ABSTRACT

Title of Dissertation: THE COPY THEORY OF MOVEMENT AND LINEARIZATION OF CHAINS IN THE MINIMALIST PROGRAM

Jairo Morais Nunes, Doctor of Philosophy, 1995

Dissertation directed by: Juan Uriagereka
Associate Professor, Department of Linguistics

This dissertation is concerned with movement operations within the Minimalist Program (Chomsky 1995). Exploring the copy theory of movement, it focuses on two issues: (i) why can traces not be phonetically realized?; and (ii) what is the theoretical status of the operation Move in a system where syntactic objects are derivationally assembled? I propose that a chain cannot surface with more than one link phonetically realized because it cannot be linearized in accordance with Kayne's (1994) Linear Correspondence Axiom. Assuming that the head of a chain and its trace(s) are nondistinct copies, they induce violations of the irreflexivity and asymmetry conditions on linear order, canceling the derivation. Deletion of all but one chain link in the phonological component (Chain Reduction) is forced upon nontrivial chains in order to permit their linearization. The choice of the links to be deleted is determined by economy considerations regarding the elimination of formal features in the phonological component. Assuming that only the chain link which is in the checking
domain of a head H is affected by a checking relation with H, the head of a chain will always have fewer formal features (if any) to be eliminated in the phonological component than its trace(s). Deletion of traces for purposes of linearization is thus more economical than deletion of the head of a chain because it requires fewer subsequent applications of deletion of formal features. As for the status of Move, I propose that it is not an operation of the computational system, but is rather a description of the interaction of the independent operations Copy, Merge, Form Chain, and Chain Reduction. Evidence for this proposal is provided by instances of "sideward movement", where a given constituent C of a syntactic object K is copied and merges C with a syntactic object L, unconnected to K. Under this analysis of movement, the linearization of chains in the phonological component constrains sideward movement in such a way that it makes it possible to subsume the core properties of parasitic gap and across-the-board extraction constructions under the properties of standard movement.
THE COPY THEORY OF MOVEMENT AND
LINEARIZATION OF CHAINS IN THE
MINIMALIST PROGRAM

by

Jairo Morais Nunes

Dissertation submitted to the faculty of the Graduate School
of The University of Maryland in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
1995

Advisory Committee:

Associate Professor Juan Uriagereka, Chairman/Advisor
Professor Norbert Hornstein
Professor David Lightfoot
Professor Stephen Crain
Assistant Professor Bonnie Dorr
DEDICATION

"Viver, o senhor sabe, é etc."
(Guimarães Rosa)

To Alvana
ACKNOWLEDGEMENTS

When I started thinking about what to write in this acknowledgement section, I had the impression that it would be longer than any chapter in this dissertation. During my five years as a PhD student at the University of Maryland, I have had the pleasure of interacting with some of the best linguists in the field and making very good friends (among them). I will not abuse the patience of the reader with all the things I would like to say, but I would like to stress my gratitude to all the many people who contributed directly or indirectly to making this journey a very enriching experience.

I would first like to thank Juan Uriagereka and Norbert Hornstein. It has been a great privilege for me to work closely with them. Attempting to explore the — many times opposite — theoretical approaches suggested by them has been one of the most exciting intellectual challenges I have faced. My deep thanks for the many hours I stole from their time and for the enthusiasm with which they entertained my ideas. I do not want to blame them for the content of what follows in this dissertation, but I was certainly very influenced by them in attempting to look for the "interesting" questions which are waiting to be asked.

I would also like to acknowledge the great influence that David Lightfoot's work and classes had in my formation as a linguist. My warmest thanks to him for his guidance in some historical work that I did and for his constant encouragement throughout the years. Thanks also to Stephen Crain and Bonnie Dorr for their comments on an earlier version of this dissertation.
The ideas to be discussed in the following chapters grew up from a short paper that I wrote inspired by Richard Kayne's seminar on antisymmetry of syntax, which was taught in the department in the Fall of 1993. Many thanks for his comments and suggestions on that early version.

Many thanks also to the other members of the department and visiting professors with whom I have interacted over the years: Luigi Burzio, Paul Gorrell, David Lebeaux, Ian Roberts, John Stonham, Ross Thornton, and Amy Weinberg.

Among my colleagues, Ellen Thompson deserves special mention. Throughout these years, I have constantly benefitted from her valuable comments and suggestions on my linguistic adventures. I am deeply grateful for her academic support and friendship.

I am also thankful to Mark Arnold for comments on much of the material presented here and for his continuous encouragement.

Special thanks to Li-Ling Chuang, Sonia Cyrino, Marcela Depiante, Ricardo Echepare, Maria Cristina Guidorizzi, Ana Maria Martins, Maria Elena Moll, Susan Powers, and Spyridoula Varlokosta.

Thanks also to Juan Carlos Castillo, Elena Herburger, Joel Hoffman, Carol Mason, Viola Miglio, Allan Munn, Kyartan Ottósson, Maria Soledad Rojas, Cristina Schmitt, and Sungki Suh.
Many thanks to Kathi Faukingham for almost making me forget that university bureaucracy exists.

Outside linguistics, I would like to thank my dear sister Maninha, for her friendship and support. Thanks also to my friends Aaron, Lee, Lynn, Maria, Samad, and Yaha. Special thanks goes to little Ian for having reminded me of things I had long forgotten.

Back home in Brazil, I would like to thank my professors of generative syntax Carlos Franchi, Charlotte Galves, Mary Kato, and Fernando Tarallo, for their encouragement. Very special thanks goes to Mary Kato for her friendship, guidance and support (although I admit that I have to work very hard to keep up with her "corujice").

Six years ago in Brazil, I took a syntax course with Eduardo Raposo, who encouraged me to pursue my doctoral studies abroad. I am grateful for his support.

I am also very thankful to my family and friends in Brazil for their solid friendship "mesmo que o tempo e a distância digam não".

The final version of this dissertation was written while I was a visiting assistant professor and postdoctoral Mellon fellow at the Linguistics Department of the University of Southern California. I would like to thank the department for the very friendly atmosphere which allowed me to finish the dissertation, without having to cope with problems of adaptation to a new environment. Special thanks to Joseph Aoun, Hajime Hoji, Jean-
Roger Vergnaud, and Maria Luisa Zubizarreta. Thanks also to the students of my Advanced Syntax course; their skepticism or enthusiasm helped me balance the tone of the dissertation.

I am very thankful to Nancy for having made my moving to Los Angeles so smooth.

During these five years, I have received financial support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), and from the Linguistic Department and the Graduate School of the University of Maryland. Their support is greatly appreciated.

In the beginning of this section, I promised not to go on forever with my acknowledgements. It is specially difficult to keep this promise in my thanks to Alvana. The difficulties inherently associated with being a graduate student were much lighter due to her emotional support and companionship. I cannot imagine how these five years would be without her presence by my side. I would like to thank her from the bottom of my heart.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter I: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter II: Theoretical Framework</td>
<td>9</td>
</tr>
<tr>
<td>II.1. Introduction</td>
<td>9</td>
</tr>
<tr>
<td>II.2. The Minimalist Program: The General Picture</td>
<td>11</td>
</tr>
<tr>
<td>II.3. The Numeration</td>
<td>14</td>
</tr>
<tr>
<td>II.4. Derivational Cost of the Operations of the Computational System</td>
<td>15</td>
</tr>
<tr>
<td>II.5. Spell-Out</td>
<td>17</td>
</tr>
<tr>
<td>II.5.1. The Version of Spell-Out to Be Adopted in This Dissertation</td>
<td>21</td>
</tr>
<tr>
<td>II.6. Feature Checking, Procrastinate, and Strong Features</td>
<td>23</td>
</tr>
<tr>
<td>II.6.2. Strong Features as Hybrid Features</td>
<td>30</td>
</tr>
<tr>
<td>II.7. Domains, Clausal Structure, and Locality</td>
<td>33</td>
</tr>
<tr>
<td>II.7.1. Domains</td>
<td>33</td>
</tr>
<tr>
<td>II.7.1.2. Some Revisions</td>
<td>37</td>
</tr>
<tr>
<td>II.7.2. The Structure of VP</td>
<td>42</td>
</tr>
<tr>
<td>II.7.3. Agr-Projections</td>
<td>43</td>
</tr>
<tr>
<td>II.7.4. Locality</td>
<td>50</td>
</tr>
<tr>
<td>II.7.5. Domains, Feature Checking, and 0-assignment</td>
<td>59</td>
</tr>
<tr>
<td>II.8. Phrase Structure and the Operation Merge</td>
<td>60</td>
</tr>
<tr>
<td>II.9. Some Definitions</td>
<td>64</td>
</tr>
<tr>
<td>II.9.1. Term</td>
<td>64</td>
</tr>
<tr>
<td>II.9.2. Domination</td>
<td>66</td>
</tr>
<tr>
<td>II.9.3. Adjunction Structures Revisited</td>
<td>67</td>
</tr>
<tr>
<td>II.10. The Copy + Merge Theory of Movement</td>
<td>70</td>
</tr>
<tr>
<td>II.10.1. Derivational Cost of the Copy Operation</td>
<td>75</td>
</tr>
<tr>
<td>II.10.2. Deriving Procrastinate</td>
<td>77</td>
</tr>
<tr>
<td>II.10.3. C-command</td>
<td>79</td>
</tr>
<tr>
<td>II.10.4. Nondistinctiveness of Copies</td>
<td>85</td>
</tr>
<tr>
<td>II.10.5. Last Resort</td>
<td>89</td>
</tr>
</tbody>
</table>
### Chapter IV: Sideward Movement and Linearization of Chains at PF  

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV.1. Introduction</td>
<td>330</td>
</tr>
<tr>
<td>IV.2. Parasitic Gaps</td>
<td>342</td>
</tr>
<tr>
<td>IV.2.1. General Approach</td>
<td>349</td>
</tr>
<tr>
<td>IV.2.2. S-Structure Licensing</td>
<td>366</td>
</tr>
<tr>
<td>IV.2.3. Anti-c-command Requirement</td>
<td>369</td>
</tr>
<tr>
<td>IV.2.4. Island Effects</td>
<td>386</td>
</tr>
<tr>
<td>IV.2.4.1. Parasitic Gaps, Wh-phrases in Situ, and Island Effects</td>
<td>393</td>
</tr>
<tr>
<td>IV.2.4.2. Categorial Restrictions</td>
<td>399</td>
</tr>
<tr>
<td>IV.2.4.3. &quot;Nonreferential&quot; NPs</td>
<td>402</td>
</tr>
<tr>
<td>IV.2.4.4. Summary</td>
<td>404</td>
</tr>
<tr>
<td>IV.3. Across-the-Board Extraction</td>
<td>405</td>
</tr>
<tr>
<td>IV.4. Sideward Movement and Case Theory</td>
<td>414</td>
</tr>
<tr>
<td>IV.5. Conclusion</td>
<td>422</td>
</tr>
</tbody>
</table>
A fundamental property of human languages is that elements are interpreted in positions different from the ones where they are phonetically realized. In (1), for instance, *John* occupies the subject position, but is interpreted as the semantic complement of *arrested*.

(1) John was arrested.

Within the framework of the Principles and Parameters Theory (see Chomsky (1981, 1986b) and Chomsky and Lasnik (1993)), which will be assumed throughout this dissertation, this displacement property of human languages is captured by means of a movement operation in the following way: given a structural configuration, an element may move to a different position in order to satisfy certain grammatical requirements, leaving behind a coindexed trace. A trace is a phonetically unrealized category which has properties relevant for the interpretation of the moved element. The moved element and its trace(s) thus form a sort of discontinuous object, which is referred to as a (nontrivial) chain. The relevant structure
underlying the sentence in (1) is therefore as in (2), where the chain \( CH = (John_i, t_i) \) is formed.

\[
(2) \quad [ \text{John}_i [ \text{was} [ \text{arrested} t_i ] ] ]
\]

A considerable amount of research within the Principles and Parameters Theory has been devoted to properly characterize the properties of movement, traces, and chains. In the recent developments of the Principles and Parameters Theory which have culminated with the proposal of a Minimalist Program for linguistic theory (see Chomsky (1993, 1994, 1995:chap. 4) and Uriagereka (forthcoming)), these questions arise anew, in face of the elimination of much of the rich theoretical apparatus previously available. In particular, only the interface levels LF and PF are assumed, and LF objects are taken to be built from the features of the lexical items which feed a derivation (the inclusiveness condition; see Chomsky (1994:8, 1995:228)). Under the standard theory of movement stemming from Chomsky (1973), however, traces and their indices are not part of the initial array, but are introduced in the course of the derivation. Thus, a new theory of movement operations is called for in the Minimalist Program.

Addressing similar issues, Chomsky (1993) incorporates the "copy theory of movement" into the Minimalist Program. According to the copy theory, a trace (also called a layered trace under this view) is a copy of the
moved element which is deleted in the phonological component, but is
available for interpretation at LF. Besides being compatible with the
inclusiveness condition, the copy theory of movement has the advantages
of allowing Binding Theory to be stated solely in LF terms and dispensing
with the operation of reconstruction (see Chomsky (1993) and section III.3
below for discussion).

With the abandonment of D-Structure as a syntactic level of
representation, Chomsky (1994) develops a "bare" X'-Theory, according to
which the operation Merge builds syntactic objects from the lexical items of a
given array in a derivational fashion. The notion of a movement operation
is then slightly changed in order to be compatible with these assumptions.
Under this view, the operation Move is a complex operation comprised of
(at least) three suboperations (see Chomsky (1993:22, 1994:fn.13, 1995:250)): (i)
a suboperation of copying; (ii) a suboperation of merger; and (iii) a
suboperation of trace deletion. In addition, Move should be followed by an
operation of chain formation relating the relevant copies.

There are several conceptual inadequacies in this picture. First, if no
explanation for why "lower" copies must be deleted in the phonological
component is provided, the notion of a trace as a primitive is being
reintroduced. To put it more generally, the simplest — therefore most
desirable — version of the copy theory of movement should take traces and
heads of chains to be subject to the same principles and be accessible to the
same operations. Any difference between heads of chains and traces, such as phonetic realization, for instance, should follow from independently motivated properties of the computational system, rather than being idiosyncratic properties of the chain links themselves.

