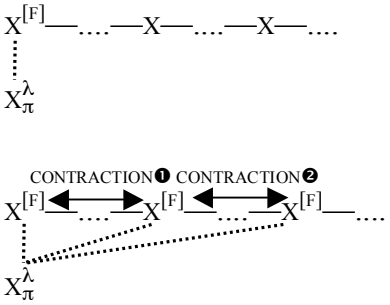
Note that this issue concerns both the A'- and A-relations. The issue arose with respect to A-relations above with respect to having κ-properties appear in all embedded positions and our suggestion that κ-licensing could be understood (for A-relations) to indicate the point in the structure where an element is pronounced. But for expletive *there* we suggested that this element was precisely one that did not allow local κ-valuation, and so it must be carried along to embedded contexts in our version of the A-type of SCM.

The idea then for an alternative view of node-identification would be to say that the features associated with a node X are "fixed" with respect to the output structure. Whatever the nature of the connection between (e.g.) WH-properties and the λ-vocabulary associated with that node, that relation keeps those properties fixed to that initial position as it is determined when the element enters the derivation.

Node identification can then still occur, under the same general conditions of subsumption as we have been assuming. But while this will identify positions in the workspace, it will not "copy" the relevant features to the lower position. To see the idea

conceive of our workspace/output distinction in terms of separate layers or tiers of structure. Consider:

(254)



I mentioned earlier on in this discussion (see Chapter One) that the WS/O-distinction allowed a way of thinking of "many" in the output structure as "one" in the workspace. This is a situation where the mapping is insisted to be one-to-one for any given stage. At the "end" of a derivation, the relevant λ-properties will be connected to lower variables via the mechanisms we have been developing above, but the syntactic information itself is "fast and fleeting". It is available for local domain construction within the workspace, and is, in situations allowing node-identification, permitted to be "carried over" to lower domains, but once the derivation is completed the workspace itself is gone, and so are the formal properties contained within it (that λ and π information is connected to).

So, how then do we end up with "one" in the workspace corresponding to "many" in the output structure? The idea here relies on the "re-spell-out" mechanism discussed earlier. If the workspace is constrained by the connectedness requirement, insisting essentially that there only be a single dominance sequence at any given stage, then branching requires spelling-out (removal from the workspace). However, we are viewing the node-identification procedure as preserving output structure relationships *so long as they do not introduce local ordering conflicts in the workspace*. This means that any

element Y that may be associated with X in our schema above in (254) will be "re-entered" into the workspace in virtue of identification of X's. And as further structure is added they will have to "re-spell-out". This yields multiples in the output for the associated Y-elements. But notice that no such "leaving" and "re-entering" is required for the X-elements as these never cause a problem for the connectedness condition (they are always present in the path).

However, we noted also in our Chapter One discussion that any such mechanism that would insist on "reintroducing" spelled-out material in virtue of the node identification process could potentially run afoul of our anti-recursion conditions, as such spelled-out branches could be arbitrarily complex. Suppose instead that the initial feature licensing relationship which holds of the top-most element is sufficient to evoke the "lowering" — that is, what matters to this process is the initially established agreement (ϕ) relationship, and that this information is "carried over" to lower domains in virtue of successive node identifications. Then, instead of reassigning syntactic/categorial information to such lowered complexes, we can say that some minimal information is assigned, perhaps just D and the relevant ϕ information, or perhaps just ϕ. The result is that the lowering that attends node identification re-introduces only an index of sorts which we take to dominate just LF-relevant vocabulary.

This allows us to keep with the idea of "pronouncing" elements in A-relations where the relevant κ-property is, without the problem of expletive-*there* spell-out raised above. And, it gives us the structures we want for *wh*-movement, with local copies of variable like elements (e.g., WH(x)...[x mother]....[x mother]....[x mother], in *whose mother did John think Bill knew Sarah met*, etc.).

Now consider some of the binding-connectivity effects discussed in Chapter 2. For example:

(255) a. Which pictures of himself did John know Bill wanted?
b. [which...himself] did John know [which...himself] Bill wanted [which...himself]

The ambiguity present here we can now attribute to the TCG derivations of SCM effects, again, like the raising cases, without the postulation of special features (M-features) driving the individual movements. The *self*-element ends up in local relationships without interveners with both of the possible antecedents.

This helps to understand cases such as those discussed in Chapter One as well, for example:

(256) a. John thought pictures of himself/*herself were on sale
b. Which pictures of himself/*herself did John think were on sale

(257) a. ??John thought Mary sold pictures of himself
b. Which pictures of himself did John think Mary sold

And reciprocal elements distribute in basically the same way as is well-known:

(258) a. The boys thought pictures of each other were on sale
b. Which pictures of each other did the boys think were on sale?

(259) a. ??The boys thought I sold pictures of each other
b. Which pictures of each other did the boys think I sold?

Impossible bindings across intervening elements suggest that a "direct" relationship between the matrix *wh*-phrase and the base position is insufficient. The well-formedness of the b-cases in (257) and (259) can be accounted for if there is a local relationship to the phrase containing *self/each-other* — and this is what the SCM-style analysis provides.

Note as well that where we have suggested that intermediate C-nodes are absent, as with the complements of raising predicates, we do not have a locally licensed "copy" that could enter into the relevant binding relations. The following examples (Abels 2003:30) illustrate:

(260) a. *Which picture of himself did Mary seem to John (Mary) to like *e*
b. Which picture of himself did it seem that John liked *e*
c. Which picture of himself did Mary think John wanted (John) to pack *e*

As Abels observes, the a-case supports the idea that raising infinitives are not CPs, since if they were they would support a potential landing site that would put the *wh*-phrase within the local environment of the NP (*John*) that could be a binder. Where we have evidence for CPs, as in the b- and c-cases, we also have the possibility of binding the *self*-form.[98]

Note as well that under the contraction view of SCM we in general expect *nested* dependencies of the sort predicted by Path Containment approaches, pioneered in the work of Pesetsky (1982), Kayne (1984), May (1985) and others.

This is a quite general property of multiple "like" dependencies. Consider the following familiar sorts of cases from Pesetsky (1982):

(261) a. What books do you know [who [PRO to persuade *e* [PRO to read *e*]]]
b. *Who do you know [what books [PRO to persuade *e* [PRO to read *e*]]]

[98] This a-case above also reveals a parallelism with a comparable copy-raising construction, suggesting that these do not involve CP's either.

a. *Which picture of himself did Mary seem to John like she wanted *e*
b. Mary seemed to John like she wanted pictures of herself
c. *Mary seemed to John like she wanted pictures of himself

We see the same kinds of nesting versus crossing effects across the range of A'-movement relationships, including within the structure of relative clauses, in topicalization, infinitival relatives, *tough*-movement, *too/enough*-movement, and comparatives (see Pesetsky 1982:269 for examples).

On the general structure of the account, it is worth pointing out that C-C nodes in the workspace will be unable to be identified in workspace contraction if subsumption does not hold. So, if we understand interrogative embedded complements as being specified for WH, then this lack of ability to contract/identify can yield for us an account of *wh*-islands.

(262) ?Who did John wonder whether Mary liked _ ?
?? Who did John wonder who Mary liked _ ?

Moreoever, if we take the identification of C-nodes to be sensitive to a more general category of operator elements, then we can extend the "impossible contraction" story to other classes of so-called non-bridge verbs, like factives for example (see Frank 2002 for some discussion along these lines).

There are, however, cases we mentioned in Chapter Two which suggest that SCM is "more cyclic" than our view predicts, in particular facts that suggest that the "edge" of *v*/VP is an intermediate landing site. I will return to these cases below, after we have discussed some possible extensions of the general architecture to local domains. The situation we are in with respect to evidence for a *v*/VP level intermediate movement is fairly straightforward: we are forced to posit more structure within local domains in order for the general approach to yield the facts.

3.4. Local Relations: Part Two

In this section I return to some of the issues raised at the beginning of this chapter regarding local relationships for simple transitives. There we noted that our feature-valuation mechanics seemed to require bi-directional valuation on the dominance ordering (to allow but upward ϕ-valuation and downward κ-valuation in D-T relations, for example), but that no locality restrictions suggested a chance for chaos when more than one θ-element would be in the same local domain. Here I pursue the possibility that this situation never arises.

3.4.1. Raising-to-Object (RtO)

Consider:

(263) John believes him to be a genius

There are two main lines of thinking regarding these constructions which differ with respect to how the accusative-marked element (*him*) is viewed with respect to the matrix/embedded clause boundary. The choice of analysis typically swings with the claims made about the categorial status of the embedded infinitival. On the one hand, there is the idea that *him* is in the lower clause, and that there is an "exceptional" process which converts the categorial status of the embedded structure from CP to TP (S' to S in traditional terms; see Chomsky 1981). On the other hand, there is the idea that these cases involve raising to an "object" position (Postal 1974). In modern views that have resurrected this Raising-to-Object (RtO) view, the categorial type of the embedded clause is usually taken to be a TP, and much is made of the similarities of these cases to the RtS sort of NP-movement discussed above.

A number of factors favor the RtO type of analysis,[99] here I will name just a few. First, passivizing *believe* targets this embedded ECM'd element, strongly suggesting matrix objecthood since, much like the impossibility of raising out such contexts, subjects of lower CPs clearly cannot undergo this process:

(264) a. He is believed to be a genius
b. *He is believed that _ is a genius

Second, binding-theoretic conditions apply to this element as if it is a matrix object, and not like a lower subject:

(265) $John_1$ believes $himself_1$/*him_1 to be a genius

However, the meaning equivalence of the following two cases insists that we understand the ECM'd nominal to be the *thematic* subject of the lower clause:[100]

(266) a. John believes Dave is a genius
b. John believes Dave to be a genius

But, on the other hand, binding conditions differ in their effects for these two kinds of complements, which again suggests that the ECM'd nominal is in the higher clause:

(267) a. $John_1$ believes $he_{1/2}$ is a genius
b. $John_1$ believes $him_{*1/2}$ to be a genius

The combination of these facts — participation in the formal processes of matrix objects but thematic association to the embedded material — strongly suggests a raising-style acccount.

[99] See Johnson (1991), Koizumi (1993, 1995), Lasnik (1995), Runner (1995), Bobaljik (1995) for recent discussions.

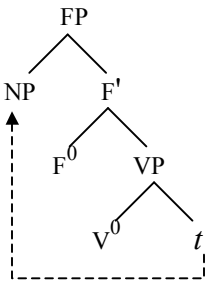
[100] Rosenbaum (1967), among others. For summary discussion and further references see Runner (*to appear*).

Another standard argument includes reference to other effects of hierarchical position as indexed by binding/scope possibilities, which suggest that the ECM'd nominal is in the higher clause (Lasnik & Saito 1991; Postal 1974):[101]

(268) a. ?The DA proved the defendants$_1$ to be guilty during each other's$_1$ trials
b. *The DA proved that the defendants$_1$ were guilty during each other's$_1$ trials

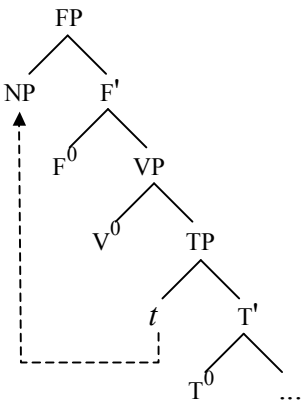
Postal's classic RtO analysis of these constructions, which seems to do pretty well with the facts, was imported into current approaches via the assumption that objective/accusative Case assignment is essentially like that of subjects, and involves movement from a base thematic position to a specifier position. In Chomsky (1991) Lasnik & Saito (1991), and Johnson (1991), among others, this was considered to be an object-related Agreement head (Agr_O). In more recent work (Chomsky 1995 and subsequent) the notion of separate agreement heads has been called into question.[102] However, the general idea of the RtO-analysis can be pictured as follows, where the nominal is understood in the general case to raise to some functional category F below matrix T to licensing its Case properties, as in (269). So the implementation of Postal-style RtO then looks like (270):

(269)



[101] As is well known, similar effects of hierarchy hold for Condition C, negative polarity licensing, etc.
[102] Though see Belletti (2001).

(270)



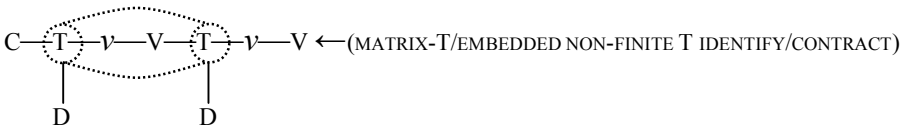
As a last review note on these constructions, an analysis like (270) also fits nicely with the various "subject-like" properties exhibited in ECM, as witnessed by expletives and idiom pieces in these positions, paralleling the properties of athematic positions in raising to subject constructions:

(271) a. I believe there to be a moron in the White House
b. I believe it to be the case that Dave left
c. I believe it that Dave left
d. I believe the shit to have hit the fan

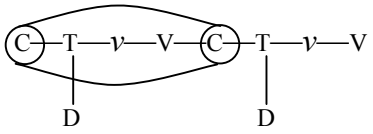
So suppose then that something like the object-raising story is correct. How can we capture it in our view of contraction? Notice that *without* some assumption about a higher functional element responsible for accusative Case in ECM, we predict that (272) should be a *subject*-raising situation if the infinitival structure is a TP.

(272) John believed him to like carrots

This ought to manifest a sequence of elements and a phase structure like (273) if there is no intervening functional element of the right sort to block the contraction:

(273) 

There are two possibilities. We could consider something like the old CP/S' plus "deletion" (i.e., ⇒TP/S) analysis of Chomsky (1981), in which case the T-T contraction illustrated above would be blocked as desired:

(274) 

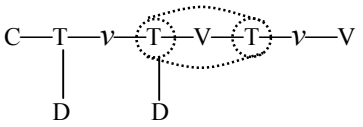
But this would be to loose all the nice properties of the object-raising analysis sketched above. Plus, its not clear how we could possibly implement the notion of S'/CP-deletion in this system, since that would put us right back in the same situation we started with regarding the undesirable contraction above in (273).

Another alternative, one consistent with the story we have told about RtS, would be to claim that there is an matrix-object-related T node above V, but below the C-T-*v* subject argument complex.

Pesetsky & Torrego (2004) suggest such an account within their general attempt to connect the presence/absence of T with Case-theory generally. Their structure for simple transitives is thus:

(275) $[_{CP}\ C^0\ [_{TP}\ T^0\ [_{vP}\ v^0\ [_{TP}\ T^0\ [_{VP}\ V^0\ ...]]]]]]$

This allows for us to consider the possibility that the matrix clause is really hiding a bit more structure. And this story would then allow us to view object-raising as T-T contraction, exactly as we did above for subject-raising, but now with "object-related" T contracting with the embedded infinitival.

(276) 

Note that our view of structural contraction forces this analysis on us. I find this interesting since this sort of iterated clause-internal "mini-clause" structure has been explicitly argued for by a number of authors under the label of the so-called Split-VP Hypothesis.

The general idea behind Split-VP includes the now fairly widely adopted view of separating/dividing the lexical shell of verbal domains into a core verbal element V (the ultimate head) and a small-*v* element, understood to introduce an external argument.

(277) $[_{vP}\ v^0 \ldots [_{VP}\ V^0 \ldots]]$

Koizumi's (1993, 1995) notion of Split-VP has it that these verbal elements are separated into distinct zones in virtue of the existence of one (or more) intervening functional elements (F-heads):

(278) $[\ \ldots [_{vP}\ v^0 \ldots [_{FP}\ F_n{}^0 \ldots [_{FP}\ F_1{}^0 \ldots [_{VP}\ V^0 \ldots]]]]\ \ldots\ ]$

There is a diverse array of analyses evoking Split-VPs in this sense in the literature, and while the exact nature of these intervening functional elements is by no means settled, there appears to be something of a growing consensus that some such division/separation approach may be correct. Candidate types evoked to label these intervening functional elements between the separated θ-elements (*v* and V) include Agr(eement), a lower T(ense), Asp(ect), a lower instance of C(omp), among others.[103] Lasnik (1995, 1999) includes arguments based on the properties of pseudogapping that support the idea of

[103] Many others, actually. Work in the minimalist program has seen no shortage of proposals arguing for functional category distinctions. I will be working with fairly blunt tools in this regard, but as mentioned earlier, the efforts in analysis which this thesis aspires to are mainly in service of developing a clear and plausible picture of the theoretical ideas and the consequences for general architecture.

Split-VP and *overt* object and verb movement in English. Runner (1995) contains arguments along these lines as well. I will not review these arguments here, and I will also not be discussing head movement in this thesis. But the conclusion which I wish to extract from this is that the iterated mini-clause analysis that our view of structural contraction appears to force upon us is by no means unprecedented, and in fact has a fair amount of independent empirical support.

Let's consider this a bit more. What prevents T-T contraction then *within* the main clause, so that external and internal arguments might either become confused (as we worried about earlier) or inappropriately identified? Note that the assumption here would be that the two T-elements within a single clause would be distinguished by κ-properties, as follows:

(279)
$$\begin{array}{l} \mathrm{C}\text{—}\mathrm{T}^{\phi:f}\text{—}\, v^{\theta[\phi:\varnothing]}\text{—}\mathrm{T}^{\phi:f}_{\kappa:n}\text{—}\mathrm{V}^{\theta[\phi:\varnothing]} \\ \quad\;\; | \\ \quad\;\; \mathrm{D}^{\phi:f}_{\kappa:n} \end{array}$$

The anti-recursion condition on workspaces will force the "subject" T-*v* structure to be spliced out when the second T is introduced. This has the welcome outcome that it appears to solve the difficulties we raised at the outset of this chapter regarding the locality of valuation. We have essentially imported the structure of the account now into the local domain of single clauses.

So we now view ECM as RtO, parallel with our RtS derivations from earlier discussion. For example:

(280)
$$\begin{array}{l} \mathrm{C}\text{—}\mathrm{T}^{\phi:f}\text{—}\, v^{\theta[\phi:\varnothing]}\text{—}\mathrm{T}^{\phi:g}_{\kappa:a\Rightarrow\varnothing}\text{—}\mathrm{V}\text{—}\mathrm{T} \\ \quad\;\; | \qquad\qquad\qquad\quad | \\ \quad\;\; \mathrm{D}^{\phi:f}_{\kappa:n} \qquad\qquad\;\; \mathrm{D}^{\phi:g}_{\kappa:\varnothing\Rightarrow\kappa:a} \end{array}$$

Assuming that the complements of "ECM verbs" (now using the term as a descriptive label) is a "defective" form of T, we would have the same node-identification and contraction for the object-raising in (280) as we did for the subject cases.

In the next section I suggest that this view of clause-internal structure can be put to work to yield some interesting properties of passives, and make some tenative suggestions regarding local binding and control phenomena.

3.4.2. Passives, Local Binding, & Control

Note that the root-first directionality for the emergence/expansion of these sequences of categories is crucial. On either a representational or bottom-up derivational view, it would be possible to view the two instances of T in these structures as undergoing contraction, resulting in the following structure where we have T-contraction 'over' an intervening C:[104]

(281) C—T—*v*—C—T—V

On the root-first view, (281) presents phase conflicts with those demanded by the necessary C-C contraction, which we can see this clearly by superimposing the two:

(282) C—T—*v*—C—T—V

Assuming for the moment that items are entered into the workspace one-at-a-time with the assumed direction/ordering given above, this means that C-C contraction will always block T-T. However, we predicts on this one-at-a-time view that in the absence of the

[104] Although in a representational implementation we wouldn't view this as literal contraction in the sense I have been entertaining, but rather just some notion of domain that would be defined over "like elements". What I am arguing here is that only on the root-first view do we get the right sort of domains.

intervening C-element we should see instances of T-T identification & contraction if there are no properties of these elements to distinguish them (so subsumption does hold).

(283) C—T—*v*—T—V

This, I propose, is exactly what happens in the case of passive. Absence of κ on object-related-T could be seen to drive a kind of "clause-internal" raising on these assumptions. Suppose that objective T can enter the derivation without κ/ϕ-properties.

(284) C—$T^{\phi:f}$— $v^{\theta[\phi:\varnothing]}$—T—V
|
$D^{\phi:f}_{\kappa:n}$

The idea of the instance of T-T contraction happening over the element *v* instantiates the idea that when accusative Case is absent, the external role could be seen as essentially spliced out of the active stretch of derivation *without* being ϕ-valued. This is, I submit, the present but unexpressed external θ-role in passives.

Such a possibility suggests that θ-roles need not, strictly speaking, be assigned. Left open (not closed off by the introduction of a local satisfied Case property) they function as implicit arguments. This view requires however that in general θ cannot be ϕ-valued *prior* to the relevant T-T identification. Suppose then in general that θ differs from the other properties we have discussed in that it is valued *when it is removed from the workspace*. On this view v^{θ} (or any θ-element) can only be properly ϕ-valued if it is "spelled-out" together with a superordinate specified ϕ-property. I assume this in what follows; to be explicit:

(285) **LAST RESORT θ-VALUATION**: θ[ϕ:∅] is valued at Spell-Out

This isn't a completely wild speculation — the general idea is that we regard θ-relationships (the ulimate "integration" of nominal elements into the emerging event structure) as a matter of the interface mapping. Here this is our WS/O-distinction, so its not unreasonable to locate our version of θ-connections to this particular mapping (WS to the "LF-relevant" properties of the derived output structure).

Note that we had to assume in earlier discussion that κ-properties are parasitic on ϕ-properties and their valuations. But we discussed a few cases where we wanted ϕ-relationships to be able to hold independently (e.g. the secondary predication case in the beginning of this chapter; the small clauses where associates of expletive-*there* may be found, etc.). What happens if there are ϕ-properties on object-related T, but no κ-properties?

In such cases I suggest, we have the possibility of connecting internal and external (or *v* and V) θ-properties. Such a structure is attractive for thinking about the properties of so-called inherently reflexive (286) and reciprocal (287) verbs.

(286) a. John washed/bathed/shaved
b. John washed/bathed/shaved himself

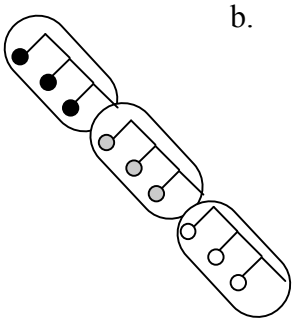
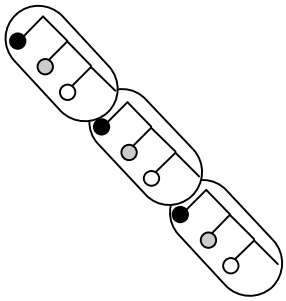
(287) a. The women met/kissed/hugged
b. The women met/kissed/hugged each other

Though this view requires that we posit an additional layer of functional structure. Consider:

(288) $\text{C—T}^{\phi:f}\text{— } v^{\theta[\phi:\varnothing]}\text{—T}^{\phi:\varnothing}\text{—V}$
$\quad\quad\;|$
$\quad\;\text{D}^{\phi:f}_{\kappa:n}$

Any ϕ-relationship between T-T would result in the splicing-out of *v* situation suggested above for passivization. How then could we manage local connections between arguments of the sort that manifest in the inherent reflexives/reciprocals?

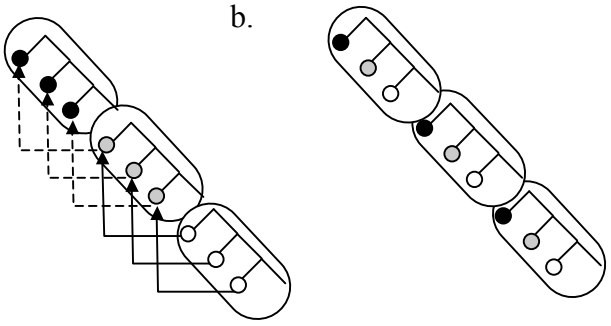
Suppose we take the idea of iterated clause-internal structure around θ-elements a step further, and view the C-T-V type structure as the general way that functional elements cluster around lexical ones, resulting in a full "stacking" view, to borrow terminology from Bobaljik (1995). He contrasts split-VP type architectures with a "leap-frogging" view. Consider an illustration. Take the light nodes to be rougly θ-related, grey to be κ/ϕ-related, and the dark nodes to be operator/A'-related:

(289) a.  b. 

The a-view is the one which makes for a leapfrogging view of domain relationships. This is a fairly common perspective. Grohmann (2003) enshrines roughly this C-T-V kind of division in his "prolific domains" view of clausal architecture, where each domain potentially decomposes into more fine-grained inventories of categories. Roughly, there is a domain where all thematic relations are computed, and this maps (by movement) to a domain where κ and ϕ and the like are licensed, and then these map (by movement) to a higher domain involving relationships of the A'-sort, including perhaps discourse related functions (e.g., like topic and focus and the like in a Rizzian "split-CP" view; see Rizzi 1997; see also Platzack 2000 for a view similar to Grohmann's).

In Bobaljik's terms, results in "leapfrogging" type movement relations to map the elements from the lowest to the highest domains:

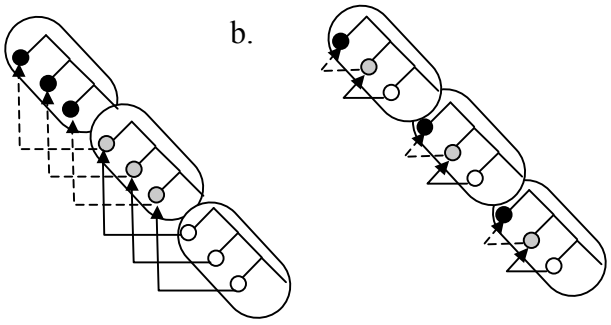
(290) a. b.



Movement relations needn't necessarily always be uniform in the fashion pictured above on the leapfrogging view — note that stranding elements in different domains in these movement relationships yields the degrees of freedom to describe various different word-order, scope/binding relations, and the like (e.g., depending also on the issues of the sort mentioned earlier in our discussion of the WS/O-distinction regarding where one keeps the λ- versus the π-relevant information).

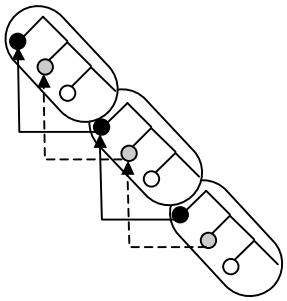
The present architecture could be seen as embracing the same general vision of functional divisions, but with a different view about how they are organized and come together. The relations indicated in the a-view above correspond to the following ones in the b-view:

(291) a. b.



In addition we have suggested certain limited ways that the b-type domains can interact across their boundaries. That is, there are a limited range of relationships of the following type:

(292)



These correspond to the relations governed by the context/node-identification, with the notion of workspace contraction serving to limit the available "viewing window" to just these domains housing distinct elements.