Deletion of traces (lower chain links) becomes even more enigmatic, if we adopt the core Minimalist assumption that economy considerations play a role in determining the set of admissible derivations in a given language or universally. Consider the relevant structure of the sentence in (1) under the copy theory of movement, which is given in (3) below. The derivation of (4a) from (3) requires one application of deletion targeting the lower copy of John, apparently being less economical than the derivation of (4b), which involves no application of deletion. Thus, if the derivations of (4a) and (4b) were to be compared in terms of economy, we would wrongly predict that the derivation of (4b) should rule out the derivation of (4a). In other words, at first sight, it appears that a derivation in which no chain link is deleted should be preferred over a derivation in which traces are deleted.

(3) \[ \text{[ John [ was [ arrested John ] ]]} \]

(4) a. John was arrested.
   b. *John was arrested John.
Another conceptual problem with the computational system as proposed in Chomsky (1994, 1995:chap. 4) is that Merge is taken to be an operation in its own right in certain cases, and a suboperation (of Move) in other cases. In an optimal system, we should in principle expect Merge to have the same theoretical status in every computation. Finally, as is emphasized by Brody (1995), who pursues a representational version of the Minimalist Program, if chain formation and Move express the same type of relation, a theory which contains both notions is redundant.

In this dissertation, I develop a strictly Minimalist version of the copy theory of movement which overcomes the conceptual problems raised above and has a broader empirical coverage than the versions developed in Chomsky (1993, 1994, 1995:chap. 4).

I propose that the fact that a chain cannot have more than only link overtly realized (see (4b)) follows from Kayne's (1994) Linear Correspondence Axiom (LCA), according to which the linear order of a PF sequence is determined by asymmetric c-command. Under the assumption that the two copies of John in (3) are "the same" (in a sense to be made precise), no linear order can be established in accordance with the LCA. Given that the verb was in (3), for instance, asymmetrically c-commands and is asymmetrically c-commanded by "the same" element, namely John, the LCA should require that was precede and be preceded by John, violating the asymmetry condition on linear order. Hence, the unacceptability of the
sentence in (4b) under the relevant derivation. Put simply, deletion of all but one link is forced upon a given chain CH in order for the structure containing CH to be linearized in accordance with the LCA. The derivations of (4a) and (4b) therefore cannot be compared for economy purposes, because only the former yields a PF object.

The next question to be addressed then is why it is the case that only traces are deleted for purposes of linearization, but not heads of chains. In other words, why does the structure in (3) surface as (5a) and not as (5b), given that both structures can be linearized in compliance with the LCA?

(5)  
   a. John was arrested.

The basic idea that I will pursue is that heads of chains become "different" (in a sense to be specified) from their trace(s) in the course of the derivation due to their participation in checking relations. For the sake of illustration, consider the Case-feature of John in the course of the derivation of (3), as shown in (6) below.

(6)  
   a. [ was [ arrested John-CASE ] ]
   b. [ John-CASE [ was [ arrested John-CASE ] ] ]
   c. [ John [ was [ arrested John-CASE ] ] ]
Assume that the Case-feature of the upper instance of John in (6b) is eliminated after being checked against the finite T head, as represented in (6c). Since Case-features are not PF objects, deletion of the lower copy of John in (6c) yields a legitimate PF object, whereas deletion of the upper copy does not; hence the contrast between (5a) and (5b). This rough idea will be technically implemented in terms of economy considerations regarding the elimination of formal features in the phonological component.

As for the other conceptual problems mentioned above (the dual status of Merge and the redundancy between Move and chain formation), I propose that Move is not a primitive operation of the computation system. It is rather a mere description of the interaction of the independent operations Copy, Merge, Chain Reduction (deletion of chain links for linearization purposes), and Form Chain. Thus, Merge is always an operation, and applications of Copy and Merge are not redundant with Form Chain. In particular, the computational system may copy a given constituent C of a syntactic object K and merge C with a syntactic object L, which has been independently assembled and is unconnected to K, as illustrated in (7). Under the standard assumption that chain links must be in a c-command relation, the instances of C in (7b) cannot form a chain.
It will be shown that Chain Reduction (deletion of chain links for purposes of linearization) constrains instances of "sideward" movement as illustrated in (7) in such a way that it makes it possible to subsume the core properties of parasitic gap and across-the-board extraction constructions to the properties of standard movement. To the extent that it succeeds, this version of the copy theory of movement will be conceptually and empirically supported.

The dissertation is organized as follows. In Chapter II, I discuss the general framework which will be assumed in the following chapters. Chapter III addresses the issue of why traces cannot be phonetically realized. Chapter IV provides empirical support to the copy theory of movement developed in chapters II and III, by analyzing parasitic gap and across-the-board extraction constructions in terms of sideward movement. Finally, some conclusions are presented in Chapter V.
CHAPTER II

THEORETICAL FRAMEWORK

II.1. Introduction

This dissertation assumes the general framework of the Principles and Parameters Theory (see Chomsky (1981, 1986b) and Chomsky and Lasnik (1993)), with particular attention to its recent developments within the Minimalist Program (see Chomsky (1993, 1994, 1995:chap. 4) and Uriagereka (forthcoming)). This chapter is devoted to presenting the specific theoretical framework I will be assuming in the following chapters.

In addition to presenting an overview of the conceptual and technical innovations introduced in Chomsky (1993, 1994, 1995:chap. 4), I will propose alternative approaches and different technical implementations for many of these innovations, some with broader implications for the system as a whole.¹ To name a few, I will show that the effects of Procrastinate (see Chomsky (1993)) can be derived without ascribing inherent cost to overt movement; that the Phrasal Uniformity Condition and the Word

¹ Part of this material was presented at the III Encontro de Estudos em Gramática Gerativa (Universidade Federal do Rio de Janeiro, August, 1995) and at the University of Southern California (Fall, 1995). I am thankful to these audiences.
Interpretation processes of Chomsky (1994) can be dispensed with; and that the relation between interpretability and accessibility to the computational system can be maintained without resort to the distinction between deletion and erasure (see Chomsky (1995:chap. 4)). The result is a considerably simpler theoretical apparatus.

I will pay special attention to the "copy theory of movement". I will propose that Move is not an operation of the computational system, but rather a description of the interaction of four operations: Copy, Merge, Form Chain, and Chain Reduction. This approach to movement operations, combined with Kayne's (1994) Linear Correspondence Axiom (LCA), will then constitute the basis for the proposals to be advanced in the chapters that follow, namely: (i) a multi-membered chain cannot be realized with all of its links overtly realized because it cannot be linearized in accordance with the LCA (chapter III); (ii) economy considerations regarding the elimination of formal features in the phonological component render deletion of traces for purposes of linearization more economical than deletion of heads of chains (chapter III); (iii) the derivation of parasitic gap and across-the-board extraction constructions involves "sideward movement", where a given constituent is copied and merged with an unconnected syntactic object (chapter IV); and (iv) deletion of chain links for purposes of linearization constrains sideward movement in such a way that it makes it possible to subsume the core properties of parasitic gap and across-the-board extraction.
II.2. The Minimalist Program: the General Picture

The Minimalist Program explores the hypothesis that the language faculty is a nonredundant and optimal system in the sense that particular phenomena are not overdetermined by linguistic principles and that the linguistic system is subject to economy restrictions of a specific type. The program also addresses the question of what conditions are imposed on the linguistic system in virtue of its interaction with performance systems ("the bare output conditions").

Earlier versions of the Principles and Parameters Theory worked with the hypothesis that the linguistic system has several levels of representation encoding systematic information about linguistic expressions. Some of these levels are conceptually necessary, since their output is the input to performance systems which interact with the linguistic system. The Minimalist Program restricts the class of possible linguistic levels of representation to only the ones which are required by conceptual necessity, namely, the ones which interface with performance systems.

Chomsky (1993, 1994, 1995: chap. 4) takes these performance systems to be the Articulatory-Perceptual System (A-P) and the Conceptual-Intentional
System (C-I).\textsuperscript{2} The linguistic levels which interface with A-P and C-I are PF and LF, respectively. Assuming that these are the only interface levels, PF and LF can be conceived of as the parts of the linguistic system which provide instructions to the performance systems. Under the Minimalist perspective, all principles and parameters of the linguistic system should be stated in either LF or PF terms, perhaps as modes of interpretation by the performance systems.\textsuperscript{3} Furthermore, all principles and parameters should make reference only to elements that function at the interface levels and to local relations among them. Linguistic expressions are then taken to be optimal realizations of interface conditions, where optimality is determined by economy conditions specified by Universal Grammar (UG).

Another assumption is that the language faculty is comprised of a

\textsuperscript{2} The term Articulatory-Perceptual System is to be understood as independent of the modality of the output system, in order to capture signed languages as well (see Chomsky (1994:fn. 4)). On the controversial assumptions that articulation and perception, on the one hand, and conceptualization and intention, on the other, each involve the same interface, and that the linguistic system only interacts with the performance systems A-P and C-I, see Chomsky (1994:fn.4, 1995:3).

\textsuperscript{3} Chomsky (1993:3) actually takes parametric variation to be restricted solely to PF, based on the poverty-of-stimulus argument that variation at LF is not readily accessible to the language learner. As Hornstein (1995:chap. 1) points out, however, this argument hinges on the questionable assumption that for the purposes of language acquisition, the relationship between LF and the C-I system is opaque, whereas the relationship between PF and the A-P system is not (see Hornstein (1994, 1995:chap. 1 and 5) for conceptual arguments as well as empirical evidence against this assumption). Although Chomsky (1995:chap. 4) continues to assume that “forms that reach the LF level must be as similar as typological variation permits – unique, if that is possible” (p.359), he proposes (p.291) that languages with and without overt wh-movement have different structures at LF and employ different interpretive devices at the C-I interface (see section II.14.3.3 for discussion). See Nunes (1995a) for an alternative analysis of wh-movement in terms of linearization at LF which takes languages with and without overt wh-movement to involve the same relevant LF structure and a single interpretive device at the C-I interface.
The lexicon specifies the items which enter into the computational system and their idiosyncratic properties, excluding whatever is predictable by principles of UG or properties of the language in question. The computational system arranges these items in a way to form a pair \((\pi, \lambda)\), where \(\pi\) is a PF object and \(\lambda\) is an LF object. If \(\lambda\) and \(\pi\) are legitimate objects (i.e. they satisfy Full Interpretation in the sense of Chomsky (1986b, 1993)), the derivation is said to converge at LF and at PF, respectively. If either \(\lambda\) or \(\pi\) does not satisfy Full Interpretation, the derivation is said to crash at the relevant level. A derivation is taken to converge only if it converges at both LF and PF.

If \(D\) is the set of permissible derivations which yield a pair \((\pi, \lambda)\), the set of convergent derivations \(C\) is thus the subset of \(D\) whose members satisfy Full Interpretation at LF and at PF. Finally, the set of admissible derivations \(A\) constitute the subset of \(C\) which is selected by economy considerations. In other words, the derivations which reach the performance systems are only the ones which converge in an optimal way.\(^5\)

In the sections that follow, I discuss specific aspects of the Minimalist

---

\(^4\) For a representational version of the Minimalist Program, see Brody (1995).

\(^5\) As Chomsky (1995:221) observes, if nonconvergent derivations could be taken into consideration for economy purposes, a derivation which employs no operation would always block any derivation which employs some operation, leading to incorrect empirical results. Thus, only convergent derivations can be compared in terms of economy.
II.3. The Numeration

The pair of legitimate objects \((\pi, \lambda)\) must meet the requirement of compatibility. After all, it is not the case that any linguistic sound can be associated with any linguistic meaning. \(\pi\) and \(\lambda\) should thus be based on the same lexical choices. In previous versions of the Principles and Parameters Theory, this compatibility requirement was ensured by D-Structure, which provided the computational system with an array of lexical items structured in a certain way. Under Minimalist assumptions, however, there is no room for a syntactic level such as D-Structure, because it is not an interface level.\(^6\)

In order for \((\pi, \lambda)\) to be formed according to Minimalist guidelines, we thus need an array of lexical items stripped of any substantive property that would make it a syntactic level of representation.

Chomsky (1994:7, 1995:225) proposes that such an array is a \textit{numeration}: a set of pairs \((LI, i)\), where \(LI\) is a lexical item comprised of (at most) phonological, semantic and formal features, and \(i\) indicates the number of times that \(LI\) is accessed by the operation \textit{Select}. Select pulls out a lexical item from a numeration, reduces its index by one, and makes this

\(^6\) For additional conceptual and empirical problems raised by the notion of D-Structure, see Chomsky (1993:sec. 3).
lexical item available for further operations of the computational system.

It is further assumed (see Chomsky (1994:8, 1995:228-229)) that the mapping from a numeration N to $\lambda$ is uniform (the operations available in the covert component are the same as the ones available in overt syntax) and inclusive ($\lambda$ is built from the features of the lexical items of N). The computations of the phonological component, by contrast, are different in nature and may involve addition of information which may not be specified by the lexical items of N (intonation structure, for instance). In this sense, the computations of the phonological component represent a departure from optimality.

II.4. Derivational Cost of the Operations of the Computational System

Among the operations of the computational system, which will be discussed in detail later on, some appear to be costless (Select and Merge) and others appear to induce computational cost to the derivation (Move, Delete, and Erase). Chomsky (1995:226) takes this distinction to hinge on whether an operation is a defining property of derivations or whether it is associated with a convergence condition on derivations.

For Chomsky (1995:225-226), a derivation is a sequence of symbolic elements S mapped from a numeration N such that the last member of S is
a pair \((\pi, \lambda)\) and \(N\) is reduced to zero (that is, for any LI of \(N\), \(i = 0\)). A derivation is said to be canceled if an illegitimate operation is performed during the computation, if the pair \((\pi, \lambda)\) is not formed, or if the numeration is not exhausted (see Chomsky (1995:225-226)). If the applications of Select are insufficient to exhaust the numeration, for instance, the derivation is canceled and no questions of convergence or economy arise. Similar considerations hold of the operation Merge, which takes two syntactic objects and replaces them with a single object (see section II.8 for details). If it is a defining property of a derivation that either \(\pi\) or \(\lambda\) are formed from a single syntactic object, the computational system must then employ sufficient applications of Merge.\(^7\) If such a requirement is not met, the derivation is canceled and no questions of convergence or economy can be raised.