Bobaljik (1995:ch3) gives a range of arguments in favor of the b-view ("stacking" in his terms) and offers a number of arguments against the a-view ("leap-frogging"). What I will consider here and below is an extension of this general way of thinking that aims to stay consistent with the general notion underlying the TCG approach as we have been developing it.

Suppose then that we have a fully iterated view of local transitive structures, this would then look as follows (with the entire presented without any licensing of properties indicated):

(293) $C^{\phi:\varnothing}$—$T^{\phi:\varnothing}_{\kappa:n}$—$v^{\theta[\phi:\varnothing]}$—$C^{\phi:\varnothing}$—$T^{\phi:\varnothing}_{\kappa:a}$—$V^{\theta[\phi:\varnothing]}$
$D^{\phi:f}_{\kappa:\varnothing}$ $D^{\phi:f}_{\kappa:\varnothing}$

I have added ϕ-properties to the C-elements, let us now consider how this view might function with respect to inherently reflexives/reciprocals. Above we noted that T-T

identification would result in splicing-out *v*, yielding a present but not κ/ϕ-connected element that would be interpreted in the output as the "implicit" external argument present in passives. Two possibilities suggest themselves on this picture. First, the derivation could proceed essentially as in previous discussion with two new additions. Take the derivation up to the introduction of *v* in (294) (all the feature valuations are pictured here on a single step for convenience):

(294) $C^{\phi:\varnothing\Rightarrow\phi:f}$—$T^{\phi:\varnothing\Rightarrow\phi:f}_{\kappa:n\Rightarrow\varnothing}$— $v^{\theta[\phi:\varnothing]}$
|
$D^{\phi:f}_{\kappa:\varnothing\Rightarrow\kappa:n}$

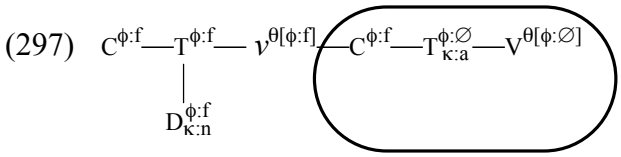
Two differences are now incorporated: (i) ϕ-properties on C, which are locally valued when D is related to T, and (ii) the θ-property is *not* valued for ϕ (this happens in the mapping to the output — whenever *v* is "spliced-out"). Subsequent addition of our hypothetical "object-related" C-element then yields:

(295) $C^{\phi:f}$—$T^{\phi:f}$— $v^{\theta[\phi:\varnothing]}$—$C^{\phi:\varnothing}$
|
$D^{\phi:f}_{\kappa:n}$

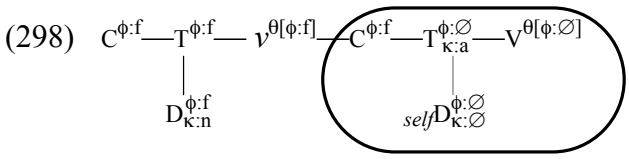
Since this new C-element subsume the higher one, we would have an instance of contraction, which I will suppose results in the following:

(296) $C^{\phi:f}$—$T^{\phi:f}$—$v^{\theta[\phi:f]}$—$C^{\phi:f}$
|
$D^{\phi:f}_{\kappa:n}$

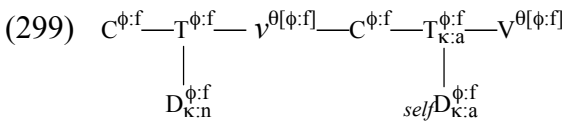
The important parts (highlighted above) are: (i) *v*-θ is valued in the mapping the output (in being voided from the workspace) and (ii) the lower C-element is now valued for the upper domain's ϕ-value. Now addition of the lower T-V structure looks as follows:

(297) 

Presence of T-κ:a (accusative) on our assumptions requires/enforces distinct D. Suppose that overt *self*-anaphors are divided into a pronominal part and the "self" part, and that the function of the "self" part is to absorb accusative κ (see Hornstein 2000 for a similar view along these lines). I assume that is arrives as a bundle with its "pronimal" part which is a bare D with unvalued ϕ, so that we have the element which I will mark as just $_{self}D^{\phi:\varnothing}_{\kappa:\varnothing}$

(298) 

ϕ-valuation then works in the expected way to yield:

(299) 

Which is thus all one ϕ-chain. The assumption is that *self*-serves to "capture" the κ-property, so it does not serve to individuate the "pronominal" part, which becomes valued by T-ϕ. Thus the two θ-elements have the same index, yielding the θ-properties of local anaphors. Note that we could alternatively view the *self* forms as coming to the derivation specified for ϕ, this would result in the usual D-T exchange of κ/ϕ-values. On the view above there is the oddity of having T-ϕ filled in by C, and then having T value *both* the κ and ϕ properties of the anaphor. On the alternative just mentioned, we would get the same result as the representation in (299), except that technically the marked ϕ-features below will not have entered into a valuation relationship:

(300) $C^{\phi:f}$—$T^{\phi:f}$— $v^{\theta[\phi:f]}$—$C^{\phi:f}$—$T^{\phi:f}_{\kappa:a}$—$V^{\theta[\phi:f]}$

$D^{\phi:f}_{\kappa:n}$ $\quad$ *self*$D^{\phi:f}_{\kappa:a}$

In our discussion of logophoric *self*-elements above, we suggested the following difference between local reflexives and logophoric-*self*, repeated here:

(301) a. **Local Reflexives**: ϕ:f—θ ... ϕ:$\varnothing$ — θ

MATCHING/VALUATION

b. **Logophoric -*self***: ϕ:f—θ ... ϕ:f—θ

LINKING

In the local context the matter is obviously quite subtle, as a distinction is being drawn between one versus two tokens of a valued feature. The intuition behind local valuation, implicit throughout, is the notion that is known as "reentrancy" in feature-based/unification frameworks,[105] is that co-valued features are literally sharing a value. I have been assuming here that this yields a kind of internal unit-hood along the dominance sequence, and what is being explored now is the possibility of such relations extending across local recursive domains. (Relations between *valued* ϕ-properties I have suggested be treated as relations on the output structure of the linking sort).

Note that regular pronouns on the present view would then plausibly be viewed as coming to the derivation with valued ϕ, and unvalued κ, like regular nominals. This would on our assumptions yield local obviation (**John*$_1$ *saw him*$_1$).

The idea for inherent reflexives would then be to say that these occur where the lower κ-property would be absent entirely, as follows:

[105] See Shieber (1985) for discussion and references.

(302) $C^{\phi:f}$—$T^{\phi:f}$— $v^{\theta[\phi:f]}$—$C^{\phi:f}$—$T^{\phi:f}$—$V^{\theta[\phi:f]}$
|
$D^{\phi:f}_{\kappa:n}$

Assume that this is how inherent reflexives work, and that something *like* this is relevant to inherent reciprocity. There are obviously complications with the latter that do not arise in former regarding plurality and how members of the denoted set are understood to participate in the relevant relation (e.g., sorting out various flavors of "strength" of the reciprocal relation). Setting aside this significant complication, what I'd like to focus on now is the following property that arises in both in passivization:

(303) a. John washed/bathed/shaved
b. John washed/bathed/shaved himself
c. John was washed/bathed/shaved

(304) a. The women met/kissed/hugged
b. The women met/kissed/hugged each other
c. The women were met/kissed/hugged

A curious fact about these kinds of predicates is that both the inherent reflexive and reciprocal readings disappear in passivization.[106] The c-cases above cannot have the b-reading which the a-cases with 'missing' direct objects obligatorily have.

So the idea here would be that the upper (nominative) κ-properties, in virtue of contraction, would create a second κ-domain, thus serving to index both the external *and* the internal role. This general line of thinking could then be understood to support the inherent reflexive/reciprocal readings.

Roughly this kind of distinction is developed by Hornstein (2000) though with rather different technical assumptions: presence/absence of accusative Case can result —

[106] These facts were pointed out to me by Ian Roberts (p.c.). See Baker, Johnson, & Roberts (1985) for an account that has a somewhat similar structure despite having little else in common with the present architecture.

in Hornstein's terms — in movement of and NP from the object to the subject θ-position, licensing reflexive readings (Hornstein does not discuss inherent reciprocals).[107] The presence of the relevant κ-property results in the two roles being distinguished, a possibility clearly permitted by these verbs types (e.g., *John washed the baby*; *The women kissed the baby*).

The view of passive offered above explains this. The move is to suggest that as object related T can lack a κ-property, it also may lack a projected/selected C-element, which otherwise serves to "shield" it from contraction with the higher T. Since these derivations individuate/index the internal role via the upper (nominative) κ, this will necessarily be distinct from v^{θ}, thus yielding the absence of the inherent reflexive/reciprocal readings for these verbs.

This is worth taking a closer look at, as it bears on the issues of what happens where in the TCG approach we've been developing here. The assumption is that "like" features in the workspace simply come to share values. So two κ features, or two ϕ features, if these are in the same workspace, then they share values. Period. (Given that we have now partitioned θ-elements into separate workspace zones).

Landau (2004) criticizes Hornstein's analysis, with which I share some assumptions (the whole story above simply stipulates absence of accusative Case) as follows. Why are Case features not potentially optional for *all* transitives? Thus the accusative κ (our "lower" C^{κ}) could be omitted with the mandatory reading then being an inherently reflexive reading, where (305)a would have to mean what (305)b does.

[107] The general idea of having a Case-distinction permit the connection between internal and external arguments is attributed to suggestions made by Howard Lasnik & Alan Munn.

(305) a. John hit
b. John hit himself

This strikes me as a reasonable question, but one that has a reasonable answer. This is, it seems to me, a bit like asking why it cannot be the case that an unaccusative verb (e.g., *arrive*) cannot end up in the syntax with an outer *v*-shell and thus manifest a structure supporting things like *John arrived the man* with some corresponding transitive or causative reading (e.g., 'John made the man arrive' or some such).

There are really two possibilities as far as I can see to address this issue. Either there is such a thing as non-compositional, not-fully-productive sort of "structure-building", or something else accounts for the lack of productivity (or promiscuity) among the decomposed bits in approaches adopting one or another view of the separation hypothesis.

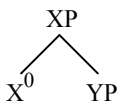
It seems fairly clear to me that differences between verbs licensing inherent reflexivity/reciprocity versus not is a *lexicon* distinction. This only means "arbitrary fact" if we presuppose a Bloomfieldian view. The question is: how does this distinction manifests in the syntax? What are the properties that must be *projected* such that we can account for the patterns that are exhibited by the relevant elements? What is being claimed here (and, as I understand things, by Hornstein as well) is that Case properties are central to how these different manifestations of a verb are projected into the syntax. Landau's objection seems to presuppose that we need to regard Case optionality in the intended sense as a matter of the workings of the syntactic system, as opposed to *consequences* for the syntactic system which could be seen to follow from alternative projection possibilities of Case/θ-properties associated with different classes of elements.

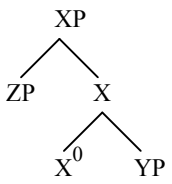
What answers the objection for unaccusatives? Clearly it must be part of the specifications of these elements that they cannot enter into a structure with a superordinate *v*-shell. Once we have made the move of doing the decomposition of the VP in this manner there's no getting the genie back in the bottle, we have to live with the somewhat more exciting ontology of the separation/decomposition view. And this, it seems, means accepting that there are some non-compositional sorts of organization involved here.

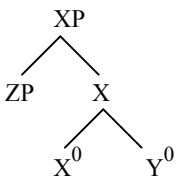
If I'm right in this line of argumentation, then connecting these sorts of "argument structure" alternations with the theory of Case (or agreement perhaps as well, and perhaps more in the functional hierarchy) then it begins to become plausible that we are looking at (as suggested earlier) paradigmatic organization, which we needn't necessarily expect to be "productive" in the manner that syntagmatic organization is. Its just a different kind of system.

This does of course presuppose that what is captured "in the lexicon" includes specifications regarding the projection possibilities *which go beyond a single phrasal projection layer*. But this is general, and not specific to the issues as they arise regarding inherently reflexive or reciprocal verbs. This is, in fact, part of the point of work elaborating on the notion of *extended* projections of the sort discussed in the work of Grimshaw (1991, 2000) and others, or so it seems to me. Given the advent of the functional projection explosion which has attended the development of minimalist theory, it is no longer conceivable that we can understand "projection" as governing the distribution of elements within a single categorial shell.

Let us ask the same question in a related domain in an analogous way that will make the issues somewhat clearer, as the issue strikes me as one worth spending some time on. Consider the following distinctions in the "lexical syntax" of various different types of elements, as conceived in the widely adopted work of Hale & Keyser (1993, 2002):

(306) a. 

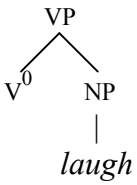
b. 

c. 

d. X^0

They distinguish between an elements which (a) take a complement, (b) take a complement and a specifier, (c) take only a specifier, and which (d) take neither a complement nor a specifier. They suggest that while these possibilities do not universally align with categories, there may be associations which predominant (e.g., they suggest that in English, the predominant realization of (a) is V, (b) is P, (c) is A, and (d) is N).