The other operations of the computational system (Move, Delete, and Erase), on the other hand, are associated with convergence conditions. If they do not apply, a derivation may be formed, but at least one object of the

\(^7\) This corresponds to the property of single-rootedness of phrase markers in standard X'-Theory. Chomsky (1993:22) takes single-rootedness to be a convergence property at PF; Chomsky (1995:226), on the other hand, takes it to be a defining property of the mapping from \(N\) to \(\lambda\). The shift is related to the fact that Chomsky (1995:chap. 4) allows lexical access in the covert component, but not in the phonological component, as discussed in section II.5 below. It is reasonable to take single-rootedness to also be required at the point where Spell-Out applies, in order for (the relevant features of) the lexical items shipped to the phonological component to be linearized in accordance with Kayne’s (1994) Linear Correspondence Axiom (see section II.17 below). I will assume this to be the case (but see Hoffman (forthcoming) and Uriagereka (forthcoming:chap. 4) for the view that single-rootedness may not be required by the phonological component).
pair \((\pi, \lambda)\) violates Full Interpretation.

The operations which define what is a possible derivation therefore have no cost, whereas the operations which are required for the pair \((\pi, \lambda)\) to be legitimate and interpreted by the performance systems are derivationally costly. This notion of derivational cost will play an important role in the theory of movement to be developed in this dissertation.

II.5. Spell-Out

Elements interpretable at the A-P interface are not interpretable at the C-I interface, and vice versa. At some point in the derivation, the computational system must then split into two parts, one forming \(\pi\) and the other forming \(\lambda\), which do not interact any further after the bifurcation. S-Structure was the point of this split in pre-Minimalist versions of the Principles and Parameters Theory. The problem from a Minimalist perspective with there being a level feeding PF and LF such as S-Structure is that, since it does not interface with any performance system, it is not conceptually necessary. Thus, every substantive property attributed to S-Structure should be restated within the Minimalist framework in either LF or PF terms.\(^8\)

\(^8\) Chomsky (1993) offers an analysis of Binding Theory solely in LF terms (see section III.2), and outlines an approach reducing the parametric variation with respect
The only thing required under Minimalist assumptions is a rule which splits the computation to form the distinct objects $\pi$ and $\lambda$. Chomsky (1993:22) dubs this operation Spell-Out. Spell-Out is free to apply at any point in a given derivation; "wrong" choices presumably cause the derivation to crash at one of the interface levels.

The computation from Spell-Out to PF is referred to as the *phonological component*, the computation from Spell-Out to LF as the *covert component*, and the computation that obtains before Spell-Out as the *overt syntax*. In addition to containing phonological rules proper, the phonological component includes a morphological subcomponent and also deals with linearization, as will be discussed below (see section II.17).

The specifics of the Spell-Out rule have to do with the internal coherence of the system regarding access to the lexical items of N after Spell-Out applies. If a real option, lexical access after Spell-Out must be of a very restricted sort; otherwise the compatibility between $\pi$ and $\lambda$ collapses.

Chomsky (1993:22) takes Spell-Out as simply the shipment of the structure assembled by the computational system to the phonological component; however, he has to stipulate that there is no lexical access after Spell-Out.

to overt vs. covert movement to PF’s inability to deal with strong features, combined with economy considerations concerning overt movement (see section II.6). Also relevant in this regard is section IV.2.2 below, where I reinterpret the S-Structure condition on parasitic gap licensing in terms of linearization in the phonological component.
Chomsky (1994:8) eliminates this stipulation by interpreting Spell-Out as a rule that strips away from the structure already formed those elements that are only relevant to \( \pi \). Phonological features are then mapped only to PF, semantic features are mapped only to LF, and formal features feed both the covert and the phonological component.\(^9\) If a lexical item is selected by the phonological component, its semantic features will cause the derivation to crash at PF; on the other hand, if the covert component has access to the numeration, phonological features will cause the derivation to crash at LF.

Underlying this approach is the idea that both components can eliminate (certain kinds of) formal features, but the covert component cannot eliminate phonological features and the phonological component cannot eliminate semantic features. Left open is the possibility of access to lexical items without phonological or semantic features after Spell-Out.

Chomsky (1995:232) observes that once Select is available before Spell-Out, according to the uniformity condition on the mapping from N to \( \lambda \), it should also be available in the covert component. Also relying on the uniformity condition, Chomsky (1995:230) further proposes that the covert component cannot eliminate phonological features, under the assumption

\(^9\) Formal features are relevant for checking operations in the covert component, and presumably for morphological reasons in the phonological component. For instance, it is probably the case that Morphology checks the compatibility between phonological and Case-features (the phonological features of \( he \) are to be associated with nominative, not with accusative Case), and relies on categorical features to determine stress assignment (the stress pattern of \( record \) depends on whether it is a verb or a noun). In section III.6.2.3, I will propose that formal features are also relevant for deletion of traces in the phonological component.
that overt operations do not eliminate them.\textsuperscript{10} Lexical access in the covert component is thus permitted only for lexical items without phonological features. Semantic and formal features, by contrast, are allowed to be eliminated in the phonological component due to the different nature of its rules (see Chomsky (1995:230)).

Since the phonological component in this system is able to eliminate semantic features, questions arise regarding lexical access to items with semantic features in the phonological component. If the lexical items \textit{John} and \textit{denied} of the sentence in (1), for instance, were selected in the phonological component and their semantic features were deleted, (1) could wrongly mean 'he left early'.

\begin{equation}
\text{(1)} \quad \text{John denied he left early.}
\end{equation}

Chomsky (1995:232) assumes that Select is inoperative in the phonological component, circumventing the potential problem posed by this derivation of (1). Hence, lexical access in the phonological component is disallowed

\textsuperscript{10} Although the uniformity condition on the mapping from N to \(\lambda\) in principle makes the same set of operations available in the overt syntax and in the covert component, it need not be the case that the same operations are actually realized before and after Spell-Out; economy conditions may block the application of certain operations before Spell-Out. Suppose, for the sake of argument, that in the mapping from N to \(\lambda\), phonological features can be eliminated by operation OP. Notice that OP would be redundant if it were to apply overtly, given that Spell-Out removes phonological features from the computation from N to \(\lambda\). In the covert component, on the other hand, OP would not be redundant and would apply as required. In this case, OP would be uniformly available during the mapping from N to \(\lambda\), but would apply only in the covert component. I will return to this issue in section III.6.2.3.
regardless of whether or not the lexical items have semantic features.

II.5.1. The Version of Spell-Out to Be Adopted in This Dissertation

As can be seen from the above presentation, there is a balance between the specific details of Spell-Out and the computation to be done in the phonological component: the more elaborate the rule of Spell-Out, the lighter the work load of the phonological component. In the system proposed in Chomsky (1995:chap. 4), Spell-Out only takes phonological features away from the computation from N to λ; the phonological component is then required to eliminate the semantic features it receives from Spell-Out (see Chomsky (1995:230)). By contrast, in the system proposed in Chomsky (1994), Spell-Out is much more elaborate in that it ships lexical features to the covert and/or the phonological component according to their relevance for the computations of each component; hence no elimination of semantic features in the phonological component is required because these features are only shipped to LF.

The economy criterion discussed in section II.4 can help us to choose between these two alternatives. As discussed in section II.4, operations which are required for a derivation to be generated are derivationally costless, whereas operations which are required for convergence purposes are derivationally costly. In the case at hand, we need to choose between
introducing a more "complex" definition of Spell-Out and allowing semantic features to be eliminated in the phonological component.

The Spell-Out rule is a defining property of a derivation. If it does not apply, the pair \((\pi, \lambda)\) cannot be obtained and no derivation is generated; hence the Spell-Out rule, even if it is "complex", is in principle costless. Elimination of semantic features in the phonological component, on the other hand, is required in order for the PF object \(\pi\) to satisfy Full Interpretation. Thus, this operation, however defined, entails derivational cost to the system. I will therefore assume the version of Spell-Out proposed in Chomsky (1994:8), according to which semantic features are shipped only to the covert component, phonological features are shipped only to the phonological component, and formal features are shipped to both components (see fn. 9).

As for lexical access after Spell-Out, I will follow Chomsky (1995:232) in taking this to be possible in the covert component, but not in the phonological component. Nevertheless, I will depart from him in not ascribing the lack of lexical access in the phonological component to the operation Select itself. If we understand Select as merely the operation which feeds the computational system with lexical items drawn from a numeration, it should in principle be available to both components. The operation Merge, however, appears to be restricted to the mapping from N
to $\lambda$.

If this is so and if the operation which linearizes the relevant features shipped to the phonological component can only apply to a single syntactic object (see fn. 7 and section II.17), lexical access in the phonological component will always cancel the derivation. If a lexical item LI is selected in the phonological component, it will not concatenate with the already assembled structure $\Sigma$, because Merge (by hypothesis) only operates in the mapping from $N$ to $\lambda$. The linearize operation will then be presented with two disconnected syntactic objects, LI and $\Sigma$, and will yield no output, canceling the derivation. Thus, there is no way for a sentence such as (1) to be derived with the meaning 'he left early'.

II.6. Feature Checking, Procrastinate, and Strong Features

Chomsky (1993:28-30) notes that a language in which inflectional heads lower and adjoin to lexical heads would look quite different at LF from languages whose lexical heads raise and adjoin to inflection heads. In order to avoid this disparity (see fn. 3), Chomsky assumes that lexical and functional heads are already inflected at the point of their insertion in the derivation.\footnote{In more current terms, to say that a lexical item LI is already inflected when it enters into a derivation amounts to saying that the formal features of LI are fully} A checking operation made available by Merge or Move then
allows lexical and functional heads to be appropriately paired (if they can be). Thus, regardless of whether an element raises overtly or covertly, the relevant features would be checked in the same way.

The problem is then to show how the parametric variation concerning overt vs. covert movement could be stated without reference to a syntactic level such as S-Structure. This problem basically involves two questions: (i) Why should all languages not have only overt movement?; and (ii) Why do some languages have overt movement?

To address the first question, Chomsky (1993:30) proposes an economy principle called Procrastinate, which states that covert movement is less costly than overt movement. It should be noted that as formulated, Procrastinate constitutes a departure from the uniformity requirement on the mapping from \( N \) to \( \lambda \) in that it ascribes an inherent difference between overt and covert operations. For current purposes, I will nonetheless assume Chomsky's (1993) original characterization of Procrastinate, leaving for section II.10.2 a discussion of how the effects of Procrastinate can be derived in accordance with the uniformity condition on the mapping from \( N \) to \( \lambda \).

specified by them. Chomsky (1995:231) classifies phonological, semantic, and formal features as intrinsic or optional. Intrinsic features encode idiosyncratic properties of a lexical item and are present in the lexicon; by contrast, optional features encode the properties that are predictable by UG or by the language in question, and are assigned to LI as it becomes part of a given numeration. A given feature may be intrinsic or optional, depending on the lexical item or on the language in question (see Lasnik (1994) and Uriagereka (1994, forthcoming:chap. 4) for related issues).
Regarding the second question, Chomsky (1993:30) proposes that the features of a lexical item may be weak or strong, and that strong features cannot be eliminated in the phonological component. Recall that strong features could not be introduced in the covert component, because no lexical access after Spell-Out was permitted in Chomsky's (1993) system. Thus, the only way to prevent strong features from reaching the phonological component was to eliminate them before Spell-Out through the checking operation made available by either Merge or Move. If Merge does not yield a convergent derivation, overt movement is then required.\textsuperscript{12} Under this view, the Extended Projection Principle (EPP) of Chomsky (1981), which requires that every clause have a subject, is reinterpreted as the tense head T having a strong nominal feature.\textsuperscript{13}

So far, approaches to strong features within the Minimalist Program have not gone much beyond mere restatements of the phenomena to be accounted for. In the system developed in Chomsky (1995:232-235), for instance, it is assumed that strong features induce cyclicity and that this

\textsuperscript{12} The fact that overt movement triggered by strong feature checking always violates Procrastinate is not a problem. Since Procrastinate is an economy principle, it chooses among competing derivations which converge. In the case at hand, if overt movement does not take place, a strong feature will reach the phonological component and the derivation will crash at PF.

\textsuperscript{13} Further refinements may require the postulation of three variants of EPP; in terms of a strong D-feature, a strong N-feature, or either (see Chomsky (1995:233)). For purposes of presentation, I will however ignore this potential distinction and assume that T has a strong D-feature which can be checked by either an NP or a DP. Accordingly, when I say, for instance, that the D-feature of a name like \textit{John} checks the EPP, that is to be taken as an abbreviation for something like 'the N-feature of \textit{John} or the D-feature of the null determiner which merges with \textit{John} checks the EPP' (see Uriagereka (1993) and Longobardi (1994) for relevant issues).
property follows from (2):

(2)   a. Nothing can join to a non-projecting category.

       b. A derivation is canceled if it has formed a structure \( \Sigma \) containing an element \( \alpha \) with a strong feature and \( \Sigma \) is not headed by \( \alpha \).

The idea is that "a strong feature merged at the root must be eliminated before it becomes part of a larger structure by further operations" (Chomsky (1995:234)). Thus, if \( F \) is a strong feature of a functional category \( X \), \( F \) must be checked while the only categories dominating \( X \) are projections of \( X \). Suppose, for instance, that \( T \) has a strong D-feature; \( T \) can have this feature checked by having a DP in its Spec before TP merges with the complementizer C and C projects. By contrast, if TP merges with C without having a DP in its specifier, (2a) prevents a DP from moving to the specifier of TP after C projects. Since the strong D-feature of \( T \) is not checked within the TP projection, the derivation is canceled, in accordance with (2b).

As for the possibility of inserting a lexical item with a strong feature but no phonological features in the covert component (see section II.14.3.3 for further discussion), Chomsky (1995:294) rules it out based on the economy principle in (3), where effect is taken to mean literal identity with respect to
the PF output, and logical equivalence with respect to the LF output.

(3) \( \alpha \) enters the numeration only if it has an effect on output.