So: how do we think about "mismatches" — that is, situations in which the wrong element somehow gets associated with projection realizations which clash with its properties? For example, Hale & Keyser treat so-called unergatives (e.g., *laugh*) as realized by monadic head-complement structures of the (a)-type above, as follows:

(307) 

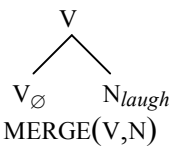
They suggest that the impossibility of such elements participating in the following transitivity alternations follows from the analysis above:

(308) a. The children laughed
b. *The clown laughed the children
(i.e., "the children laughed because of the clown)

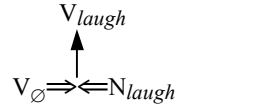
They note that this property is shared by analytic expressions *make trouble*:

(309) a. The cowboys made trouble
b. *The beer made the cowboys trouble
(i.e., the cowboys made trouble because of the beer)

Thus the transitivity alternation impossibilities are understood to follow because these elements are in their surface "object-less" form in a sense already transitive. However, H&K assume that the relevant structures such as the one above for *laugh* do not strictly speaking exist at any level of syntactic structure. Rather, they assume that there is a process of *conflation* which happens as a "concomitant" of Chomsky's MERGE. That is, there are not two operations like (310)a, but rather simply MERGE and the consequences of MERGE (conflation) for items of this particular type, as in (310)b:

(310) a. 

MERGE(V,N) + b. V_{laugh} CONFLATE(V,N)

(311) 

MERGE&CONFLATE(V,N)

So, why isn't this process totally general? What stops impossible applications involving "nominal" elements that do not fall into the unergative class? Take the nominal element *cigar*; why shouldn't it manifest the properties that *laugh* exhibits in virtue of undergoing this kind of MERGE+CONFLATION?

(312) *John cigared
(meaning perhaps: 'John had a cigar', it doesn't matter for the present point)

This would be an instance of the kind of mismatch raised above. That is, generally, once we have made some of these distinctions between predicates a matter of structure, what is it that keeps the relevant elements in their appropriate bins?

Alternative derivational routes of the sort explored above regarding passives and reflexive/reciprocal verbs are like the options of entering into a MERGE+CONFLATION derivation versus one with simply MERGE. What these alternative derivations *do* is to partition the space of structural possibilities *given a set of distinctions provided as input*. This is to say that "lexicon information is syntactically represented". The job of syntactic theory is to provide a *principled* partition of this space that captures the relevant properties of the structural alternatives *given* the properties projected by given items.

I will leave this to the side for now, observing that the suggested analysis rather large issues. I have offered a general direction for thinking about their solution, and that's about it. However, there is one final further point raised above that requires discussion.

Nothing in what I have said about these reflexive/reciprocal verbs distinguishes *between* the two classes. The relevant claims being made here are that (i) absence of κ-properties make it so we do not have local obviation, so that the external and internal roles can become associated, and (ii) that the properties of passive derivations make it so the external role will be not κ-associated, but that since the internal role will be, these two roles will be distinct. This accounts for the *similarities* of reflexive/reciprocal verbs under passivization (why these inherent readings go away for both types of verbs).

But I have nothing further to add here about how the semantics works for these cases such that we can discriminate between the effects of this kind of external/internal argument connection for the two classes (i.e., the fact that *the women kissed* does not mean 'the women kissed themselves'). I do think that the present system makes promising distinctions which can be taken as a good basis upon which to build in this respect. I will leave this as an open question, taking the current account to provide part of the final solution (the part which captures the similarities — leaving the differences up to an unspecified semantic story for now).[108]

We can, however, capitalize on the general logic deployed for reflexives to provide a schema for analyses of control phenomena. Taking note of the collection of properties shared by obligatory control and reflexivization (Hornstein 2000), I suggest here that the same general notion of C-C identification and contraction is at work.

That control predicates take C-complements is a widely held assumption in theories of control. One fairly clear source of evidence for this sort of analysis is based in the fact that, as pointed out in Landau (2000), control predicates — but never raising predicates — can be seen in many languages to manifest overt complementizers. Moreover, in languages which manifest this distinction overtly, those predicates which appear to be ambiguous between the control vs. raising type, when they manifest a complementizer, *only* manifest the control reading. So assuming these predicates to take

[108] Juan Uriagereka points out cases like "they scratched" which seem to be vague between a reflexive, reciprocal, or an implicit (distinct) object reading. This is one of many cases to address. There is also the possibility of "making" a reciprocal verb by passivizing certain ditranstives. For example, *He introduced John to Mary* versus *John was introduced*, which with a plural subject (*they were introduced*) is ambiguous between an implicit object reading (i.e. they were introduced (to the audience)) or a reciprocal (each other) reading. The present suggestion for heading into these issues is to continue to explore the appeal to sameness/difference in the form of κ/ϕ properties.

C-complements, suppose that we say that these similarly manifest absence of an individuating κ-property that we have argued to be the driving factor in inherent reflexivization.

Let us begin with a familiar contrast (Rosenbaum 1967) between raising and control:

(313) a. Dave persuaded a doctor to examine Bill
b. Dave expected a doctor to examine Bill

(314) a. Dave persuaded Bill to be examined by a doctor
b. Dave expected Bill to be examined by a doctor

These examples differ in the interpretative properties of the embedded infinitival clause, depending on active/passive voice. The b-cases are ECM/object-raising, and are basically synonymous. The a-cases, involving control, are not; they differ as to who is being persuaded (*Bill* or *a doctor*).

Another familiar contrast involves idiom chunks, which raising (a-cases) but not control (b-cases) allows (i.e., to the extent the b-cases are ok they must not be idiomatic):

(315) a. The cat seemed to be out of the bag
b. #The cat tried to be out of the bag

(316) a. John expected the cat to be out of the bag
b. #John persuaded the cat to be out of the bag

As mentioned earlier, Hornstein (2000) argues for an approach which reduces control to raising/movement. Under this view the salient difference between the two sorts of construction is simply how many θ-roles are hanging around. Whereas in raising an element moves from a θ-position to a Case position, in control elements are understood to move from θ-to-θ, picking up "θ-features" as they go. Thus, the above contrasts are

straightforwardly accounted for in terms of whether the idiom chunks, which cannot receive thematic roles without loosing their idiomaticity, have to move through a position where getting a θ-feature is avoidable.

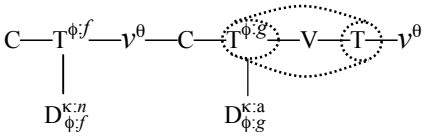
Similarly, the active/passive difference between object raising and object control shown above simply differ as to whether the passivized NP end up in a theta-position (control) versus not (raising).

Manzini & Roussou (2000) develop an idea similar in spirit to Hornstein's raising/control reduction, though with a rather different technical implementation. They suggest that in both raising and control the "moved" element is rather simply inserted into its surface position, and from there it "attracts" θ-features (conceived as aspectual features). The raising/control difference turns on simply whether a single θ-feature or more than one such feature is attracted to the "controller".

Let's consider some possible structures for some core cases. Consider first the difference between the object raising and the control manifestations of a verb like *expect*:

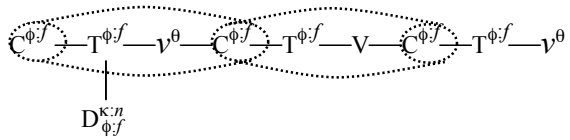
(317) a. John expected him to leave
b. John expected to leave

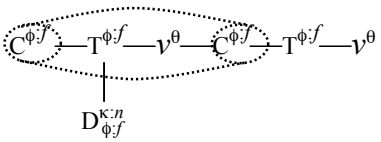
The object raising version we expect to evoke the following structure (using a shorthand notation indicating the relevant T-T identification):

(318) 

What about the control case? If control complements generally involve a C-layer, then we can view the extension of the extention of the ϕ-properties into these non-finite domains

just as we did in the case of local reflexives above, which would look as follows for subject control. We could assume either of the following:

(319) 

(320) 

Both of these would result in the extension of the matrix subject domain so that it end up in a local relation with the lower v^{θ}. The two possibilities would differ in whether we would find reason to maintain the object-related C and T elements in absence of either object-θ or the θ-less κ/ϕ properties argued to be present in object raising cases. I will not pursue this issue, though I cannot at present see a reason to maintain the more complicated structure.

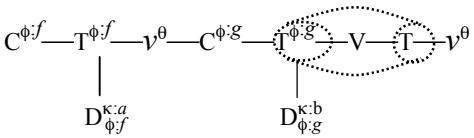
On this view, note that we would be assuming the *subject* element is itself not brought into a local relationship with the lower role, as with raising. Rather, only its ϕ-feature is. Consider in this connection the often noted lack of reconstruction effects in control (but not raising).

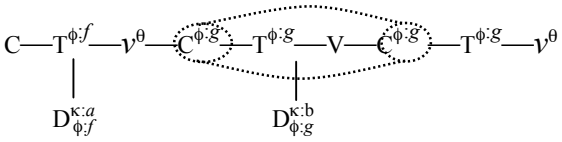
(321) a. Someone from New York is likely to win the lottery
b. Someone from New York is eager to win the lottery

As noted in May (1985), the a- and b-case above differ in whether they admit a reading with *someone* scoping low. That is, the raising (a-) case is ambiguous between meaning that some *particular* person from New York is likely to win, versus a low scope reading paraphrasable as "It is likely that someone from New York will win the lottery", where it

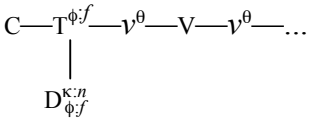
is clear that we do not have any particular person in mind. Importantly, the control (b-) case above has no such reading; it only exhibits the higher scope interpretation in which we have a particular person in mind who is eager to win. This raising/control difference follows on the assumption that computing scope requires the actual scope bearing element to be in the relevant local domain. (However, these facts as an argument for the present view should be treated with caution; see our discussion above regarding Lasnik's arguments about specificity).

Note that object control (e.g., *John persuaded Mary to leave*) now gets a parallel derivation to the one offered for object-raising, only involving C-contraction as with the subject-control cases above. Consider:

(322)  OBJECT RAISING

(323)  OBJECT CONTROL

Another possibility for analysis of control in the present terms is the equivalent of a bare-VP complement:

(324) 

Obligatory control verbs with gerundival complements might be of this sort.

(325) John tried [$_{vP}$ eating the pie]

Such cases would thus involve direct relationships between *v*'s, somewhat akin to the suggested story at the beginning of this chapter for secondary predication (*John arrived sad*).

Note that of course what has been offered in this section is just a sketch. Nonetheless, two key points emerge that will be important to pursue further. First, there can be no straightforward *direct* control/raising assimilation in the present architecture. But, second, the issues are now perhaps a bit more subtle. Given that we have reconceived movement in general as "agree-type" feature/category relationships, all localizable relations fall into this general bin in one way or another. The general appeal to agreement (ϕ) properties sketched above for a potential account of control relationships is consistent with Landau's (2000) view, but as we do not recognize a separate "movement-type" of relation, its not clear that raising and control aren't being brought closer together in terms of being subserved by the same general mechanisms (albeit in different ways). These issues need to be more carefully pursued within the TCG framework to see how things turn out, but the general format that the system makes available for analysis suggests that at least a partial control/raising unification may be feasible (so we may have a position intermediate between those advanced by Hornstein and Manzini/Roussou on the one hand, and views of the sort championed by Landau on the other). At any rate, I leave these matters for future investigation.

3.5. Clausal Unithood & *Wh*-Again,...

The discussion in the previous section is of course quite speculative. However, we noted above that positing an object-related T-element is not without precedent, nor is the general "split-VP"/"stacking" approach.

Note as well that in a recent thesis, Butler (2004) argues for an general view of *phase*-hood with roughly the kind of iterated CP-structure that our view of contraction requires. In addition to the development of his own arguments, he points to a number of other places in the literature where similar kinds of assumptions have been shown to bear fruit in syntactic analysis.[109]

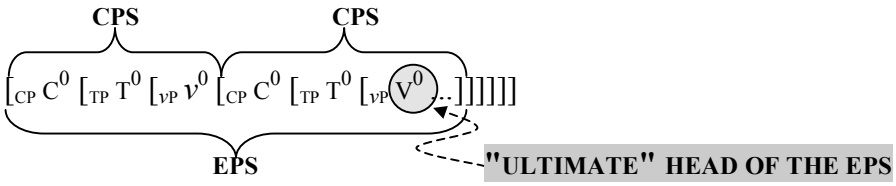
Iterated clause-internal sub-structures of this kind I have in mind were also proposed by Demuth & Gruber (1994), who distinguish between Basic Projection Sequences (BPS's) and Lexical Projection Sequences (LPS's). To avoid confusion with references to Chomsky's Bare Phrase Structure (also "BPS"), and to suggest the connection with the organization of what I earlier referred to as Core Licensing Properties (CLPs), let us refer to the sort of objects that Demuth & Gruber call Basic Projection Sequences as instead CORE PROJECTION SEQUENCES (CPSs), and to make an explicit connect to Grimshaw's (1991, 2000) proposals, call the analogue of their LPS instead an EXTENDED PROJECTION SEQUENCE (EPS).

Demuth & Gruber's proposals differ somewhat in detail from what I proposed here (or what Butler proposes for example), but the ideas are all very similar. On D&G's

[109] Butler's articulation of phases is quite detailed, and motivated by connections to a particular view of the syntax-semantics interface relevant to understanding quantification, scope, and the like, building on ideas of Beghelli & Stowell (1997), Belletti (2001, 2003), Jayaseelan (2001) among others. I refer readers to Butler's thesis for further discussion and references.

view their BPS's iterate to form LPS's. An LPS is simply a series of BPS's with a kind of ultimate lexical/thematic head at the bottom of the lowest BPS.

We can illustrate the idea as it is relevant to what has been suggested here with reference to our proposed iterated CP-TP-*v*/VP structures as follows, using our new terminology for the relevant units (CPS/EPS):[110]

(326) 

EPS's then might be seen as a series of CPS's which bottom out in a major lexical category. It may be that the typical case is that an EPS is at most two CPS's (as in (326) above) though further issues not examined here may force us to conclude otherwise (the structure of ditransitives, causatives, and many other matters).

Of course, we would like to have some idea of what makes a series of CPS's "hang-together" to form an EPS. There are a couple of things we might say on this score which require further investigation but which seem like the right sort of ideas. First, recall from our discussion of contraction and node-identification in §3.3.2 the idea the following general schema that we used to explain how features might stay "fixed" to positions in the output structure, despite being implicated in lower domains. What we

[110] The reasons for my changing terminology are not just to avoid confusions with references to Chomsky's theory of phrase structure. Demuth & Gruber actually understand their BPS's/LPS's to bottom out in a *thematic* element, with the higher BPS's understood to be *athematic*, so there is only one of these "thematic" BPS's in a given LPS on their view. See their paper for discussion and interesting analysis of compound tenses in Bantu languages. Given the fit of this general idea with Split-VP ideas I will not be importing this aspect of their story. Here lowest elements of both sub-units (now: CPS's), that is *v* and V, are both thematic elements. And, while *v* has sometimes been entertained as a member of the "functional" category inventory, I will here regard it as essentially functional (perhaps "semi-lexical").

require is perhaps some further property which could be seen to run *through* a series of CPS's, in a sense serving to hold them together. This would be something analogous to the way ϕ-properties have been suggested to "hold together" a series of distinct elements forming our CPS's.

One plausible candidate, which we might or might not wish to view as "part of th syntax" in any direct way, are variables and quantifiers associated with *eventualities* (the "*e*" variable casually referred to in earlier discussion). It seems plausible to say that something like Kratzer's (1996) event identification might serve to unify two such CPS structures into a single "EPS" (i.e., some way in which the event variables in the two separate domains are linked/identified).

And, just as there are properties marking the edges of CPS's, we might examine other properties that might serve to group our CPS's into larger units. Finiteness and Force (Rizzi 1997) might be such properties. Realized as categories, these could be elements which serve to mark off our larger stretches of structure equivalent to the traditional clause.

However these matters are pursued, I will close this chapter with reference to one last class of facts that our view of SCM *does not* predict unless we take the idea of recursive structure into the clause in the way suggested above. In particular I am referring to the Fox examples discussed in Chapter Two. Consider:

(327) a. √ [Which of the papers that he$_1$ gave Mary$_2$] did every student$_1$ [$_{vP}$ __√__ [ask her$_2$ to read __*__ carefully?

b. * [Which of the papers that he$_1$ gave Mary$_2$] did she$_2$ [$_{vP}$ __*__ [ask every student$_1$ to revise__*__ ?

Instead of moving to the 'edge' of *v*P, here we have a uniform approach of C-C contraction which serves to bring the *wh*-element into the local configurations that are necessary to account for the possible and impossible interpretations in these cases.

However, recall as well from Chapter 2 the following cases (Legate 2000):

(328) a. √ [At which of the parties that he$_1$ invited Mary$_2$ to] was every man$_1$ [$_{vP}$ √ [introduced to her$_2$ *?

b. * [At which of the parties that he$_1$ invited Mary$_2$ to] was she$_2$ [$_{vP}$ * [introduced to every man$_1$ *?

(329) a. √ [At which charity event that he$_1$ brought Mary$_2$ to] was every man$_1$ [$_{vP}$ √ [sold to her$_2$ *?

b. * [At which charity event that he$_1$ brought Mary$_2$ to] was he$_2$ [$_{vP}$ * [sold to every woman$_1$ *?

These cases, as noted in our earlier discussion, are problematic for Chomsky's (1999) view of phases as just C and *v*, as they seem to involve passives.[111] These facts are incompatible with the present view as well. Recall our account of passives denied the presence of an object-related C-element to derive the "suppression" of *v*-θ (its splicing-out of the workspace in virtue of T-T contraction).

However, there is another line of argumentation available to us given the general perspective of domains. Note that the cases above are passives of ditransitives. What we have claimed in terms of the stacking of independent thematic domains with independent functional structure (perhaps the extreme of the "stacking" view) ought perhaps to hold of indirect objects as well. What is required to capture the facts above is some position that

[111] Legate provides similar examples with unaccusatives, though to build the cases she requires a special sort of unaccusative that takes more than one internal argument. I'm uncertain about the classification of the verb she uses (see her paper for the cases), but should the argument turn out to be ok, I think the story I run in the main text for the passive case will carry over (should it turn out to be sustainable!).

the *wh*-phrase must move which is below the subject (e.g., so the pronoun within it may be bound by *every man* as in the a-cases above) but *above* the indirect object (so as to avoid obviation in the a-cases above).

Suppose that this position is the edge of the indirect object's domain, conceived uniformly with the subject and direct objects cases. To get the facts, we need an outer C-type layer surrounding the prepositional phrase, i.e.:

(330) C—T—*v*—C—T—V—C?—P—D
subj. direct obj. indirect obj

However, an extra C-layer may not actually be required. What are prepositional elements anyway? In classical X-bar theoretic feature decompositions of categories they were typically regarded as negatively specified for both "n" and "v" properties. Interestingly, Pesetsky & Torrego (2004), in addition to positing an object-related T-element, also suggest the possibility that (at least some) prepositions may be a "type of T-element". They hint at a connection in terms of connections between the elements in terms of the functions the play in the semantics of time and space, but the interesting point from the present perspective is the possibility of their belong to a general type including T.

Note as well that there is often discussion of prepositional/complementizer type relations with, for example, worries about whether prepositional-looking elements that introduce clausal structures (e.g., *before John ate the pizza*) are really of the P or C type (e.g., Lasnik & Saito 1991 argue for a C-type analysis of cases of this sort).

A number of interesting possibilities arise here that deserve more attention that I can devote here, but let me make another general point. The implementation of the TCG ideas that have been pursued in the present chapter suggest a research strategy aimed at

re-evaluating how we partition syntactic classes. The actual "labels" we have deployed in our discussion and analyses are classical ones, and so might naturally evoke some suspicion (which is not unreasonable, e.g., "there aren't any clause internal *complementizers*!!").[112]

But in the spirit of pointing a direction for such potential reclassifications, the present point is that it is not entirely crazy to think that C, T, and P might be fruitfully viewed as members of a larger class.

At any rate, if the general line of thinking is on the right track, we might then discover classes of the "P-type" which would relate to C, T, or perhaps to both types via the node-identification mechanism developed here. We might find reason to attribute different features to this super-class to discriminate possible/impossible identifications along the lines that have been suggested for the subject/object situation and for cross-clausal relations above (e.g., two C-wh's cannot identify, etc.).

For Legate's data above, if this is correct then it might be that the *wh*-element moves to the edge of the indirect-object domain (C-"P" identification). This view would be interesting as well to explore with respect to passivization and ditransitives.

However, these are topics for another day.

3.6. Conclusions: The Take-Home Message of TCG

Here are the core points to take home. First: the analyses that TCG supports with respect to core "classical" cases of SCM, in particular raising-to-subject and *wh*-movement of the typical clause-edge-to-edge variety, should be taken to be the central result.

[112] Though, again, see Bulter (2004), Belleti (2001, 2003), Jayaseelan (2001).

If nothing else about this dissertation is correct, the ability of the present system to provide a basic platform for understanding SCM *without* postulating what we called M-features is something that should be attempted to be maintained in any further pursuit of this enterprise. The effect of this architecture is to reduce a wide-ranging general class of superficially non-local dependencies to local ones. In the local domains themselves, we see only independently motivated "core licensing properties" (CLPs) at work in establishing the key relationships.

What allows us to dispense with M-features (i.e., "the EPP") is the mechanics of node identification. Generally, the idea is that intermediate positions of the SCM sort exist only because of (i) the existence of core local-type relation hold in the matrix clause and (ii) the general fact that lower "like-elements" constitute informational supersets of matrix contexts. Crucial to executing this intuition is the WS/O-distinction, which allows us to separate-out local computation of relationships from the resultant/derived output in a way that allows non-local relations in the output to be maintained within a local workspace.

What I have sketched in the present chapter is one possible implementation of a more general set of ideas. However, I have suggested that some of the specifics yield an interesting story, both about some particulars and in the general form of the answer that is provided regarding clause-structure and dependency relationships.

We are now in a position to address as well an issue that I have left unaddressed throughout — the assumption that these derivations work "top-down". Numerous technical aspects of the presentation hinge on this assumption, but the general logic of the SCM story is where the distinction is clearest. The necessity of a top-down derivation

goes hand-in-hand with the denial of M-features. If there are no such properties, then it is not possible in general to have an element associate directly with intermediate positions, nor is it possible to "move" to them. Note as well that offering a format within which an eliminative agenda regarding EPP-type properties in favor of a view where local licensing is handled in terms of CLPs distinguishes the present approach from the views of TAG that we discussed in Chapter 2, where elementary tree-local movement is not generally understood in this way.

Needless to say this is just the setup for an analytical investigation into the wider range of cases, and cross-linguistic differences, that have been taken to motivate EPP-features and the like. What I have offered here is a start on what strikes me as the most serious challenge — getting rid of non-CLPs as motivators for intermediate movements.

We could, presumably, attempt to motivate a "bottom-up" view along one of the lines mentioned in Chapter 2 (e.g., non-feature-driven movement to avoid crashing the derivation), with a *wh*-element starting in a base position, but this bottom/embedded domain will not itself be a "phase" on the node-identification and contraction view — as the relevant "like element" won't arrive until the top of the next highest clause. Moreover, *other* like elements (e.g., the V selecting the embedded clause) will arise first, and in virtue of the anti-recursion restriction on the workspace, will force a splicing out of the intervening material. This *could* be taken to motivate movement directly to a superordinate VP-adjoined type position, skipping a lower C, but this would seem to be contrary most of the empirical evidence reviewed in Chapter 2. Also, this would be a mixed-view system of displacement, which would involve both standard movement and

the node-identification mechanism.[113] So, it turns out on the present view that top-down structure expansion and eliminating M-features go hand-in-hand.

Returning to TAG approaches, one might ask whether the adjoining/substitution mechanics could be put to work in ways similar to what has been developed here in appealing to the reduced Brody-type structures that collapse intra-phrasal projection-level distinctions. This seems possible, though I have not investigated the matter.

The general outlook here on the relationship between TCG, TAG, and the MSO-type systems discussed in Chapter One is that they constitute a family of closely related approaches. The introduction of the WS/O-distinction at the outset of this work is designed to form a general background context within which various aspects of the different approaches might be mixed/matched and then tested against the facts of human language. What I have offered here is an outline of one such approach.

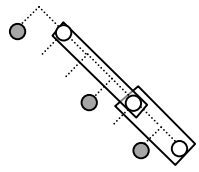
And there are numerous issues which have not even been scratched. In order to concentrate on the key properties of interest here regarding recursion, the node-identification view of lowering/copying, and the like, issues regarding head-movement and modification have been completely avoided. This is a serious omission, and should be one of the first areas to be developed in any continued thinking on this general approach.

[113] Note that we did suggest something close to such mixed view in our discussion of A' to A-position movement earlier in the chapter — but the details were developed to bring this case within the general logic of identifying properties on the dominance sequence as a way of deriving copying of elements in the output.

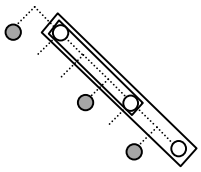
3.7. Closing

I wish to close with some general points and a few open questions. First, consider again our earlier discussion regarding chain structures, where we suggested that Chomsky's (1995) "technical options" (including our third previously undiscussed option) regarding chains are not in fact different possible ways of talking about the same thing, but rather simply different things. In our dominance encoded feature-relations, suggested that the following three schemas are actually different chain types:

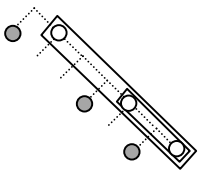
(331)



Binding/Control (connections between A-relations)



A'-relations



Some A-relations

Simple A-relations were understood to involve two or three node connections, linking up a κ-marked nominal with θ by means of ϕ-properties. These sequences I suggested, might be themselves linkable, via node-identifications involving the top-most members of each such sequence. This was the basic shape that was suggested for local reflexives (perhaps also inherently reflexive verbs), and was offered as a schema for the structure of control relations as well. This is the picture of chains that we see in the left-most schema in (331). Local A'-relations, it was suggested are best understood as local WH-κ relationships, which are themselves connected to θ via ϕ-relationships. This is the picture in the middle schema in (331). Finally, the rightmost schema above manifests the structure I have attributed to (e.g.) passives. There a T-T contraction was argued to

splice-out intervening *v*, leaving it in the workspace unassociated with any other element (this was suggested to be the present but unexpressed external role). The picture then is of a lower κ/ϕ-θ relationship which is connected to a higher position (e.g., nominative T).

There is a central idea in play here that our discussion has not adequately touched upon. The mechanics we have been working with assume a single formal ordering dimension, that admits to branching, along which feature-licensing relationships are characterized. The ideas just discussed regarding different "constituency structures" for chain relationships has been key.

This is just a sketch. Putting the system to work in more in-depth and rigorous analysis is what is now required. What the present work has accomplished is to set the stage for a novel type of approach that I have argued has the right general structure to provide principled accounts of SCM phenomena at the least, and perhaps has consequences for other concerns.

In general, I wish to stress again here that I believe it is best to view the present approach along with the others that have been discussed alongside it (TAG and MP/MSO appraoches) as a family of closely related ideas. The efforts here have brought out some differences between these approaches, but the hope is that they have also been brought somewhat closer together.