According to Chomsky (1995:294),

Output conditions enter into determination of the numeration itself; they affect the operation that constructs the numeration from the lexicon. (...) Insofar as its presence [the presence of a strong feature; JMN] is motivated only by PF manifestation, it cannot be inserted covertly, under (3) [my numbering, JMN], or it would not have been in the numeration at all. (Chomsky (1995:294))


It seems to me that if compared to the proposal laid out in Chomsky (1993), for instance, the analysis of strong features in terms of (2) and (3) represents a considerable step back. First, this approach takes strong features to induce cyclicity. However, cyclicity is arguably a general property of the system which is independent of strong feature checking. Chomsky (1995:294), for instance, assumes that an operation which merges two root structures \( \alpha \) and \( \beta \) is conceptually much simpler than an operation which
merges $\alpha$ with a constituent of $\beta$. Thus, regardless of whether or not strong features are involved, cyclic operations are in principle to be preferred over noncyclic ones. This is even more so, if Move is to be taken as a description of the interaction of Merge with other independent operations, as I will argue below.

The fact that some movement operations take place overtly, in violation of Procrastinate, arguably follows from some property of the phonological component; otherwise, optimality considerations would only allow covert movement. The analysis offered in Chomsky (1995:294) masks this phonological nature of strong features by relying on the global economy principle in (3), even though it assumes that the presence of strong features is motivated "only by PF manifestation" (see citation above).

The principle in (3) itself does not fit well within the overall system outlined in Chomsky (1995:227-228), which strives for reducing computational complexity. According to (3), in order for the system to determine whether the presence of a feature $F$ in a numeration $N$ is legitimate, it may be necessary to consider every possible derivation starting with $N$ and verify if $F$ causes some effect on the output of at least one of these derivations. This, nonetheless, is as global a condition on possible numerations as is possible. Furthermore, as pointed out to me by Norbert Hornstein (p.c.), the introduction of (3) comes close to restipulating the
properties of D-Structure in the numeration.\textsuperscript{14}

The principle in (3) also appears to make incorrect predictions for languages without morphological case distinction. Since Case-features are uninterpretable regardless of the categorial features they are associated with (see sections II.14, for further discussion), they do not have an effect on the LF output, under the assumption that the presence or absence of Case-features does not change the logical structure with which the C-I system works. Thus, according to (3), Case-features must have an effect on the PF output. That is plausible in languages with a rich morphological case system, for it is reasonable that in these instances, the phonological realization of a nominal element is linked to its Case-feature (see fn. 9). In languages with no morphological case, on the other hand, no effect on the PF output is observed. Hence, Case-features should not be part of any numeration in these languages in compliance with the principle in (3). This consequence of (3) is undesirable, given the body of empirical evidence stemming from Vergnaud’s original proposal that abstract Case should be dissociated from morphological case (see section II.15 for an additional problem for (3) regarding computation of economy).

Rather than an economy principle governing the licensing of a strong

\textsuperscript{14} I should be noted, however, that (3) may derive the requirement that a derivation must exhaust the numeration (see section II.4). If (3) holds, then every lexical item of a numeration will have to be pulled out from the numeration in order to have an effect on the output. A more notational observation is that (3) also blocks the presence of a lexical item LI with index \( i = 0 \) in an initial numeration.
feature in a given numeration, something along the lines of (3) seems to be a reasonable learning strategy that a child may employ in order to set parameters concerning the presence of strong features in a given language. Below I suggest an alternative view of strong features.

II.6.2. Strong Features as Hybrid Features

A more appealing proposal concerning strong features is made in Chomsky (1994:9), which suggests that a checked strong feature is stripped away by Spell-Out. Under this view, the phonological component is able to eliminate checked strong features, but the covert component is not. If a strong feature is not taken away from the computation from N to \( \lambda \) by Spell-Out, the derivation crashes at LF.

Notice that in order for this approach to be consistent with the uniformity condition on the mapping from N to l, strong features must have some phonological property. Otherwise, a strong feature should be checked covertly in compliance with Procrastinate. If strong features have some phonological property, on the other hand, strong feature checking in the covert component is not enough to ensure LF convergence. By assumption (see section II.5), the covert component is not able to eliminate phonological features. Thus, even after being checked in the covert component, a strong feature still violates Full Interpretation at LF because its
phonological part is uninterpretable at this level. The impossibility of inserting lexical items with a strong feature but no phonological features in the covert component is then derived locally, without a global condition such as (3).

The idea that strong features have some phonological property is also consistent with Chomsky's (1994) version of Spell-Out adopted here. If Spell-Out is sensitive to the relevance of lexical features for the covert and the phonological components, as assumed in section II.5.1, its stripping strong features from the mapping from N to $\nu$ entails that strong features have some phonological property uninterpretable at LF.

Left open in this elaboration of Chomsky's (1994) suggestion is the question of why strong features trigger movement operations to begin with. In other words, why can they not simply be shipped unchecked to the phonological component by Spell-Out? As a speculative answer, I would like to suggest that a strong feature is a hybrid formal-phonological feature which is a member of the set of phonological features of a given lexical item. If this is so, it is possible that the phonological component cannot eliminate the formal part of a strong feature. Thus, if a strong feature is not checked in the overt syntax, it causes the derivation to crash at PF because its formal part violates Full Interpretation at PF.

This approach admittedly rests on the stipulation that checking operations can "see" only the formal part of a strong feature, whereas the
operations of the phonological component which eliminate unchecked formal features (see section III.6.2.1) cannot. This stipulation may perhaps be eliminated if we combine some natural assumptions about the rules of the phonological component with the Move-F theory to be discussed in section II.12.

In the version of Spell-Out I have adopted (see section II.5.1), lexical items are split into sets of features, which are then shipped to the relevant components. It is thus plausible to assume that the phonological component employs a set of rules which specifically deal with the set of formal features, and another distinct set of rules which specifically deal with the set of phonological features. If strong features are members of the set of phonological features, as suggested above, their formal part cannot be deleted by the rule which applies to the set of formal features and deletes unchecked features (see section III.6.2.3). Under the assumption that the rules which are designed to apply to the set of phonological features are not able to eliminate the formal part of strong features, a derivation containing a strong feature will fail to meet Full Interpretation at PF and will crash unless an overt checking operation renders the formal part of a strong feature invisible at PF.

Checking operations, in contrast, are independently taken to be blind to anything other than formal features (see section II.12). The general idea under a Move-F approach is that morphological reasons require that overt
movement carry semantic and phonological features along with the relevant formal features which are subject to a checking operation, but the checking operation itself is blind with respect to semantic and phonological features (see section II.12 for detailed discussion). Thus, a checking operation is able to ignore the phonological part of a strong feature and operate solely with its formal part.

If the asymmetry between operations of the phonological component and checking operations can be independently motivated along the lines suggested above, the fact that strong features require overt checking will follow.\textsuperscript{15} I return to the issue of strong features in section III.6.3.

II.7. Domains, Clausal Structure, and Locality

II.7.1. Domains


In the system outlined in Chomsky (1993), grammatical relations are specified in terms of the notions defined below (see Chomsky 1993:11-14), where \( \alpha \) is either a trivial head H or the head of a nontrivial head chain

\textsuperscript{15} The proposal that a strong feature is a hybrid formal-phonological feature is consistent with Chomsky's (1995:232) claim that strong features should be checked only by categorial features. However, the reason why a strong feature cannot be checked by Case or \( \phi \)-features remains to be explained.
CH = (Hi, ti):\textsuperscript{16}

(4) \textit{Max}(\alpha):

The least full-category maximal projection dominating \alpha.

(5) \textit{Domain of } \alpha (\delta(\alpha)):

The set of categories contained in \textit{Max}(\alpha) that are distinct from and do not contain \alpha.

(6) \textit{Complement Domain of } \alpha (\textit{Compl}(\delta(\alpha))):

The subset of \delta(\alpha) reflexively dominated by the complement of \alpha.

\textsuperscript{16} Unless specified otherwise, (categorial) domination and containment, which are defined in (i) and (ii) (see May (1985) and Chomsky (1986a)), are to be taken as irreflexive.

(i) \textit{Categorial Domination} \\
\alpha \text{ dominates } \beta \text{ iff every segment of } \alpha \text{ dominates } \beta.

(ii) \textit{Containment} \\
\alpha \text{ contains } \beta \text{ iff some segment of } \alpha \text{ dominates } \beta.

Although extensionally equivalent to the corresponding definitions in terms of reflexive domination, the revisions of the notions of minimal, internal and checking domains that I suggest below (cf. (12), (14) and (15)) will not make reference to reflexive domination for reasons having to do with how domination should be naturally stated in bare X'-Theory terms (see section II.9.2 and fn. 36 below).
(7) *Residue of* $\alpha$ (*Res*(\(\alpha\))): 

The subset $K$ of $\delta(\alpha)$ such that $K$ is the set-theoretic complement of $\text{Compl}(\delta(\alpha))$.

(8) *Minimal Domain of* $\alpha$ (*Min*(\(\delta(\alpha)\))): 

The smallest subset $L$ of $\delta(\alpha)$ such that for any $\gamma \in \delta(\alpha)$, some $\beta \in L$ reflexively dominates $\gamma$.

(9) *Internal Domain of* $\alpha$ (*Int*(\(\delta(\alpha)\))): 

The smallest subset $M$ of $\text{Compl}(\delta(\alpha))$ such that for any $\gamma \in \text{Compl}(\delta(\alpha))$, some $\beta \in M$ reflexively dominates $\gamma$.

(10) *Checking Domain of* $\alpha$ (*Check*(\(\delta(\alpha)\))): 

The smallest subset $N$ of $\text{Res}(\alpha)$ such that for any $\gamma \in \text{Res}(\alpha)$, some $\beta \in N$ reflexively dominates $\gamma$.

Let us see how these definitions work by considering the phrase-marker in (11), where the head $H$ adjoins to the head $X$: 
Let us first consider the two-segment category $X ([X_2, X_1])$. $\text{Max}(X)$ is the two-segment category $\text{XP} ([\text{XP}_2, \text{XP}_1])$; $\delta(X) = \{\text{GP}, [\text{UP}_2, \text{UP}_1], \text{WP}, \text{KP}, H_i, \text{HP}\}$ and whatever these categories dominate; $\text{Compl}(\delta(X))$ is $\text{HP}$ and whatever it dominates;\(^{17}\) $\text{Res}(X) = \{\text{GP}, [\text{UP}_2, \text{UP}_1], \text{WP}, \text{KP}, H_i\}$ and whatever these categories dominate; $\text{Min}(\delta(X)) = \{\text{GP}, [\text{UP}_2, \text{UP}_1], \text{WP}, \text{KP}, H_i, \text{HP}\}$; $\text{Int}(\delta(X))$ is $\text{HP}$ (see fn. 17); and $\text{Check}(\delta(X)) = \{\text{GP}, [\text{UP}_2, \text{UP}_1], \text{WP}, \text{KP}, H_i\}$.

Regarding the head $H$, we have to consider two different derivationally defined domains: when $H$ enters into the derivation, and after it raises and

---

\(^{17}\) The term *complement* in the definition of complement domain in (6) presumably means 'sister of a (nonmaximal) head'. If so, the complement of $X_1$ in (11) before adjunction of $H$ is $\text{HP}$ and whatever $\text{HP}$ dominates. Matters are not so clear after $H$ adjoins to $X_i$, forming the two-segment category $[X_2, X_1]$. Is $\text{HP}$ still the complement of $X_1$, does it become the complement of $[X_2, X_1]$, or does it become the complement of both? If the definition of $c$-command in (13) is to be assumed for the reasons discussed below, $\text{HP}$ in (11) will be the complement of both $X_1$ and $[X_2, X_1]$.
forms the chain $\text{CH} = (H_i, t_i)$. Before it moves, $\text{Max}(H)$ is HP; $\delta(H) = \{ZP, YP\}$ and whatever these categories dominate; $\text{Compl}(\delta(H))$ is YP and whatever it dominates; $\text{Res}(X)$ is \{ZP\} and whatever it dominates; $\text{Min}(\delta(H)) = \{ZP, YP\}$; $\text{Int}(\delta(H))$ is YP; and $\text{Check}(\delta(H))$ is ZP. After H raises, the newly established notions for the chain $\text{CH} = (H_i, t_i)$ are as follows: $\text{Max}(\text{CH})$ is the two-segment category XP ([XP$_2$, XP$_1$]); $\delta(\text{CH}) = \{GP, [UP$_2$, UP$_1$], WP, KP, ZP, YP\}$ and whatever these categories dominate; $\text{Compl}(\delta(\text{CH}))$ is ZP and YP and whatever they dominate; $\text{Res}(\text{CH}) = \{GP, [UP$_2$, UP$_1$], WP, KP\}$ and whatever these categories dominate; $\text{Min}(\delta(\text{CH})) = \{GP, [UP$_2$, UP$_1$], WP, KP, ZP, YP\}$; $\text{Int}(\delta(\text{CH}))$ is \{ZP, YP\}; and $\text{Check}(\delta(\text{CH})) = \{GP, [UP$_2$, UP$_1$], WP, KP\}$.

II.7.1.2. Some Revisions

The checking domain of a given head $X$ in Chomsky’s (1993) system, as defined in (10), is a heterogenous set. It contains (i) the specifier of $X$, which enters either into a $\theta$-relation or a checking relation with $X$ (see discussion in section II.7.4); (ii) heads adjoined to $X$, which enter into checking relations with $X$; (iii) maximal projections adjoined to the specifier of $X$, which perhaps enter into a checking relation with $X$;\textsuperscript{18} and (iv) maximal

\textsuperscript{18} For instance, if languages such as Bulgarian involve adjunction of a wh-phrase to another wh-phrase in Spec of CP, as proposed by Rudin (1988), we could say that
projections adjoined to $X'$ or $XP$, which appear to enter neither into a $\theta$-relation nor into a checking relation with $X$. Chomsky (1995:319) restricts the notion of checking domain by proposing that an element adjoined to a nonminimal category $X$ (that is, a maximal or an intermediate projection; see section II.8 below) is not in the checking domain of the head of $X$. In order to obtain this result, I will assume the definition of minimal domain in (12), incorporating Chomsky's (1995:319) restriction in (12i):\footnote{Of course, Chomsky's (1995:319) restriction could also be imposed on the definitions of residue or checking domain. I will, instead, reformulate the notion of minimal domain with an eye to the revision of internal and checking domains to be suggested in (14) and (15) below.}

the adjoined wh-phrases allow a strong feature in $C$ to be multiply checked (see section II.7.3 below for the hypothesis that in some languages a strong feature may remain visible after being checked). If these languages actually involve multiple Specs of CP, as proposed by Koizumi (1994), we should restrict the checking domain of a head $H$ to the specifier(s) of $H$ and the elements adjoined to $H$. In this dissertation, I will be assuming this more restrictive notion of checking domain (see section II.13.1 for further discussion). Accordingly, the relation between $\text{Max} (\beta)$ and $\gamma$ in the definition of minimal domain in (12) will be stated in terms of containment. However, if it turns out that Bulgarian-type languages must be analyzed as involving multiple adjunction to a wh-phrase in the Spec of CP, the definition of minimal domain in (12) should then be formulated only in terms of domination, as shown in (i):

\begin{enumerate}
\item \textit{Minimal Domain of $\alpha$ ($\text{Min} (\delta (\alpha))$):}
The subset $K$ of $\delta (\alpha)$ such that for any $\gamma \in \delta (\alpha)$,
\begin{enumerate}
\item $\gamma$ is not adjoined to nonminimal projection of $\alpha$; and
\item if some $\text{Max} (\beta) \neq \text{Max} (\alpha)$ dominates $\gamma$, then $\text{Max} (\beta)$ dominates $\text{Max} (\alpha)$.
\end{enumerate}
\end{enumerate}
(12)  Minimal Domain of $\alpha$ (Min($\delta(\alpha)$)):

The subset $K$ of $\delta(\alpha)$ such that for any $\gamma \in K$,

(i) $\gamma$ is not adjoined to a nonminimal projection of $\alpha$; and

(ii) if some $\text{Max}(\beta) \neq \text{Max}(\alpha)$ contains $\gamma$, then $\text{Max}(\beta)$ dominates $\text{Max}(\alpha)$.

Let us now reconsider the notion of complement domain given in (6). As mentioned in fn. 17, when applied to heads, the term *complement* in (6) appears to be synonymous to 'sister of a (nonmaximal) head'. This becomes less clear, however, when the complement domain of head chains such as $\text{CH} = (H_i, t_i)$ in (11) is computed. Chomsky (1993) takes ZP in a structure such as (11) to be a member of the complement domain of CH, although ZP is a sister of neither $H_i$ nor $t_i$. If complement domain is to be so construed, the notion of complement should be stated in terms of c-command, rather than sisterhood. Observe that the categories which belong to the complement domain of $\text{CH} = (H_i, t_i)$ are c-commanded by $H_i$.

The question then is what the relevant notion of c-command is, if this approach is to be pursued. Crucially, $H_i$ in (11) should be prevented from being c-commanded by the two-segment category $[X_2, X_1]$; otherwise, it would be in the complement domain, rather than in the checking domain of $[X_2, X_1]$, as desired. The expected results can be obtained if the c-command
definition in terms of exclusion given in (13) below is adopted (see Kayne (1994:18) and Chomsky (1994:30)), where maximal projections and heads are accessible to the computational system, but not intermediate projections. According to the definition in (13), the two-segment category \([X_2, X_1]\) does not c-command \(H_i\) in (11) because one of its segments dominates \(H\).

(13) **C-command:**

Where \(\alpha\) and \(\beta\) are accessible to the computational system, \(\alpha\) c-commands \(\beta\) iff:

(i) \(\alpha \neq \beta\);

(ii) no segment of \(\alpha\) dominates \(\beta\); and

(iii) every category dominating \(\alpha\) dominates \(\beta\).

If the definition of c-command in (13) is assumed, the notion of internal and checking domains can be reformulated as in (14) and (15), where \(\text{Min}(\delta(\alpha))\) is to be understood as in (12):

(14) **Internal Domain of \(\alpha\) (Int(\(\delta(\alpha)\))):**

The subset \(L\) of \(\text{Min}(\delta(\alpha))\) such that for any \(\beta \in L\), \(\beta\) c-commands no member of \(\text{Min}(\delta(\alpha))\).
(15) **Checking Domain of α (Check(δ(α))):**

The subset M of Min(δ(α)) such that M is the set-theoretic complement of Int(δ(α)).

According to the revisions in (12), (14), and (15), the minimal domain of the two-segment category [X₂, X₁] in (11) is \{[UP₂, UP₁], H, HP\} and its checking domain is \{[UP₂, UP₁], H\}; the minimal domain of CH = (Hi, ti) is \{[UP₂, UP₁], ZP, YP\} and its checking domain is \{[UP₂, UP₁]\}. The minimal and checking domains of H before raising, as well as the internal domains of [X₂, X₁], H, and CH = (Hi, ti) remain the same.

As will be discussed in section II.7.4 below, Chomsky (1995:356) proposes that the computation of locality for movement operations may take into consideration only the domains specified by trivial head chains. Uriagereka (forthcoming:chap. 6), on the other hand, argues that domains determined by nontrivial head chains are relevant for issues related to idioms and word formation. Since these issues will not be directly relevant for what follows, I will assume the definitions in (12), (13), and (15) with α being restricted to trivial heads chains.
II.7.2. The Structure of VP

Chomsky (1993:7, 19) takes the structure of IP to be as represented in (16) below, which incorporates the VP-Internal Subject Hypothesis (see Zagona (1982), Fukui and Speas (1986), Kitagawa (1986), Kuroda (1988), and Koopman and Sportiche (1991), among others) and Chomsky's (1991) version of the Split-Infl Hypothesis (see Pollock (1989) and Belletti (1990)).

(16)

```
  AgrsP
    Agrs'  TP
       Agrs  AgroP
              T
                   Agro'  VP
                          SU
                               V
                                    OB
```

VP-Internal Subject Hypothesis sketched in (17):

\[
\begin{array}{c}
\text{vP} \\
\text{DP}_1 \quad \text{v'} \\
\quad \text{v} \quad \text{VP} \\
\quad (\text{DP}_2) \quad V' \\
\quad \quad \text{V} \quad \text{DP}_3
\end{array}
\]

In (17), v is a light verb which takes VP as complement. The DPs dominated by VP are internal arguments and the DP in the Spec of vP is the external argument. The causative or agentive role of the external argument is taken to be determined by the v-VP configuration (see Chomsky (1995:315)). The structure in (17) is extended to transitive verbs with a single argument and unergative verbs (see fn. 35); unaccusative verbs, on the other hand, only make use of the lower VP shell in (17).

II.7.3. Agr-Projections

The postulation of an Agro projection is motivated only theory-internally, because Agro receives no independent interpretation at the
interface (see Iatridou (1990) for relevant discussion). In the system proposed in Chomsky (1993), for instance, the object raises to Spec of AgroP at some point in the derivation to check its Case and $\phi$-features against the category in (18), which is formed after the verb adjoins to Agro.

(18) $[\text{Agro} \ V \ [\text{Agro}, \text{Agro}] ]$

Chomsky (1995:355) suggests that only functional projections which receive an interpretation at the interface should be postulated. The challenge then is to account for the structural and checking properties ascribed to Agro in Chomsky’s (1993) system, which are summarized in (19), without relying on a projection of Agro:

(19) a. Agro mediates the checking relation between the Case- and the $\phi$-features of the verb, and the Case- and the $\phi$-features of the object.  
   b. Agro provides a structural position for an object to check its features at LF.  
   c. Agro provides a structural position for an object to check its features overtly in languages with overt object movement.

Chomsky (1995:351) dispenses with (19a) and (19b) by allowing the DP
object to enter into a checking relation directly with the light verb or the complex verbal head formed by adjoining the main to the light verb, as shown in (20). 20

(20) \([v\ V\ [v\ v\ ]]\]

According to the definitions of checking domain discussed in sections II.7.1.1 and II7.1.2, if the object adjoins to the light verb or to the complex verbal head at LF, it will be in the checking domain where the relevant features can be checked (see sections II.12, II.13, and III.4.1.5.3 for further discussion). If the object adjoins to (20) at LF, for instance, it will be in the checking domain of both the light verb and the verb chain (Vi, ti). Hence, no position associated with Agro needs to be postulated for this checking operation to obtain at LF (see sections II.7.4 and II.14, and fn. 22 below for some refinement).

The only remaining property ascribed to Agro to be accounted for is thus the extra specifier position that it appears to make available in languages with overt object movement. In order to also dispense with (19c), Chomsky (1995:352) assumes that a given head may license more than one specifier depending on its feature composition (see Ura (1994)). He proposes

---

20 The light verb presumably has a strong V-feature (see Chomsky (1995:321)), triggering overt verb movement within the VP-shell and yielding the category in (20).
(p. 352) that in languages with overt object movement, the light verb \( v \) in (17) has a strong feature, triggering movement of the object to its "outer" specifier.\(^{21}\) If this line of reasoning is correct, Agro can be dispensed with entirely.

Chomsky (1995:351, 354) uses similar considerations to dispense with Agrs, as well. In the system outlined in Chomsky (1993), the subject raises to Spec of AgrsP and checks Case and \( \phi \)-features against the category in (21), formed by the adjunction of T (or \([T [\text{Agro} V [\text{Agro}]] [T T]]\)) to Agrs.

\[
(21) \quad [\text{Agrs} T [\text{Agrs Agrs}]]
\]

By allowing the subject to enter into a checking relation directly with the category in (22) below, which is formed by raising the complex verbal head in (20) to T, Chomsky (1995:351) eliminates the need for an Agrs projection in most cases. In English, for instance, the subject raises overtly to Spec of TP to check the strong D-feature of T. This movement also allows the subject and T to check their Case-features. At LF, the complex verbal head in (20) adjoins to T, forming (22); the subject is then in the checking domain of the chain \((v_i, t_i)\), allowing the \( \phi \)-features of this chain to be

\(^{21}\) If the subject moves to check this feature, the derivation will not converge, as will be shown in section II.7.4.
checked against the $\phi$-features of the subject.$^{22}$

\[(22) \quad [T \, [v \, V \, [v \, v \, ]] \, [T \, T \, ]]\]

Like Agro, Agrs seems to make an extra Spec position available in instances of transitive expletive constructions (see Ottósson (1989) and Jonas and Bobaljik (1993), among others), such as the one illustrated in (23) below (from Jonas and Bobaljik (1993:76)), where the subject appears to be in Spec of TP and the expletive in Spec of AgrsP (see Jonas and Bobaljik (1993) and Chomsky (1994)):

\[(23) \quad Tað lásu einhverjir stúdentar bókina.\]

there read some students the-book

'Some students read the book.'

---

$^{22}$ As mentioned in section II.7.2, unaccusative verbs presumably do not make use of the light verb shell. Given that unaccusative verbs have $\phi$-features but not an accusative Case-feature, it is reasonable to assume that in general, the main verb carries $\phi$-features, whereas the light verb carries a Case-feature. Since the light verb is also involved in external $\theta$-role assignment, this assumption is compatible with Burzio's (1986) Generalization, according to which a verb assigns accusative Case only if it assigns an external $\theta$-role.

As will be discussed in section II.4, Chomsky (1995:280) defines the checking operation in such a way that it allows a specifier of a head H to check the features of H and of any other head adjoined to H. Under this approach, the subject is thus able to check the $\phi$-features of the main verb even after the verbal complex in (20) adjoins to T, yielding (22).
Chomsky (1995:354) proposes that strong features may be parametrized in terms of the number of times that they can be accessible to the computational system after being checked (on the distinction between deletion and erasure, see section II.14 below):²³

MSCs [multiple specifier constructions; JMN] appear only when the EPP holds. The question of their nature arises, the, only when T already has a strong [nominal-] feature, which is deleted when checked by DP or NP in [Spec, T]. Suppose that the derivation has reached the stage of TP with T strong, and the numeration contains an unused expletive Exp. Then Exp can be inserted by Merge to satisfy the EPP, and we have an ordinary expletive-associate construction. The strong features of T deletes and furthermore erases, since the derivation converges. Hence overt MSCs exist only if T has a parametrized property (...) which allows a -Interpretatable feature (in this case, the strong [nominal-] feature) to escape erasure when checked. If the option is selected, then there must be a multiple-Spec construction, with \( n+1 \) Specs if the option is exercised \( n \) times. In a language with EPP but no MSCs, the strong feature of T is introduced into the derivation with \( n = 0 \), hence erased when checked. In Icelandic, the descriptive facts indicate that \( n = 0 \) or \( n = 1 \); in the latter case, T has two Specs. (Chomsky (1995:354-355))

In contrast to English, where the strong D-feature in T is inaccessible

---

²³ See also Chomsky (1995:374) for a formulation of this parameter in terms of "forced" vs. "unforced" violations of Procrastinate.
after being checked, the strong feature of T in Icelandic may be accessible once checked. In the derivation of (23) under this approach, the subject checks the strong D-feature of T, which remains accessible to the computational system; the expletive then merges with TP, and checks the strong feature of T again, finally rendering it inert for the computational system and for interpretation. This analysis predicts, however, that the order obtained after this double checking should be as in (24), rather than the order in (25), which is the one observed in (23) under the assumption that the verb is adjoined to T. Chomsky (1995:368) suggests that the order in (25) is the result of phonological rules having to do with the V2 nature of Icelandic, but that in the mapping from N to λ, the configuration is indeed the one in (24).

(24)   Exp-Subject-T

(25)   Exp-T-Subject

To summarize, the basic IP skeleton proposed in Chomsky (1995:chap.