Consider again both the Chamorro agreement facts and the S-V inversion cases from Spanish discussed in Chapter 2:

(332) ¿ Qué pensaba **Juan** [que le había dicho **Pedro** [que había publicado **la revista**]]]
what thought Juan that him had told Peter that had published the journal
'What did John think that Peter had told him that the journal had published?'

(333) Hafa s***in***angani-n Juan as Dolores [*t* ni m***in***alago'ña [*t* pära un-taitai *t*]]?
WHAT? WH[OBJ2].*tell Juan* OBL *Dolores* COMP WH[OBL].*want*-AGR FUT WH[OBJ].AGR-*read*
"What did Juan tell Dolores that he wants you to read?"

I noted in Chapter Two that the Spanish facts at least have been subject to some controversy. In this connection I mentioned the work of Baković (1995), who documents the following dialect variation:

(334) a. No inversion with any *wh*-phrases (Suñer 1994)
b. Inversion with argument *wh*-phrases only (Torrego 1984; Suñer 1994)
c. Inversion with all but reason *wh*-phrases (*por qué*/"why") (Goodall 1991a,b)
d. Inversion with all *wh*-phrases in matrix clauses; all but reason *wh*-phrases in subordinate clauses (Baković's survey)
e. Inversion with all but reason *wh*-phrases in matrix clauses; only argument *wh*-phrases in subordinate clauses (Baković's survey)
f. Inversion with argument *wh*-phrases in matrix clauses; no inversion in subordinate clauses (Baković's survey)

The general conclusion of Baković's research into these matters is that there is a scale which to a first approximation tracks a hierarchy of *wh*-elements, ordered on a more-to-less "referential" (or perhaps "argumental") scale. Dialects variation appears to be systematic if Baković is right. First, it is possible to have a dialect with no inversion at all. However, if there is inversion and if it is allowed with *wh*-elements position X in a more-to-less referential/argumental continuum, then it is allowed with all the others higher in the hierarchy. Moreover, there appears to be a "subset" relationship, with respect to the matrix/subordinate distinction, such that embedded clause inversion possibilities with respect to this hierarchy of *wh*-elements is always a subset of what is possible in the matrix.

Interestingly, Chung (1994) reports that the local agreement facts in Chamorro are *optional* for referential *wh*-elements, taking the relevant classes of elements to be those picked out in the work of Cinque (1991) (see also Pesetsky 1987 on "D-linking").

What sort of mechanics should be deployed to account for their syntactic properties? Petesky (1987) offers a story under which relies on a division between "regular" *wh*-movement, and a kind of unselective binding by a matrix Q-morpheme, of the sort offered in the work of Baker (1979) to provide a grounding for the D-linked/non-D-linked distinction. Suppose something like this is correct. There are then *two* types of relationships that in principle can account for the connection between a *wh*-element and potentially distant (embedded) thematic information. A local SCM mechanism, and a potentially long-distance kind of (semantic?) relation.

But why should this be? Why should there be *two*? If it can happen long-distance, why not always that way? Or why not always linked-local?

Note that our WS/O-distinction offers a reasonable place to hang this difference — we could understand unselective binding to be a semantic relation (perhaps of the linking sort we discussed for logophors earlier, realizing an unselective binding relation) holding over output structures, and also have the edge-to-edge linked-local style relation as mediated by the syntactic workspace. My suspicion is that the right way to approach these issues should involve a close examination of the learnability of the distinctions. In general our view here has been of the narrow syntactic computation as itself constituting the interface between the lexicon and a PF/LF output structure. The learnability of core local relations ought to fall within a correspondingly local view of where learners find the information needed to acquire grammar (something along the lines of Lightfoot's Degree-

Zero "plus a little"; Lightfoot 1989). If the ideas here regarding node-identification and contraction are generally on track, learners needn't have access to anything more than roughly (traditional) clause-sized objects to acquire the relevant distinctions that pertain as well to linked-local/SCM-type relationships, as these I have argued fall out from the basic mechanics. But there is still a serious problem to be faced, one with two facets: (i) its just not true that *all* dependencies are reducible to local domains, that this is the case is what motivates views like that introduced by Pesetsky (1987), as mentioned above, and (ii) its just not true that matrix level generalizations carry over to embedded domains. Regarding (ii), what seems to be the case is that something like Ross's (1973) "Penthouse Principle". Ross offered the following metaphor "whose truth is borne out in myriad cases of Real Apartment Life":

(335) **The Penthouse Principle:** More goes on upstairs than downstairs.

Life, it seems, is always more exciting in the Penthouse; anything happening downstairs is sure to be a trendy copy of things that have already been done upstairs. So we needn't strain our capacity for decoding metaphors, Ross translates into terms more linguistic:

(336) No syntactic process can apply only in subordinate clauses.

This seems to be borne out by many of the SCM-effects discussed in Chapter Two. Inversions, *wh*-copying, and local agreement all seem to be required "at the top" if they hold in embedded domains, and where they hold in embedded domains, they must hold "all the way down" (typically no domains may be skipped). These matters strike me as important to investigations concerned with Degree-*N* learnability and may be a route that might help us to better understand SCM-type effects. Recall that *wh*-copying shows up as

well in L1-acquisition of English (see §2) which is a case where the adult/target grammar does not generally permit such copying. This suggests that learners are not necessarily conservative in the sense of waiting for positive evidence to alter their grammars to allow multiple PF-spell-outs of this kind. And if children do not have access to direct negative evidence that these constructions are *not* part of the target grammar, then something else must be in play to allow them to converge on the correct target. The following questions then can frame further inquiry into these matters:

(337) a. Is the Penthouse Principle *true*?
b. If it is true, *why*? *In virtue of what*?
c. How does whatever underlies this principle effect the presence/absence of SCM-type effects?

Part of the answer to questions a/b I believe lies in how we understand the domains that learners have access to in order to find evidence to sort out where their target language lies in the UG-governed array of possibilities.

Question-c is then framed in our TCG approach (deploying the WS/O-distinction) in terms of how the mechanisms of unselective binding or the like arise, making truly long(er)-distance relationships possible.

I will leave these matters here, noting in closing that the general structure of our account makes room for three kinds of dependencies: (i) those that are purely local, (ii) those that are linked-local, and (iii) those that are non-local. Further, I have suggested here a way that natural language grammars might reduce the (ii)-type to the (i)-type. How the (iii)-type fits in to this view is a job for future inquiry.

Last, consider some remarks from Fodor (1977) regarding grammars and derivational directionality, which will allow us to sum-up some of the key ideas discussed in this dissertation in a more general way:

> We might suppose that we could isolate the issue of directionality by comparing two (imaginary) grammars, G_1 and G_2, which are identical except that the rules of G_1 are the inverses of the rules of G_2 (i.e., the input to each rule of G_1 is the output of the corresponding rule of G_2, and vice-versa), and the order of application of the rules in G_1 is the inverse of the order of application of the corresponding rules in G_2. The set of structural representations constituting the derivation of a given sentence would be identical in both grammars, but these structures would be generated in reverse order. But now notice that where G_1 has a deletion rule, the corresponding rule in G_2 will be an insertion rule; where G_1 has a rule moving a constituent to the left, the corresponding rule of G_2 will move that constituent to the right.

G_1		Derivation		G_2
step 1: *d* is moved to the left of *bc*	↓	abcd	↑	step 2: *d* is moved to right of *bc*
step 2: *a* deletes before *d*	↓	adbc	↑	step 1: *a* is inserted before *d*
	↓	dbc	↑	

> The difference is direction is inevitably accompanied by a difference in the operations that particular rules perform. [We might] consider the possibility of constraining the rules of a grammar so that they can perform certain types of operations but not others. We might then be able to decide between the grammars G_1 and G_2 on this basis. But let us temporarily abstract from this issue by supposing that the rules of G_1 and G_2 all conform to the definition of a possible rule of grammar. Could there, nevertheless, be some reason for preferring either G_1 or G_2? The consensus of opinion, even among those who agree about almost nothing else, appears to be that there would be no significant difference between the two grammars — as long as the old confusion of grammars with psychological models of speech production and perception is avoided.

The important bit in this stretch of passage is the observation that "difference in direction is inevitably accompanied by a difference in the operations that particular rules perform". It is such a difference between bottom-up and left-to-right/incremental assembly that Phillips (1996, 2003) exploits in that the incremental system seems to allow us to make reference to "units" that we need for analysis but which are unavailable in a bottom-up characterization.

We can excise the following main issues from Fodor's passage:

(338) **INDEPENDENCE:** considering ordering of operations in grammar and performance-theoretic systems as independent conceptually, what considerations might lead us to consider one or another possible view of the ordering of combinatory operations in the grammar?

(339) **CORRESPONDENCE:** suppose we were to have strong reasons for thinking one or another global ordering for syntactic structure assembly was correct, how ought we think about correspondence relationships to operations in parsing and production — how ought we think about GRAMMAR as embedded in "time"?

The sub-part of the passage from Fodor (1977) above contrasting two toy grammars G and G' ends with the fairly reasonable assertion that it is difficult to see how we could find reason to think there was any real difference between such grammars.

Setting to the side for the moment the present work and the work of Phillips and others, the state-of-affairs regarding such questions about directionality in more recent theory is even a bit more difficult, if anything, than it was at the time of Fodor's writing. Following the above-quoted discussion, Fodor goes on to contrast two ways of understanding a rule like *wh*-movement implicated in (340) and (343), one including a rule of "*wh*-fronting" and another including a rule of "*wh*-backing". The latter rule would map (342) to (341) and (345) to (344), while the former would do the reverse:

(340) Who do you expect to murder Jemima?

(341) *Q* You *Pres* expect [*WH+pro* murder Jemima] WH-FRONTING ↑ WH-BACKING
(342) *Q WH+pro* you *Pres* expect [murder Jemima] ↓

(343) Who do you expect to murder?

(344) *Q* You *Pres* expect [*PRO* murder *WH+pro*] ↓ WH-FRONTING ↑ WH-BACKING
(345) *Q WH+pro* you *Pres* expect [*PRO* murder]

If direction of derivation does not fundamentally matter then there should be no non-trivial differences between these derivations. Of course, any present-day comparison

between two such visions of transformational operations differs quite a bit from the theoretical situation that obtained in the days before the assumption of *structure preservation* (Kimball 1972, Emonds 1985).[114] Without structure preservation the rule of *wh*-backing appears to require a bit more in the way of additional assumptions that *wh*-fronting does.[115] That is, without traces or some kind of marker or variable or the like designating the internal position to which *wh*-backing would displace the *wh*-element, some additional mechanism would be required to filter out misapplications.

In contrast, the rule of *wh*-fronting has a salient target: the "edge" of the sentence.[116] Of course, in Fodor's example things are rigged in favor of *wh*-fronting given the presence of an abstract *Q*-morpheme (a "trigger" morpheme[117]), but with no corresponding marker for the base (trace/copy or thematic) position. If we embrace structure preservation, then our contrast between *wh*-fronting and *wh*-backing looks like this, where we simply switch the direction of the arrows on our "movement" notation (assuming *copies* rather than traces for the moment):

---

[114] Kimball was, to my knowledge, the first to observe the interest of a restriction stated in terms of the *kinds* of objects produceable in principle by phrase-structure rules and the *kinds* of objects produceable by transformational operations. The idea of structure preservation really caught on, however, following the work of Emonds (1970, 1985), who used this as a general restriction which enabled him to differentiate between structure-preserving and non-structure-preserving operations (e.g., his root transformations were of the latter type). The development of this notion with respect to traces/copies put this notion on even more solid ground, though it fell somewhat into the background as a general kind of constraint on operations and the structure of the grammar. See Newmeyer (1986) for a good discussion of this history.

[115] Fodor observes this, but includes no discussion of "structure-preservation" in this context.

[116] That is, even in absence of "trigger morphemes" or a designated landing site for such transformations, *wh*-fronting appears to require less information to apply correctly. This is all assuming, of course, that the transformational approach to these matters is the right way to go. There exists context-free grammars (e.g., GPSG) which incorporate rules with complex ("slash") symbols which make the kind of asymmetry Fodor is pointing to irrelevant. Such grammars, like the context-free rules we examined above, can operated trivially either from terminals to "S" or from "S" to terminals. Whether or not the kind of asymmetries Fodor is pointing to here still can be found in more recent models of syntax is one way of stating the major theme of this work.

[117] Or, for more recent versions of the "trigger morpheme" idea, consult almost any recent article which has the words "minimalism" or "derivation" or "checking" in the abstract key-words line.

(340) Who do you expect to murder Jemima?

(341)' *Q WH+pro* you *Pres* expect [*WH+pro* murder Jemima] ***wh*-backing**

(342)' *Q WH+pro* you *Pres* expect [*WH+pro* murder Jemima] ***wh*-fronting**

(343) Who do you expect to murder?

(344)' *Q WH+pro* you *Pres* expect [*PRO* murder *WH+pro*] ***wh*-backing**

(345)' *Q WH+pro* you *Pres* expect [*PRO* murder *WH+pro*] ***wh*-fronting**

In more modern terms, we have a difference between an operation which raises the *wh*-element, leaving a null copy or a trace, versus an element which lowers a null copy/trace. Or, more neutrally, we have some minimal formal way of establishing a dependency between two positions in a structure, and attendant (interpretative?) processes which sort out where to pronounce and interpret what.