---

24 As pointed out to me by Norbert Hornstein (p.c.), the correlation between transitive expletive constructions and V2 properties in Icelandic receives support from the fact that the acceptability of these constructions degrades when they occur in embedded clauses. If this line of reasoning is to be pursued, however, it seems less stipulative to take Icelandic matrix C to have a strong feature, rather than taking the strong feature of T to be optionally inaccessible after being checked.
4), which will be assumed here, is as in (26), with extra Specs for TP and vP being dependent on whether the head of these projections has a strong nominal feature:

\[(26)\]

\[
\begin{array}{c}
TP \\
\downarrow \\
T \\
\downarrow \\
\text{vP} \\
\downarrow \\
\text{DP}_1 \\
\downarrow \\
\text{v'} \\
\downarrow \\
\text{VP} \\
\downarrow \\
(DP_2) \\
\downarrow \\
V' \\
\downarrow \\
V \\
\downarrow \\
\text{DP}_3
\end{array}
\]

II.7.4. Locality

In the system proposed in Chomsky (1993), the object moves to Spec of Agro, crossing (the trace of) the subject in Spec of VP; in turn, the subject may move to Spec of Agrs skipping the object in Spec of Agro, as illustrated in (27) below.\(^{25}\) Both instances of movement should yield a relativized

\(^{25}\) For presentation purposes, I omit the TP projection in (27), which is orthogonal to the issue under discussion.
minimality violation in the sense of Rizzi (1990).26

(27) AgrsP
    SU    Agrs'
    Agrs  AgroP
    OB    Agro'
    V+Agro  VP
    tSU    V'
             tv    toB

Chomsky (1993:17-19) circumvents this problem by relying on the notion of equidistance in (28), where \( \alpha \) and \( \beta \) are targets of movement for a category \( \gamma \):

(28) *Equidistance:*

If \( \alpha \) and \( \beta \) are in the same minimal domain, they are equidistant from \( \gamma \).

---

26 Chomsky (1993:43, 49, fn. 52) points out that indices should not be treated as entities and should be replaceable by the structural account of the relation they annotate. I will however continue to use traditional indexing in certain representations, for expository purposes.
Take the derivation of a regular transitive sentence in English. The subject moves overtly to Spec of Agrs without any problems, because the object does not raise overtly in English. At LF, the verb moves to Agro, forming the chain $\text{CH} = (V_i, t_i)$, as shown in (27). The minimal domain of this chain is comprised of Spec of Agro, Spec of VP and the object. Since at this point of the derivation, Spec of Agro and Spec of VP are in the same minimal domain, they are equidistant from the object, according to (28). Thus, the intervening Spec of VP with the trace of the subject does not block the movement of the object to Spec of Agro.

On the other hand, if the subject has accusative Case and moves to Spec of Agro, the object will not be able to move to Spec of Agrs even if it has nominative Case and if Agro moves up. There is no step in this derivation in which a minimal domain contains Spec of Agrs, Spec of Agro and Spec of VP (the potential targets of movement for the object), making it possible for the object to cross the subject in Spec of Agro and its trace in Spec of VP. When Agro moves up, for instance, Spec of VP is not a member of the minimal domain of the chain $\text{CH} = (\text{Agro}_i, t_i)$. Since the trace of the subject in Spec of VP c-commands the object, it is "closer" to Spec of Agrs than the object. Movement of the object over the subject in Spec of Agro would therefore yield a violation of the condition that a derivation should take the "shortest steps", which Chomsky (1994:14) dubs the Minimal Link.
Condition. This analysis therefore correctly rules out the sentence in (29) with the meaning 'he saw her', because the object cannot check its features:

(29) *Him saw she.

It should be observed that the definition of equidistance in Chomsky (1993:17) is not an absolute notion with \(\alpha\) and \(\beta\) ranging over arbitrary nodes in a phrase-marker; rather, equidistance is taken to be a relative notion with \(\alpha\) and \(\beta\) being restricted to potential targets of movement. This is necessary in order to rule out an alternative analysis of (29). If the notion of equidistance were absolute, the subject and object within VP would be equidistant from all of the above Specs, because they are within the minimal domain of the verb. Thus, a nominative object could move directly to Agrs, triggering subject agreement, followed by movement of an accusative subject to Spec of Agro, triggering object agreement.

Notice that cyclicity is not sufficient to rule out the undesirable outcome of this derivation. Had the movements proceeded cyclically, the subject would move first to Spec of Agro, and the verb would then adjoin to Agro, forming the chain \(CH = (V_i, t_i)\), whose minimal domain is comprised of the subject in Spec of Agro, the trace of the subject in Spec of VP, and the object. Since these categories are in the same minimal domain, they should be equidistant from any higher Spec, according to an absolute notion of
equidistance. The object would then be allowed to move to Spec of Agrs, without giving rise to a Minimal Link Condition violation. The unacceptability of (29) would thus be unaccounted for.

Recall that according to the VP structure adopted by Chomsky (1995:321), subjects are inserted in the minimal domain of a light verb, whereas objects are inserted in the minimal domain of a main verb (see section II.7.2). This difference allows Chomsky (1995:356) to formulate an absolute notion of equidistance by computing only minimal domains determined by trivial head chains. The notion of *closeness* is then defined as in (30) below (see Chomsky (1995:356), where $\tau$ is the target of raising and $K$ is the head with which the moved element enters into a checking relation, and the Minimal Link Condition is defined in terms of closeness, as given in (31) (see Chomsky (1995:311)).

\begin{equation}
(30) \quad \text{Closeness:}
\end{equation}

$\alpha$ is closer to $K$ than $\beta$ is iff:

- (i) $\alpha$ c-commands $\beta$; and
- (ii) $\alpha$ is not in the same minimal domain as $\tau$ or $\beta$.

---

27 On the reason for treating movement of $\alpha$ to the minimal domain of $K$ as attraction of $\alpha$ by $K$, see fn. 69 below.
(31) **Minimal Link Condition:**

K attracts α only if there is no β, β closer to K than α, such that K attracts b.

In section II.10.6 below, I will slightly reformulate the definitions of closeness and the Minimal Link Condition in face of the theory of movement to be developed in this dissertation. For the moment, let us consider how the data considered so far can be accounted for, given the definitions in (30) and (31). As a starting point, take the derivational step where the VP in (32) has been assembled.

(32) 
```
          vP
         / \  
        SU  v' 
           /   
          v   VP
             /   
            V   OB
```

Suppose that the language under consideration has overt object movement. Under the framework laid out in Chomsky (1995), that would mean that the light verb of (32) has a strong D-feature (see section II.7.3), triggering overt object movement to the "outer" Spec of v, as illustrated in
The question then is whether the subject in (33) does not prevent the object from moving to the outer Spec of \( vP \). According to the definition of closeness in (30), SU in (33) is not closer to \( tOB \) than OB, because SU and OB are in the minimal domain of the light verb. Hence, no Minimal Link Condition violation arises when the object crosses the subject in structures such as (33).

Consider now the movement of the subject to Spec of TP across the object in the outer Spec of the light verb, as illustrated in (34) below. Since OB and \( tSU \) are in the minimal domain of the light verb, OB is not closer to the Spec of TP than \( tSU \). Hence, according to the definition of closeness in (30), movement of the subject from the inner Spec of \( vP \) to the Spec of TP can proceed across the object in the outer Spec of \( vP \), without yielding a
Minimal Link Condition violation.

By contrast, if a subject with accusative Case moves from the inner to the outer Spec of vP, as illustrated in (35), movement of an object with nominative Case to the Spec of TP will violate the Minimal Link Condition. In (35) SU and tsu both c-command to and neither of them is in the same minimal domain as OB or to (Recall that minimal domains of head chains are not computed). Hence, the unwanted case in (29) is ruled out (see Chomsky (1995:sec. 4.10.2) and section III.4.1.4.3 for further discussion). By restricting the notion of minimal domain to trivial head chains, Chomsky (1995:chap. 4) is thus able to formulate an absolute notion of equidistance,
which entails that head movement is irrelevant for standard A-
movement.\textsuperscript{28,29}

\begin{equation}
(35)
\begin{array}{c}
\text{TP} \\
\text{OB} \quad T' \\
\text{T} \quad \text{vP} \\
\text{SU} \quad \text{v'} \\
\text{t}_{\text{SU}} \quad \text{v'} \\
\text{v} \quad \text{VP} \\
\text{V} \quad \text{t}_{\text{OB}}
\end{array}
\end{equation}

\textsuperscript{28} As Chomsky (1995: sec. 10.2) acknowledges, this analysis loses Holmberg’s (1986) Generalization, according to which overt object movements is contingent on overt verb movement.

\textsuperscript{29} Chomsky (1995:358) also entertains the possibility that the external \(\theta\)-role may be assigned to the outer Spec of \(\text{vP}\) after an object moves to the inner Spec of \(\text{vP}\). Chomsky (1995:358) notes that if it were allowed, the notion of closeness could be simplified along the lines of (i) below. As far as I can see, nothing that follows hinges on this possibility, which I will not explore here.

(i) Closeness
\(\alpha\) is closer to the target \(K\) than \(\beta\) is, if \(\alpha\) c-commands \(\beta\).
II.7.5. Domains, Feature Checking, and q-assignment

We have been assuming that in a clause structure such as the one in (33) the object raises to check its features against the verbal complex formed by adjoining the main verb to the light verb (see fn. 20 and 22). Hidden in this discussion was the assumption that the object and the main verb could not check their features in situ. The question then is whether this is a reasonable assumption to make, given that the main verb and the object are in an extremely local relation in (33). In fact, if locality were only measured by the number of nodes intervening between two categories, the relation between the verb and the object in (33) would be much more local than after V adjoins to V and the object moves to the outer Spec of the light verb.

Although it is plausible that the object must check its Case-feature against the light verb (see fn. 22), the issue still arises with respect to the checking of φ-features in an unaccusative construction, for instance. The unaccusative verb carries φ-features (see fn. 22), which will eventually be checked against the φ-features of its internal argument. What then prevents a checking relation between these two elements without object raising?

Addressing similar problems, Chomsky (1995:312) observes that “θ-theory is virtually complementary to the theory of checking”, which for the current purposes can be interpreted as meaning that θ-assignment is in
complementary distribution with feature checking. If this is true, the object of an unaccusative verb cannot check its $\phi$-features in situ, because it is $\theta$-marked in this position. Similarly, the subject in (32) cannot enter into a checking relation with the light verb despite being in its checking domain (Recall that the subject is $\theta$-marked in the Spec of the light verb). To the extent that is on the right track, this complementarity is arguably related to the fact that $\theta$-role is not a formal feature (see Chomsky (1994:39)).

II.8. Phrase Structure and the Operation Merge

Some parts of standard $X'$-Theory (see Chomsky (1970, 1981, 1986a), (Jackendoff (1977), and Stowell (1981), among others) are not compatible with the Minimalist Program. The categoricity of the lexical item saw in the phrase-structure illustrated in (36) below, for instance, is redundantly encoded by the formal features of the terminal node saw and by the node V, which immediately dominates it. In addition, bar-distinctions in a nonbranching structure such as $[\text{NP} [N [N \text{he } ] ] ]$ are determined neither by the lexical features of he, nor by the relational properties between he and the other constituents of (36), which is at odds with the inclusiveness condition on the mapping from N to $\lambda$.\footnote{Of relevance here is also Kayne (1994), who derives several stipulations of standard $X'$-Theory from his Linear Correspondence Axiom, which governs the}
Chomsky (1994, 1995: chap. 4) develops a "bare" version of X'-Theory which overcomes the problems mentioned above. The operation Merge of the computational system takes two lexical items a and b selected from a numeration and forms a new object K, constituted by a and b. If K is formed by "substitution", it has the form $K = \{g, \{a, b\}\}$, where g is the label of K indicating its relevant properties at the interface levels. The label of K is determined by either a or b. If a, for instance, is the constituent of K which determines its properties, we say that a projects and is the head of K, which is then represented as $K = \{a, \{a, b\}\}$. The working hypothesis is that between the two possibilities for the label of K (a or b), only one allows a convergent

---

mapping from phrase-structure to linear order (see section (II.17)).
derivation in language L or universally.\footnote{Chomsky (1994, 1995:chap. 4) shows that the label of K can be uniquely determined in the case of Move (see section II.11), leaving open the question of whether the label of K can also be uniquely determined in the case of Merge. The assumption is that "wrong" choices for the label of structures formed by Merge cause the derivation to crash. Since I will take Move to be actually decomposed into the operations Copy, Merge, Form Chain and Chain Reduction (see section II.10), Chomsky’s results will be reinterpreted in terms of the operation Form Chain, rather than in terms of Move (see section II.13).}

Under this perspective, the phrase-marker in (36) is formed in the following way: after being selected from a numeration, \textit{saw} and \textit{it} merge and \textit{saw} projects, forming the object \(X\) in (37a); \textit{he} is then selected from the numeration and merges with \(X\), which projects, and the object \(Y\) is formed, as shown in (37b) (an abbreviation of (37c)). Given that \(Y\) is interpreted as a phrase with the relevant properties of \textit{saw}, the label of \(Y\) is taken to be the head of \(X\) rather than \(X\) itself.\footnote{Chomsky (1995) represents complex objects such as \(Y\) in (37b) as \(\{H(X)\}, \{\text{he, } X\}\). Where \(H(X)\) stands for the head of \(X\). When the head of a category can be identified in the representation, I will follow Uriagereka (forthcoming) in representing the label of a projection by underlining the lexical item that determines it, as in (37b). Differently from Uriagereka, however, I will keep the lexical items underlined as more complex objects are formed, as illustrated in (37c). The intuition behind this notation is that when the computational system moves a complex element which is part of a larger syntactic object, the moved element will behave in accordance with its label.}

\begin{align*}
(37) & \quad a. \quad X = \{\textit{saw}, \{\textit{saw}, \textit{it}\}\} \\
& \quad b. \quad Y = \{\textit{saw}, \{\text{he, } X\}\} \\
& \quad c. \quad Y = \{\textit{saw}, \{\text{he, } \{\textit{saw}, \{\textit{saw, } \textit{it}\}\}\}\} \\
\end{align*}

The notions of minimal, maximal, and intermediate projections are
now derivationally and relationally defined.\textsuperscript{33} A category that does not project any further is a maximal projection; a category that is not a projection at all is a minimal projection (a lexical item); any other projection is an intermediate projection, invisible at the interface and for computation (see section II.13.2 for discussion).\textsuperscript{34} Thus, before *saw* and *it* merge in (37a), each lexical item is a minimal maximal projection. After they merge forming X, *saw* loses its maximal projection status and becomes a minimal projection; *it* does not change its phrasal status; and X is a maximal projection. *He* is a minimal maximal projection before and after merging with X, which becomes an intermediate projection after merger (see Epstein (1995)). Finally, Y is a maximal projection.\textsuperscript{35}

Merging $\alpha$ and $\beta$ by adjunction yields a different kind of object: $K = \{<\gamma, \gamma>, \{\alpha, \beta\}\}$, a two-segment category (see section II.9.3 below for some revision). If $\alpha$ projects, for instance, $K = \{<H(\alpha), H(\alpha)>, \{\alpha, \beta\}\}$.

To sum up, under bare X'-Theory, the redundancy concerning categorial features existing in standard X'-Theory is eliminated; phrasal

\textsuperscript{33} The idea that the different types of projections are derived from structural relations among categories is due to Muysken (1982).

\textsuperscript{34} Adjunction structures aside (see section II.9.3), a complement is conceived of as the sister of a minimal nonmaximal projection, and the specifier as the sister of an intermediate projection. In addition to standard notions of head, (nonminimal) maximal projection, and intermediate projection, Chomsky (1995:chap. 4) makes use of the notion $X^{\text{omax}}$, which refers to the object formed by the head X of a projection XP and other heads or features adjoined to X (see section II.14. for discussion).

\textsuperscript{35} As Chomsky (1995:249) observes, a system which allows projections which are both minimal and maximal has the necessary tools to account for the behavior of clitics, which appear to move as heads and as maximal projections.
status is determined configurationally and derivationally; and there are no vacuous nonbranching projections.\textsuperscript{36}

II.9. Some Definitions

II.9.1. Term

In standard X'-Theory representations, the functioning elements correspond to nodes in a phrase-marker. In bare X'-Theory, the functioning elements correspond to terms, as defined in (38) (see Chomsky (1995:247)).

\textbf{(38) Term:}

For any syntactic object $K$,

(i) $K$ is a term of $K$;

(ii) if $L$ is term of $K$, then the members of the members of $L$ are terms of $K$.

Consider the structure in (39), for instance.

\textsuperscript{36} In order to capture the distinction between unaccusative and unergative verbs without the correspondent of a nonbranching $V'$ projection, Chomsky (1994:13; 1995:352) relies on Hale and Keyser's (1993) theory, according to which unergatives are actually transitives.
According to the base step in (38i), M in (39) is a term of M. If M is a term of M, then the members of the members of M are also terms of M, according to recursive step in (38ii). M has two members: the label $\alpha$, which is irrelevant for our purposes because it has no members, and the set $\{\delta, \{<\alpha, \alpha>, \gamma, \{\alpha, \{\alpha, \beta\}\}\}\}$, whose members are $\delta$ and $L$. Thus, $\delta$ and $L$ are terms of M by (38ii). $L$ in turn has two members: the label $<\alpha, \alpha>$, which is not relevant because it is not a set, and the set $\{\gamma, \{\alpha, \{\alpha, \beta\}\}\}$, which has $\gamma$ and $K$ as members. Given that $L$ is a term of M, and $\gamma$ and $K$ are members of a member of $L$, $\gamma$ and $K$ are also terms of M, according to the recursive step of (38). Finally, the set $K$ has two members: the label $\alpha$, which is irrelevant in virtue of not having any members, and the set $\{\alpha, \beta\}$. Thus, by the recursive step, $\alpha$ and $\beta$ are also terms of $L$ and M.

Notice that the label of a syntactic object formed by adjunction such as $L$ in (39) is constructed from the term that projects, $\alpha$ in the case of $L$, but is not
identical to it (or to any other term).

II.9.2. Domination

Nunes and Thompson (forthcoming) adapt Chomsky’s (1994) definition of term in (38) to define domination relations in terms of the set-theoretic objects used in a bare phrase-structure representation, as shown in (40):

\[(40) \quad \text{Domination:} \]

\[
\text{Given a syntactic object } \alpha \text{ such that } \alpha \text{ has the form } \{\gamma, \{\delta, \mu\}\} \text{ or } \{<\gamma, \gamma>, \{\delta, \mu\}\}, \alpha \text{ dominates } \beta \text{ iff:}
\]

(i) \( \beta \in A \) and \( A \in \alpha \); or

(ii) \( \alpha \) dominates \( A \) and \( A \) dominates \( \beta \).

Let us see how this definition works by reconsidering the structure in (39). In (39), \( \alpha \) and \( \beta \) are members of the set \( \{\alpha, \beta\} \), which is a member of the category \( K = \{\alpha, \{\alpha, \beta\}\} \). Hence, according to the base step of (40), \( K \) dominates \( \alpha \) and \( \beta \). In turn, \( K \) and \( \gamma \) are members of the set \( \{\gamma, \{\alpha, \{\alpha, \beta\}\}\} \), which is a member of \( L = \{<\alpha, \alpha>, \{\gamma, \{\alpha, \{\alpha, \beta\}\}\}\} \). \( L \) then dominates \( K \) and \( \gamma \) by (40i).
Since L dominates K and K dominates \( \alpha \) and \( \beta \), as we have just seen, L dominates \( \alpha \) and \( \beta \) by the recursive step in (40ii). Finally, \( \delta \) and L are members of the set \( \{ \delta, \{<\alpha, \alpha>, \{\gamma, \{\alpha, \beta}\}\}\} \), which is a member of the set \( M = \{\alpha, \{\delta, \{<\alpha, \alpha>, \{\gamma, \{\alpha, \beta}\}\}\}\} \). Thus, M dominates \( \delta \) and L by the base step and all the elements dominated by L by the recursive step, namely, \( \gamma, K, \alpha, \) and \( \beta \).\(^{37}\)

**II.9.3. Adjunction Structures Revisited**

The notation \( K = \{<\alpha, \alpha>, \{\alpha, \beta]\}\) is taken by Chomsky (1994:15, 1995:248) to correspond to the two-segment category \([\alpha_2, \alpha_1]\) of the standard X'-Theory representation in (41) below. Alternatively, one could take \( K = \{<\alpha, \alpha>, \{\alpha, \beta]\}\) to be a segment built from \( \alpha \), corresponding to \( \alpha_2 \) in (41), and the two-segment category corresponding to \([\alpha_2, \alpha_1]\) in (41) to be actually \( L = [K, \alpha] \), whose segments are \( K \) and the category that projects.

\(^{37}\) Empirical reasons having to with Binding Theory (see Lasnik and Uriagereka 1988:fn. 14) and linear order (see Kayne 1994:chap. 3, fn. 8) may require that dominance be an irreflexive relation. As pointed out by Nunes and Thompson (forthcoming), nothing to this effect needs to be stipulated, though. If dominance is defined along the lines of (40), it will inherit irreflexivity from set membership (a set is not a member of itself). See Nunes and Thompson (forthcoming) for discussion.
The distinction between these two interpretations becomes relevant when domination relations are computed, as shown below. Let us assume the definition of domination for categories given in (42) (see May (1985) and Chomsky (1986a)), and the definition of c-command (in terms of categories) given in (43) (see Chomsky (1994:30)):\footnote{38 I will henceforth ignore the exclusion requirement of the definition of c-command given in (13ii), which is orthogonal to the issues under discussion.}

(42) \textit{Categorial domination:}

A category $\alpha$ dominates $\beta$ iff every segment of $\alpha$ dominates $\beta$.

(43) \textit{C-command:}

Where $\alpha$ and $\beta$ are accessible to the computational system, $\alpha$ c-commands $\beta$ iff:

(i) $\alpha$ does not dominate $\beta$;  
(ii) $\alpha \neq \beta$; and  
(iii) every category dominating $\alpha$ dominates $\beta$.  

\begin{figure}[h]
\centering
\begin{tikzpicture}
    \node (a) {$\alpha_1$};
    \node (b) [right of=a] {$\beta$};
    \node (c) [above of=a] {$\alpha_2$};
    \draw (a) -- (c) -- (b);
\end{tikzpicture}
\caption{(41)}
\end{figure}
The relevant questions here are: (i) in an adjunction structure such as
\[ K = \{<\alpha, \alpha>, \{\alpha, \beta}\} \], does \( K \) dominate \( \alpha \)?; and (ii) in this structure, is \( K \) a segment of a two-segment category, or is it a two-segment category? In order to answer these questions, let us consider verb movement to \( T \), as represented in (44).

\[
(44) \quad M = \{T, \{L, K\}\} \\
\quad L = \{<T, T>, \{V, T\}\} \quad K = \{V, \{\ldots\}\} \\
\quad V \quad T \\
\quad \text{Vi} \quad \text{ti} \quad \text{...} \\
\]

Under the standard assumption that c-command is a condition on chains (see section II.10.3), it must be the case that the moved verb in (44) is able to c-command its trace. In order for this to hold, every category dominating \( V \) should also dominate its trace, according to (43). If \( L \) is a two-segment category, as proposed by Chomsky (1994, 1995), \( V \) will not be able to c-command its trace: according to (40), \( L \) dominates \( V \), but not the trace of \( V \). If \( L \) is a segment of the two-segment category \([L, T]\), on the other hand, it does not prevent \( V \) from c-commanding its trace. The question then is whether the two-segment category \([L, T]\) does.

According to (42), in order for the category \([L, T]\) to dominate \( V \), both of its segments must dominate \( V \). \( L \) dominates \( V \) by the base step of (40): \( V \) is a
member of the set \{V, T\}, which is a member of the set \( L = \{<T, T>, \{V, T\}\} \). T, however, does not dominate V. Given that one of the segments of \([L, T]\) does not dominate V, the whole two-segment category does not dominate V either.\(^{39}\)

Based on these considerations, I will henceforth treat a structure such as \( K = \{<\alpha, \alpha>, \{\alpha, \beta\}\} \) as a segment of a two-segment category \( L = \{K, \alpha\} \).\(^{40}\)

II.10. The Copy + Merge Theory of Movement

Given the syntactic object \( \Sigma \) with terms \( K \) and \( \alpha \), as shown in (45a), the operation \textit{Move} targets \( K \), raises \( \alpha \), and merges \( \alpha \) with \( K \), forming \( \Sigma' \), as shown in (45b). \( \Sigma' \) differs from \( \Sigma \) in that \( K \) is replaced by \( L = \{\gamma, \{\alpha, K\}\} \) or

\(^{39}\) A similar argument can be constructed for sentences which arguably involve noncyclic adjunction of a relative clause such as (ia) (see section III.2 for details), represented in (ib). Assuming that the wh-phrase in the Spec of CP has the structure as in (ic), it must be the case that \( M = \{<\text{which, which}>, \{K, L\}\} \) is a segment of the two segment category \([M, K]\) and not a category in its own right; otherwise, the moved wh-phrase cannot c-command its trace. See Nunes and Thompson (forthcoming) for discussion.

(i) a. Which portrait that Picasso painted did he like?
   b. [ [ [ which portrait ], [ that Picasso painted ] ] did he like t_i ]
   c. \( M = \{<\text{which, which}>, \{K, L\}\} \)

\[ \begin{align*}
K &= \{\text{which, [which, portrait]}\} \\
L &= \{\text{that Picasso painted}\}
\end{align*} \]

\(^{40}\) Since head movement, for instance, can target a head with another head adjoined to it, it must be the case that two-segment categories are accessible to the computational system and are able to c-command their traces (see Nunes and Thompson (forthcoming) for discussion).
\( L = \{ \langle \gamma, \gamma \rangle, \{ \alpha, K \} \} \), depending on whether movement proceeds by substitution or adjunction (see Chomsky (1993:22, 1994:fn. 13, 1995:sec. 4.4)).

(45) a. \( \Sigma = \{ ??, \{ \beta, K \} \} \)

\[ \begin{array}{c}
\beta \\
K \\
\ldots \alpha \ldots
\end{array} \]

b. \( \Sigma' = \{ ??, \{ \beta, L \} \} \)

\[ \begin{array}{c}
\beta \\
L = \{ ??, \{ \alpha, K \} \} \\
\alpha \\
K \\
\ldots \alpha \ldots
\end{array} \]

The operation Move also forms the chain \( CH = (a_i, t_i) \), a two-element pair where \( t \) is the trace of \( a \). As will be discussed in detail in sections III.2 and III.3, Chomsky (1993) incorporates into the Minimalist framework the "copy theory of movement", according to which a trace is a copy of the

\[ \text{\footnotesize \cite{Uriagereka:forthcoming:chap.4, NunesThompson:forthcoming}} \]

for discussion of this "overwriting" process that takes place in the mapping from \( \Sigma \) to \( \Sigma' \) in (45).
moved element which gets eliminated at PF but remains present at LF (see Chomsky (1993:35)). The inner workings of the operation Move thus encompass: (i) a suboperation of copying; (ii) a suboperation of merger; (iii) a suboperation forming a chain between two copies; and (iv) a suboperation deleting traces for PF purposes.

The complexity of the operation Move is a historical residue of the description of generalized transformations as binary or singulary operations: binary transformations target two disconnected syntactic objects, whereas singulary transformations target two constituents of a single syntactic object (see Chomsky (1975, 1993) and Kitahara (1995)). In a framework which takes Merge to be an operation of the computational system, as in Chomsky (1994, 1995:sec. 4.4), it is an odd fact from a conceptual point of view that Merge is an operation in its own right and is also part of the inner workings of Move. This becomes clearer when economy computations are considered. Merge and Move are taken to be comparable for economy purposes (see Chomsky (1994, 1995:chap. 4) and section II.15 for discussion), even though Merge is also a suboperation of Move.

One of the main goals of this dissertation is to show that Move is not a primitive operation of UG. I propose that Move is merely a description of the effects of the interaction of four different operations which correspond to the suboperations in (i)-(iv) above: Copy, Merge, Form Chain, and Chain
Reduction.\textsuperscript{42} I will refer to this approach as the Copy + Merge theory of movement.

The Copy + Merge theory of movement has two major conceptual advantages over an approach which takes Move to be a complex operation. First, it will not be the case that Merge will behave as a full operation in certain cases and as a suboperation in others; Merge will always be taken as an operation and, as such, will be comparable to other operations in terms of economy (see section II.10.1 for the comparison between "lexical insertion" and "movement"). And second, by analyzing Move in terms of Copy and Merge, we will be able to eliminate Procrastinate as a primitive of UG and derive its effects in compliance with the uniformity condition on the mapping from N to \( \lambda \) (see section II.10.2).