Given structure preservation, in other words, the issue of derivational directionality becomes quite a bit foggier than it was at the time Fodor's book was published (27 years ago). Consider some further remarks which Fodor makes on *wh*-backing, which reveal the present point about structure preservation nicely (despite including no explicit mention of structure preservation as such; **bold** emphasis mine):[118]

> *WH*-backing knows which noun phrase to move, but how it could know where to move it? What indicates that there is an appropriate gap for the interrogative pronoun to move into at the end of [(343)] but not at the end of [(340)]? **A gap, after all, is just a nothing**. **Two words are adjacent that otherwise would not have been.** The information that determines where there is a gap, and which gap has to be filled by the *WH*-Backing transformation, is information about the deep structures of these sentences, and about other transformations that do and do not apply in their derivations. [...] for the *WH*-Backing transformation to apply correctly, it would need information about structures in the derivation of a sentence which are 'deeper' than the one on which it operates, i.e. structures which are generated only AFTER *WH*-Backing itself has applied. By contrast, the standard *WH*-Fronting rule is self-sufficient; it can apply correctly without 'looking ahead' to later stages of the derivation. The reason for this difference is that *WH*-Fronting parallels, while *WH*-

[118] This passage also raises questions about "look-ahead" which have become relevant in much current derivational syntactic theory.

> Backing opposes, the direction of flow of information between structures like [(341)] and [(342)]. Before *WH*-Fronting applies, the position of the interrogative pronoun indicates its syntactic and semantic role in the sentence. But this information is lost when all interrogative pronouns are moved into the same position at the front of the sentence. Thus, [(341)] contains more information (in this respect) than [(342)]; [(341)] contains enough information to determine [(342)], but [(342)] does not contain enough to determine [(341)]. Other transformations (e.g., Passive, Particle Movement) apparently involve no loss of information and hence determine a unique output when applied in either direction. An asymmetry in information content between two adjacent structural representations in a derivation thus gives some content to the notion of the direction of the rule (p110-112).

With the notion of structure preservation in place, the informational asymmetry Fodor points to with respect to *wh*-fronting versus *wh*-backing no longer holds — at least, not obviously.

But these aspects of Fodor's discussion regarding informational asymmetries serve to bring some issues into the foreground rather clearly for our purposes here, even if they rely on now outdated assumptions. Note the part of the above passage regarding the issue of *wh*-backing "opposing" the "flow of information" and the matter of information loss in derivations. These issues become relevant — though in a different way — in the context of current derivational minimalist syntactic theory if we consider the notions of "phases" and the like within MSO-systems, which exhibit what we have called 'expand/contract dynamics'. So we now have encountered a possible motivation — albeit quite general and abstract — for pursuit of one or another global directionality for syntactic derivations: the potential existence of *informational asymmetries*.

In this work we have built an approach to grammar which capitalizes on such asymmetries, suggesting that what underlies SCM-type effects is an informational superset relation (subsumption) which holds between positions hosting "intermediate" positions and matrix positions where core licensing properties are related. This is thus an argument for a particular direction of derivation that focuses on purely competence-

theoretic issues. What of the matters of *correspondence* then — that is, the relation between grammar and parser?

I have not addressed matters of performance in this work at all, but to wind up this closing discussion, at least the following two points are of interest for further pursuit in the present framework. First, as mentioned in Chapter One, the TCG mechanics make available a grammar-based conception of how "displaced" elements may be buffered in a sense (kept in the workspace) so that they may be integrated in some lower domain. But our above discussion regarding the possibility of other mechanisms that handle such dependencies differently raises the question of how these two different sets of mechanisms — one that is WS-local, and one that functions over the output structure — may or may not interact in on-line processing. Second, the top-down perspective as it has been wed here with our workspace ordering and distinctness constraints, suggests some ways of beginning to think about how categories/features and ordering information might be mapped to "time". Chomsky's term "phases" may turn out to be particularly apt in the sense that we might investigate ways in which categories/features could be understood to have a *duration* — that is, a time-course in which these properties are "active" in on-line processing. The general view of categories/features and ordering pursued here suggests that sameness/difference may matter for determining an abstract sort of "chain constituency" that is important to understanding local structures and how they may overlap (or not). This may be potentially translatable onto a time-axis in a way that preserves the informational groupings that same/different properties have been suggested to effect along the dominance ordering. This is a fairly general, somewhat vague

suggestion, to be sure, but I believe something along these lines may be the key to understanding the various ways that we might understand grammar as relating to time.

References

Abels, K. 2003 *Successive Cyclicity, Anti-Locality, and Adposition Stranding*. Doctoral dissertation, University of Connecticut, Storrs.

Abney, S. & M. Johnson. 1991. Memory Requirements and Local Ambiguities for Parsing Strategies. *Journal of Psycholinguistic Research* 20(3):233-250.

Adger, D. & G. Ramchand 2003. Merge vs. Move: Wh-Dependencies Revisited. Ms. University of London.

Aoun, J. 1985. *The Grammar of Anaphora*. Cambridge, MA: MIT Press.

Asudeh, A. 2004. Resumption as Resource Management. Doctoral dissertation. Stanford University.

Bach, E. 1986. The Algebra of Events. *Linguistics & Philosophy* 9:5-16

Baker, C. L. 1970. Notes on the Description of English Questions: The Role of an Abstract Question Morpheme. *Foundations of Language* 6, 197-219.

Baker, M., K. Johnson, & I. Roberts. 1987. Passive Arguments Raised. *Linguistic Inquiry* 20: 219-251

Baković, E. 1995. A Markedness Subhierarchy in Syntax: Optimality and Inversion in Spanish. Ms. Rutgers University.

Barss, A. 2001. Syntactic Reconstruction Effects. In C. Collins & M. Baltin (eds.) *The Handbook of Contemporary Syntactic Theory*, Oxford: Blackwell, 670-696

Beghelli, F. & T. Stowell. 1997. Distributivity and Negation: the syntax of *each* and *every*. In A. Szabolcsi (ed.) *Ways of Scope Taking*, Dordrecht: Kluwer, 71-107.

Belletti, A. 1988. The Case of Unaccusatives. *Linguistic Inquiry* 19:1-34.

Belletti, A. 2001a. Agreement Projections. In C. Collins & M. Baltin (eds.) *The Handbook of Contemporary Syntactic Theory*, Oxford: Blackwell, 483-510.

Belletti, A. 2001b. 'Inversion' as Focalization. In A. Hulk & J.-Y. Pollock (eds.) *Subject Inversion in Romance and the Theory of Universal Grammar*, Oxford: Oxford University Press, 60-90.

Belletti, A. 2003. Aspects of the low IP area. In Luigi Rizzi (ed.) *The Structure of IP and CP*: *The Cartography of Syntactic Structures*, Oxford: Oxford University Press.

Benmamoun, A. 1998. Spec-Head Agreement & Overt Case in Arabic. In D. Adger et al. (eds.) *Specifiers: Minimalist Approaches*, Oxford: Oxford University Press, 110-125.

Berwick, R. 1985. *The Acquisition of Syntactic Knowledge*. Cambridge, MA: MIT Press.

Berwick, R. & A. Weinberg 1985. *The Grammatical Basis of Linguistic Performance*. Cambridge: MIT Press.

Bittner, M. 1994. *Case, Scope, & Binding*. Dordrecht: Kluwer.

Bitter, M. & K. Hale. 1996. The Structural Determination of Case and Agreement. *Linguistic Inquiry* 27(1), 1-68.

Bobaljik, J. 1995. Morphosyntax: The Syntax of Verbal Inflection. Doctoral dissertation. MIT.

Bobaljik, J. 2002. Floated Quantifiers: Handle with Care. In L. Cheng & R. Sybesma (eds.) *The Second State of the Article Book.* Berlin: Mouton de Gruyter.

Boeckx, C. 2000. EPP Eliminated. Ms. University of Connecticut.

Boeckx, C. 2003. *Islands & Chains*. Amsterdam: John Benjamins.

Boeckx, C. & N. Hornstein. 2003. Reply to 'Control is not Movement'. *Linguistic Inquiry* 34/2: 269-280.

Boeckx, C. & N. Hornstein. 2004. Movement under Control. *Linguistic Inquiry* 35/3: 431-452.

Bošković, Z. 2002. A-Movement and the EPP. *Syntax* 5: 167-218.

Bresnan, J. 1971. Contraction and the Transformational Cycle. Ms. MIT, Cambridge, MA.

Brody, M. 2000. Mirror Theory: Syntactic Representation in Perfect Syntax. *Linguistic Inquiry* 31(1) 29-56.

Brody, M. 2003. *Towards an Elegant Syntax*. London: Routledge.

Butler, J. 2004a. Phase Structure, Phrase Structure, & Quantification. Doctoral dissertation, University of York.

Butler, J. 2004b. On Having Arguments and Agreeing: A Semantic EPP. *York Papers in Linguistics* 1, 1-27.

Castillo, J.C., J. Drury, & K. K. Grohmann. Localizing Procrastinate. Ms. University of Maryland, College Park.

Castillo, J.C., J. Drury, & K. K. Grohmann. 1999. Merge over Move and the Extended Projection Principle. In S. Aoshima, J. Drury, & T. Neuvonen (eds.) *University of Maryland Working Papers in Linguistics* 8, 63-103.

Castillo, J. C. & J. Uriagereaka. 2000. A Note on Successive Cyclicity. In M. Guimarães, L. Meroni, C. Rodrigues, & I. San Martin (eds.) *University of Maryland Working Papers in Linguistics* 9, 1-13.

Chametzky, R. 1996. *A Theory of Phrase Markers and the Extended Base*. Albany: SUNY Press.

Chametzky, R. 2000. *Phrase Structure: From GB to Minimalism*. Oxford: Blackwell.

Chomsky, N. 1955/1975. *The Logical Structure of Linguistic Theory*. New York: Plenum.

Chomsky, N. 1965. *Aspects of a Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. 1973. Conditions on Transformations. In S. Anderson & P. Kiparsky (eds.) *A Fetschift for Morris Halle*, New York: Holt, Reinhart, & Winston, 232-286.

Chomsky, N. 1980. *Rules and Representations*. New York: Columbia University Press.

Chomsky, N. 1981. *Lectures on Government & Binding*. Dordrecht: Foris.

Chomsky, N. 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA: MIT Press.

Chomsky, N. 1986a. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, N. 1986b. *Knowledge of Language: Its Nature, Origin, & Use*. New York: Praeger.

Chomsky, N. 1993. A Minimalist Program for Linguistic Theory. In K. Hale & S. J. Keyser (eds.) *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, Cambridge, MA: MIT Press, 1-52.

Chomsky, N. 1994. Bare Phrase Structure. *MIT Occasional Papers in Linguistics*, 5.

Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. 1998. Minimalist Inquiries: The Framework. *MIT Working Papers In Linguistics*.

Chomsky, N. 1999. Derivation by Phase. Ms., MIT.

Clements, 1975. The Logophoric Pronoun in Ewe: Its Role in Discourse. *Journal of West African Languages* 10: 141-177.

Collins, C. 1997. *Local Economy*. Cambridge, MA: MIT Press.

Collins, C. 2000. Eliminating Labels. Ms. Cornell University.

Davidson, D. 1967. The Logical Form of Action Sentences. In N. Rescher (ed.) *The Logic of Decision and Action.* U. of Pittsburgh Press.

Demuth, K. & J. Gruber. 1994. Constraining XP Sequences. Ms. Brown University and UQAM.

De Villiers, J., T. Roeper, and A. Vainikka. 1990. The acquisition of long-distance rules. In L. Frazier and J. de Villiers (eds.) *Language Processing and Language Acquisition*, Dordrecht: Kluwer, 257-297.

den Dikken, M. & A. Szabolcsi. 2002. Islands. In L. Cheng & R. Sybesma (eds.) *The Second State of the Article Book.* Berlin: Mouton de Gruyter.

DiScuillo, A. & E. Williams. 1987. On the Definition of Word. Cambridge, MA: MIT Press.

Drury, J. 1998a. Root-First Derivations: Atomic Merge & The Coresidence Theory of Movement. Ms. University of Maryland, College Park.

Drury, J. 1998b. The Promise of Derivations: Atomic Merge & Multiple Spell-Out. *Groninger Arbeiten zur germanistischen Linguistik* 42, 61-108.

Drury, J. 1999. The mechanics of π-derivations. In S. Aoshima et al. (eds.) *UMWPiL* 8.

Drury, J. 2000. Some Thoughts on Generalizing Transmission. In M. Guimarães, L. Meroni, C. Rodrigues, & I. San Martin (eds.) *University of Maryland Working Papers in Linguistics* 9, 1-13.

Drury, J. 2005. The Order of Form. Ms. Georgetown University, Washington D.C.

Du Plessis, H. 1977. Wh-Movement in Afrikaans. *Linguistic Inquiry* 8, 723-726.

Emonds, J. 1970. *Root and Structure-Preserving Transformations*. Indiana University Linguistics Club Publication.

Emonds, J. 1976. *A Tranformational Approach to English Syntax*. New York: Academic Press.

Emonds, J. 1985. *A Uniform Theory of Syntactic Categories*. Dordrecht: Foris.

Epstein, S., E. Groat, R. Kawashima, & H. Kitahara. 1998. *A Derivational Approach to Syntactic Relations*. New York: Oxford University Press.

Felser, C. 2004. Wh-Copying, phases, and successive cyclicity. *Lingua* 114, 543-574

Fodor, J. D. 1977. *Semantics: Theories of Meaning in Generative Grammar*. Hassocks, Eng.: Harvester Press.

Fodor, J. 1983. *Modularity of Mind*. Cambridge, MA: MIT Press

Fodor, J. 2000. *The Mind Doesn't Work That Way*. Cambridge, MA: MIT Press.

Fox, D. 1998. *Economy and Semantic Interpretation*. Cambridge, MA: MIT Press.

Frampton, J. & S. Guttman. 2000. Agreement is Feature-Sharing. Ms. Northwestern University.

Frank, R. 1992. *Syntactic Locality and Tree Adjoining Grammar: Grammatical, Acquisition, and Processing Perspectives*. Doctoral dissertation. University of Pennsylvania.

Frank, R. 2002. *Phrase Structure Composition and Syntactic Dependencies*. Cambridge, MA: MIT Press.

Frank, R. & A. Kroch 1995. Generalized Transformations and the Theory of Grammar. *Studia Linguistica* 49, 103-151.

Fukui, N. 1986. A Theory of Category Projection and its Applications. Doctoral dissertation, MIT.

Fukui, N. & M. Speas. 1986. Specifiers and Projections. In N. Fukui *et al*. (eds.) *MIT Working Papers in Lingusitics 8.*

Grimshaw, J. 1991. Extended Projections. Ms. Brandeis University.

Grimshaw, J. 1997. Projection, Heads, and Optimality. *Linguistic Inquiry* 28, 373-422.

Groat, E. & J. O'Neil 1996. Spell-Out at the LF Interface. In W. Abraham, S. Epstein, H. Thrainsson, & J.-W. Zwart (eds.) *Minimal Ideas*, Amsterdam: John Benjamins, 305-327.

Grohmann, K. K. 2003. *Prolific Domains*. Amsterdam: John Benjamins.

Grohmann, K. K., J. Drury, & J. C. Castillo. 2000. No More EPP. In R. Billerery & B.D. Lillegaugen (eds), *WCCFL 19: Proceedings of the 19th West Coast Conference on Formal Linguistics*. Cascadilla Press.

Guimarães, M. 1999. Phonological Cascades & Intonational Structure in Dynamic Top-Down Syntax. Ms. UMCP

Guimarães, M. 2004. *Derivation and Representation of Syntactic Amalgams*. Doctoral dissertation. UMCP.

Hale, K. & S. J. Keyser 1993. On Argument Structure and the Lexical Expression of Syntactic Relations. In K. Hale & S. J. Keyser (eds.) *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, Cambridge, MA: MIT Press.

Hale, K. & S. J. Keyser 2002. *Prolegomenon to a Theory of Argument Structure*. Cambridge, MA: MIT Press.

Hallé, F., R. A. A. Oldeman, & P. B. Tomlinson. 1978. *Tropical trees and forests: An architectural analysis.* Berlin: Springer-Verlag.

Heck, F. and G. Müller. 2000. Successive cyclicy, long-distance superiority, and local optimization. *Proceedings of WCCFL 19*, 101-114.

Henry, A. 1995. *Belfast English and Standard English: Dialect Variation and Parameter Setting*. Oxford: Oxford University Press.

Herburger, E. 2000. *What Counts*. Cambridge, MA: MIT Press.

Hiemstra, I. 1986. Some aspects of Wh-Questions in Frisian. *North-Western European Language Evolution* (NOWELE) 8, 97-110.

Higginbotham, J. 1985. On Semantics. *Linguistic Inquiry* 16/4: 547-593.

Higginbotham, J. 1988. Contexts, Models, and Meanings: A Note on the Data of Semantics. In R. Kempson (ed.) *Mental Representations: The Interface Between Language and Reality*, Cambridge: Cambridge University Press, 29-48.

Hornstein, N. 2001. *Move! A Minimalist Theory of Contrual*. Oxford: Blackwell.

Huang, J. 1982. *Logical Relations in Chinese and the Theory of Grammar*. Doctoral dissertation, MIT.

Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Jackendoff, R. 1977. *X-bar Syntax*. Cambridge, MA: MIT Press.

Jayaseelan, K. A. 2001. IP-internal Topic and Focus Phrases. *Studia Linguistica* 55: 39-75.

Jenkins, L. 2000. *Biolinguistics*. Cambridge: Cambridge University Press.

Johnson, K. 1991. Object Positions. *Natural Language & Linguistic Theory* 9: 577-636.

Kayne, R. 1975. *French Syntax*. Cambridge, MA: MIT Press.

Kayne, R. 1984. *Connectedness and Binary Branching*. Dordrecht: Foris.

Kayne, R. 1989. Facets of Romance Past Participal Agreement. In P. Beninca (ed.) *Dialect Variation and the Theory of Grammar*, Dordrecht: Foris, 85-103.

Kayne, R. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Kayne, R. & J.-Y. Pollock 1978. Stylistic Inversion, Successive Cyclicity, and Move NP in French. *Linguistic Inquiry* 9: 595-621.

Kimball, J. 1972. Cyclic and Linear Grammars. In J. Kimbal (ed), *Syntax and Semantics 1*, New York: Seminar Press, pp 63-80.

Klima, E. 1964. Negation in English. In J. Fodor & J. Katz (eds.) *The Structure of Language*. Englewood Cliffs, NJ: Prentice-Hall: 246-323.

Koizumi, M. 1993. Object Agreement Phrases and the Split-VP Hypothesis. In J. Bobaljik & C. Phillips (eds.) *Papers on Case & Agreement I, MITWPL* 18.

Koizumi, M. 1995. Phrase Structure in Minimalist Syntax. Doctoral dissertation. MIT.

Kratzer, A. 1996. Severing the External Argument from its Verb. In J. Rooryck & L. Zaring (eds.) *Phrase Structure and the Lexicon*, Dordrecht: Kluwer.

Kulick, S. 2000. Constraining Non-Local Dependencies in Tree Adjoining Grammar: Computational and Linguistic Perspectives. Doctoral dissertation. UPenn.

Landau, I. 2000. *Elements of Control*. Dordrecht: Kluwer.

Landau, I. 2004. Movement out of Control. *Linguistic Inquiry* 34(3), 470-498.

Lasnik, H. 1995. A Note on Pseudogapping. In *Papers on Minimalist Syntax MITWPL* 27, 142-163.

Lasnik, H. 1999. *Minimalist Analysis*. Oxford: Blackwell.

Lasnik, H. & M. Saito. 1991. *Move-α.* Cambridge, MA: MIT Press.

Lasnik, H. & J. Uriagereka. Forthcoming. *A Course in Minimalist Syntax*. Oxford: Blackwell.

Lightfoot, D. 1989. The Child's Trigger Experience: Degree-0 Learnability. *Behavioral & Brain Sciences* 12: 321-334.

Manzini, R. & A. Roussou. 2000. A Minimalist theory of A-movement and Control. *Lingua* 110, 409-447.

May, R. 1985. *Logical Form*. Cambridge, MA: MIT Press.

McCloskey, J. 2000. Quantifier Float in *Wh*-Movement in an Irish English. *Linguistic Inquiry* 31:1, 57-84.

McDaniel, D. 1989. Partial and Multiple *Wh*-movement. *Natural Language and Linguistic Theory* 7, 565-604.

McDaniel, D., B. Chui, & T. Maxfield. 1995. Parameters for Wh-Movement Types: Evidence from Child English. *Natural Language and Linguistic Theory* 13, 709-753.

McKinnon, R. & L. Osterhout. 1996. Constraints on Movement Phenomena in Sentence Processing: Evidence from Event-Related Brain Potentials. *Language & Cognitive Processes* 11, 495-524.

Muysken, P. 1982. Parameterizing the notion "Head". *Journal of Linguistic Research* 2, 57-75.

Nunes, J. 1995. The Copy Theory of Movement and Linearization of Chains in the Minimalist Program. Doctoral dissertation. University of Maryland.

Parsons, T. 1990. *Events in the Semantics of English*. Cambridge, MA: MIT Press.

Pesetsky, D. 1982. *Paths & Categories*. Doctoral dissertation. MIT.

Pesetsky, D. 1987. *Wh*-in-Situ: Movement and Unselective Binding. In E. Reuland & A. G. B. ter Meulen (eds), *The Representation of (In)definiteness*. Cambridge, MA: MIT Press.

Pesetsky, D. & E. Torrego. 2001. T-to-C Movement: Causes and Consequences. In M. Kenstowicz (ed.) *Ken Hale: A Life in Language*, Cambridge, MA: MIT Press, 355-426.

Pesetsky, D. & E. Torrego. 2002. Tense, Case, and the Nature of Syntactic Categories. Ms. MIT & U. of Massachusetts Boston [in J. Gueron & J. Lecarme (eds.) *The Syntax of Time*, Cambridge, MIT Press].

Phillips, C. 1996. *Order and Structure*. Doctoral dissertation. MIT.

Phillips, C. 2003. Linear Order and Constituency. *Linguistic Inquiry* 34, 37-90.

Postal, P. 1974. *On Raising*. Cambridge, MA: MIT Press.

Postdam, E. & J. Runner 2003. Richard Returns: Copy Raising and Its Implications. *Proceedings of CLS* 2001.

Quine, W.V.O. 1940. *Mathematical Logic*. Harvard University Press.

Reinhart, T. 2000. The Theta System: Syntactic Realization of Verbal Concepts. *OTS Working Papers in Linguistics*.

Reinhart, T. & E. Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24: 657-720.

Reuland, E. & M. Everaert. 2001. Deconstructing Binding. In C. Collins & M. Baltin (eds.) *The Handbook of Contemporary Syntactic Theory*, Oxford: Blackwell, 634-669.

Resnik, P. 1992. Left Corner Parsing & Psychological Plausibility. *Proceedings of the 14th International Conference on Computational Linguistics* (COLING '92). Nantes, France.

Rizac, M. 2004. *Elements of Cyclic Syntax*: *Agree and Merge*. Doctoral Dissertation. University of Toronto.

Rogers, A. 1971. Three kinds of physical perception verbs. *Papers from the Eighth Regional Meeting of the Chicago Linguistics Society*, Chicago: CLS, 303-315.

Rosenbaum, P. 1967. *The Grammar of English Predicate Complement Constructions*. Cambridge, MA: MIT Press.

Ross, J. R. 1973. The Penthouse Principle and the Order of Constituents. *Papers from the Comparative Syntax Festival* (CLS 9).

Runner, J. 1995. *Noun Phrase Licensing and Interpretation.* Doctoral dissertation. University of Massachusetts.

Runner, J. To Appear. The Accusative Plus Infinitive Construction in English. In B. Hollebrandse and R. Goedemans (eds.) *The Syntax Companion*. Oxford: Blackwell.

Sabel, J. 2000. Partial Wh-Movement and the Typology of Wh-Questins. In U. Lutz, G. Müller, A. von Stechow (eds), *Wh-Scope Marking*. John Benjamins, Amsterdam and Philadelphia, pp 409-446.

Saddy, D. 1991. Wh-scope Mechanisms in Bahasa Indonesia. In L. Cheng & H. Demirdache (eds), *MIT Working Papers in Linguistics 15*. MITWPL, Cambridge MA: pp 183-218.

Sauerland, U. 1995. The Lemmings Theory of Case. Paper presented at the 7th Student Conference in Linguistics (SCIL 7), University of Connecticut.

Schein, B. 1992. *Plurals and Events.* Cambridge, MA: MIT Press.

Sells, P. 1987. Aspects of Logophoricity. *Linguistic Inquiry* 18, 445-479.

Shieber, S. 1986. *An Introduction to Unification-Based Approaches to Grammar*. CSLI: University of Chicago Press.

Takahashi, D. 1994. *Minimality of Movement*. Doctoral dissertation. University of Connecticut.

Terada, H. 1999. Successive Cyclicity and Incremenality. *English Linguistics* 16/2: 243-274.

Tesneire, L. 1959. *Elements de Syntaxe Structurale*. Klincksieck, Paris.

Thornton, R. 1990. Adventures in Long-Distance Moving: The Acquisition of Complex Wh-Questions. Doctoral dissertation, University of Connecticut.

Torrego, E. 1984. On Inversion in Spanish and Some of its Effects. *Linguistic Inquiry* 15: 103-129.

Ura, H. 1998. Checking, economy, and copy-raising in Igbo. *Linguistic Analysis* 28: 67-88.

Uriagereka, J. 1998. *Rhyme & Reason*: *An Introduction to Minimalist Syntax*. Cambridge, MA: MIT Press.

Uriagereka, J. 1999. Multiple Spell Out. In S. Epstein & N. Hornstein (eds.) *Working Minimalism*, Cambridge, MA: MIT Press, 251-282.

Uriagereka, J. 2002a. *Derivations: Exploring the Dynamics of Syntax*. London: Routledge.

Uriagereka, J. 2002b. Spell Out Consequences. Ms. UMCP.

van Riemsdijk, H. 1998. Categorial Feature Magnetism: The Endocentricity and Distribution of Projections. *Journal of Comparative Germanic Linguistics* 2: 1-48.

Vermeulen, R. 2002. Multiple Nominative Constructions in Japanese. In M. Cazzloi-Goeta et al. (eds.) *Fifth Durham Postgraduate Conference in Theoretical and Applied Linguistics: Conference Proceedings* 224-233.

Watanabe, A. 1995. The Conceptual Basis of Cyclicity. In R. Pensalfini & H. Ura (eds.) *MITWPL 27*, 269-291.

Weinberg, A. 1999. A Minimalist Theory of Human Sentence Processing. In S. Epstein & N. Hornstein (eds.) *Working Minimalism*, Cambridge, MA: MIT Press, 282-314.

Williams, E. 1994. *Thematic Structure in Syntax*. Cambridge, MA: MIT Press.

Zwart, C. J.-W. 1993. *Dutch Syntax: A Minimalist Approach*. Doctoral dissertation. University of Groningen.

Zwart, C. J.-W. 1996. Shortest Move vs. Fewest Steps. In W. Abraham, S. Epstein, H. Thrainsson, & J.-W. Zwart (eds.) *Minimal Ideas*, Amsterdam: John Benjamins, 305-327.