Empirical evidence for the Copy + Merge theory of movement will be provided by derivations where applications of Copy and Merge are dissociated from applications of Form Chain and Chain Reduction. The approach proposed here allows derivations where a term T is copied from the structure K and merges with the structure L, unconnected to K, as illustrated in (46).

\textsuperscript{42} As will become clear, the operation Form Chain to be proposed here is different from the operation Form Chain proposed in Chomsky (1993:15), which converts a structure such as (ia) into (ib) in a single step, yielding the chain CH = (\textit{John}, \textit{t}', \textit{t}).

\begin{enumerate}
\item a. \textit{e seems \{ e to be likely \} \{ John to win \} ]}
\item b. \textit{John seems \{ t' to be likely \} \{ t to win \} ]}
\end{enumerate}
I will refer to the sequence of derivational steps such as the one in (46) as *sideward movement*, borrowing the term from Chomsky (1995:253). The sideward movement illustrated in (46) is not possible in a framework which takes Move as a complex operation involving the suboperations in (i)-(iv) above; the two instances of T cannot form a chain because they do not enter into a c-command relation. I will show that constructions involving noncyclic adjunction of relative clauses (see section II.10.3), parasitic gap constructions (see section IV.2), and constructions involving across-the-board extraction (see section IV.3) should be analyzed in terms of sideward movement, providing empirical evidence against taking Move as a "singulary transformation" with the suboperations in (i)-(iv), and in favor of the Copy + Merge theory of movement.

Brody (1995) has pointed out that if movement operations and chain formation express the same type of relation, a theory which contains both is

---

43 Uriagereka (forthcoming:chap. 4) refers to a sequence of derivational steps such as the one in (46) as *paracyclic movement*. 
redundant. Given that the applications of Copy and Merge in (46) are not automatically associated with chain formation, this redundancy is overcome and an independent operation of Form Chain is required. The conditions on the application of Form Chain are discussed below.

In this chapter I will focus the discussion on the operations Copy, Merge and Form Chain of the Copy + Merge theory of movement, leaving a detailed discussion of Chain Reduction to chapters III and IV. I will nonetheless continue to use familiar terms such as movement and raising, for instance, when nothing bears on the issue. Below I examine how the properties which were previously ascribed to Move should be interpreted under the Copy + Merge theory of movement.44

II.10.1. Derivational Cost of the Copy Operation

As will be discussed in detail in section II.15, an operation $O$ blocks an operation $O'$ at a given derivational step if both operations lead to convergent derivations and $O$ is more economical than $O'$. In the system

44 Chomsky (1995:257 and 268) takes Last Resort and the Minimal Link Condition to be part of the definition of Move; hence, the derivation is canceled if they are violated. If these conditions as well as c-command are to be interpreted as conditions on the application of Form Chain, as I will propose below, failure to meet them entails that a chain cannot be formed, but not necessarily that the derivation should terminate or crash. The conditions under which failure to form a chain may or may not yield an unacceptably result are discussed in detail in chapters III and IV. For current purposes, I will simply translate Chomsky’s assumptions into the Copy + Merge framework and assume that if a copied element does not form a chain with the original term at some derivational step, the derivation terminates.
proposed in Chomsky (1994, 1995:chap. 5), overt insertion of a lexical item pulled out from a numeration in a structure $\Sigma$ is thus less costly than overt movement of a term from within $\Sigma$, because overt movement violates Procrastinate (see section II.6).

Under the Copy + Merge theory, lexical insertion corresponds to the sequence of derivational steps involving Select and Merge, where movement (for current purposes) corresponds to the sequence of derivational steps involving Copy and Merge. The fact that overt lexical insertion is less costly than overt movement must then be due to the difference in derivational cost between Select and Copy.

As discussed in section II.4, Select is a defining property of a derivation; if it does not exhaust the numeration, no derivation is generated (i.e. the derivation is canceled). If Copy does not apply, on the other hand, a derivation may be generated, even though $l$ or $p$ may not be a legitimate object. In other words, Copy is related to convergence conditions. Assuming Chomsky's (1995:225-226) proposal that operations which are defining properties of derivations are costless whereas operations associated with convergence conditions are derivationally costly (see section II.4), Copy is costly and Select is costless. Thus, regardless of whether movement operations (the sequence of derivational steps involving Copy and Merge) take place overtly or covertly, they will always be more costly than lexical
insertion (the sequence of derivational steps involving Select and Merge).

II.10.2. Deriving Procrastinate

As pointed out in section II.6, by establishing an inherent asymmetry between overt and covert movement, Procrastinate is at odds with the uniformity requirement on the mapping from N to \( \lambda \). Once it is established that Copy is derivationally costly, as seen in section II.10.1, the effects of Procrastinate can now be derived in compliance with the uniformity condition, as shown below.

Given that Select is costless and Copy is costly, the computational system will always prefer the sequence of derivational steps involving Select and Merge (lexical insertion) over the sequence involving Copy and Merge (movement), if both options lead to convergence. Suppose then that the computational system has assembled a structure \( \Sigma \) before Spell-Out in an optimal way. In other words, sequences of derivational steps involving Copy and Merge were employed only to check strong features, and only in cases when a sequence involving Select and Merge would not yield a legitimate pair \((\pi, \lambda)\). If Select has already exhausted the numeration, the sequence of derivational steps involving Select and Merge is no longer available.\(^{45}\) The

\(^{45}\) On the lexical insertion of null complementizers, see section II.14.3.3 below.
next step then is either to apply the Spell-Out rule or to resort to the sequence of derivational steps involving Copy and Merge, now starting to check "weak" features.

As discussed in section II.5.1, Spell-Out is a defining property of a derivation; if it does not apply, the pair \((\pi, \lambda)\) is not formed and the derivation is canceled. Spell-Out therefore is derivationally costless and should be preferred over Copy, if the two options lead to convergence. In the case at hand, both options allow the derivation to converge, since strong features have already been checked. Thus, the computational system applies the Spell-Out rule, and the remaining applications of Copy and Merge which are necessary for the derivation to converge take place in the covert component.

Therefore, it need not be the case that overt movement is inherently more costly than covert movement, as stipulated by Procrastinate. Rather, this asymmetry follows from the fact that once strong features are checked and the numeration is exhausted, Spell-Out is less costly than (the sequence of derivational steps involving) Copy (and Merge). Furthermore, these general economy considerations allow us to eliminate Procrastinate as a principle of UG, while deriving its effects without violating the uniformity condition on the mapping from \(N\) to \(\lambda\).46 In the sections that follow, my use

---

46 Nunes (1994a) and Kitahara (1995) derive the effects of Procrastinate from the fact that overt movement requires extra computations in the phonological component having to do with deletion of traces. Although these approaches are
of the term *Procrastinate* should thus be understood only as a descriptive term summarizing the comparison between Copy and Spell-Out in terms of derivational cost.

### II.10.3. C-command

Let us consider the standard assumption that a moved element must c-command its trace (see Chomsky (1981:333)). Epstein (1995) argues that c-command is a property of Merge and Move, and proposes the derivational notion of c-command in (47).47

\[(47) \quad \text{C-command:} \]

\[\alpha \text{ c-commands all and only the terms of the category } \beta \text{ with which } \alpha \text{ was paired by Merge or Move in the course of the derivation.} \]

---

consistent with the uniformity condition on the mapping from N to λ, they rely on global computations; ate every step in the derivation before Spell-Out, the computational system "looks ahead" in order to determine whether or not a given operation will bring derivational cost to the phonological component. The proposal developed above, on the other hand, is able to derive Procrastinate by considering a single derivational step.

47 Under the Copy + Merge theory of movement, (47) should be simplified along the lines of (i), which is actually Epstein’s (1995) preliminary definition.

\[(i) \quad \text{C-command:} \]

\[\alpha \text{ c-commands all and only the terms of the category } \beta \text{ with which } \alpha \text{ was merged in the course of the derivation.} \]
Assuming Epstein's (1995) theory, it would be redundant to assume a c-command condition on chains if a chain were formed whenever a movement operation took place (see Brody (1995)). Under the Copy + Merge theory of movement, however, applications of Copy and Merge do not go hand in hand with applications of Form Chain. As mentioned above, the Copy + Merge theory of movement in principle allows instances of sideward movement, as illustrated in (46) and repeated below in (48).

\begin{align*}
(48) & & \text{a.} & & K & & L \\
& & & & \ldots T \ldots \\
& & & & \\
& & & & \text{b.} & & K & & M = \{??, \{T, L\}\} \\
& & & & \ldots T \ldots & & T & & L
\end{align*}

In (48), T enters into a c-command relation with L in accordance with (47), but does not c-command its "trace" within K. Thus, if it can be argued that sideward movement is an option permitted by UG (see Uriagereka (forthcoming:chap. 4)), we will have evidence that a c-command condition on chain formation may be necessary, regardless of whether or not c-command is a property of Merge.

Let us put this issue aside for a moment and consider Chomsky's (1995:sec. 4.3) discussion of the cyclicity condition on Merge operations.
According to him, an operation merging \( \gamma \) and either of the constituents of \( K = \{a, \{a, \beta\}\} \) actually embeds \( \gamma \) in a structure already formed. Such an operation is taken to be considerably more complex than an operation merging \( \gamma \) and \( K \); thus, the simpler option is chosen.

Assuming this to be true, consider the sentence in (49a) below, which in principle can have two different derivations. Under the cyclic derivation, the relative clause merges with the wh-phrase before movement and is then pied-piped when the wh-phrase moves to Spec of CP to check the strong wh-feature of the interrogative complementizer \( Q \) (see section II.14.3.3 for discussion), as illustrated in (49b) (irrelevant traces omitted). Under the noncyclic derivation, on the other hand, the relative clause merges with the head of the wh-chain after movement proceeds, as shown in (49c). According to Chomsky's simplicity arguments presented above, the computational system should not allow the structure in (49c), given that Merge has not proceeded cyclically. However, as will be discussed in detail in section III.2, the fact that \( \text{Picasso} \) and \( he \) can be interpreted as coreferential in (49a) strongly suggests that the relative clause has not merged with the wh-phrase before movement took place. Under the standard theory of Move, we thus have conceptual reasons not to permit structures such as (49c) and empirical evidence pointing in the opposite direction.
a. Which portrait that Picasso painted did he like?

b. [[ which portrait ] [ that Picasso painted ] ] did+Q he like

[ [ which portrait ] [ that Picasso painted ] ]

c. [[ which portrait ] [ that Picasso painted ] ] did+Q he like

[ which portrait ]

In a Copy + Merge analysis of structures such as (49a), on the other hand, conceptual considerations meet empirical demands very naturally. At some point in the derivation of (49a) under the relevant interpretation, we have the two unconnected phrase-structures in (50a) and (50b), which have been independently assembled. The phrase which portrait is then copied, but instead of merging with the object in (50a), becoming the specifier of the interrogative complementizer Q, it merges with the relative clause in (50b), forming the object in (50c); finally, the object in (50c) merges with the object in (50a), yielding the structure in (49c) and allowing Q to check its strong wh-

---

48 Assembly of unconnected phrase-markers is independently required for any phrase involving a complex specifier or a complex adjunct (see chapter IV for relevant discussion). For instance, consider the partial derivation of (ia), after the verb has already been merged with the object, forming the object K in (ib). Suppose that we merge man with K, yielding the object L in (ic), and then we merge the and L, forming the object M in (id). In this derivation, the and man do not form a constituent. Thus, in order for the correct structure to be formed, the and man should merge independently of K, as shown in (ie), and the resulting object should then merge with K, as shown in (if).

(i) a. The man saw the woman.
b. K = {saw, [saw, the, [the woman]]}
c. L = {saw, [man, K]}
d. M = {saw, [the, L]}
e. N = {the, [the, man]}
f. O = {saw, [K, N]}
feature.

(50) a. [ did+Q [ he like [ which portrait ] ] ]

b. [ that Picasso painted ]

c. [ [ which portrait ] [ that Picasso painted ] ]

To the extent that we can account for data such as (49a) without resorting to noncyclic operations as in the standard analysis of these constructions, the Copy + Merge theory of movement seems more attractive than a Move-based approach. I will examine more data in chapter IV, where I propose an analysis of parasitic gap and across-the-board extraction constructions in terms of sideward movement.\(^{49}\) Right now, let us return to the discussion of the c-command condition on chain formation, which prompted this digression.

The question was whether such a condition would not be redundant, if Epstein's (1995) theory were assumed. The Copy + Merge analysis of (49a) has shown that it is not the case that whenever a copied element α merges with a given structure K in a convergent derivation, α will c-command its trace. Although the copied phrase *which portrait* ends up c-commanding its

\(^{49}\) See Uriagereka (forthcoming: chap. 4), where an analysis of overt head movement in terms of sideward movement is discussed. Given T and VP, for instance, the head of VP can be copied and merged with T forming the object \(K = \{<\text{T}, \text{T}>, \{\text{T}, V]\}\), which then merges with VP, forming \(L = \{\text{T}, \{K, VP]\}\). In no step in this derivation is cyclicity violated.
trace at a later point in the derivation, it does not do so at the point where it merges with the relative clause (see fn. 39).

I will thus assume that the "representational" notion of c-command, as defined in (43) and repeated below in (51), is a condition on the application of Form Chain, as stated in (52).

(51) \textit{C-command:}

Where \( \alpha \) and \( \beta \) are accessible to the computational system, \( \alpha \) c-commands \( \beta \) iff:

(i) \( \alpha \) does not dominate \( \beta \);

(ii) \( \alpha \neq \beta \); and

(iii) every category dominating \( \alpha \) dominates \( \beta \).

(52) \textit{Form Chain:}

The syntactic objects \( \alpha \) and \( \beta \) can form the nontrivial chain \( \text{CH} = (\alpha, \beta) \) only if \( \alpha \) c-commands \( \beta \).

In section II.13.2, it will be shown that the standard assumption that chain links must be in a c-command relation, as stated in (52), provides us with the means to advance a considerable simplification of the systems proposed in Chomsky (1994, 1995:chap. 4) regarding the determination of the
II.10.4. Nondistinctiveness of Copies

If the copy theory of movement is assumed, one needs to determine whether two terms with the same set of features are to be interpreted as distinct, or as nondistinct and therefore able to form a chain. Chomsky (1994:7, 1995:sec. 4.2.1) proposes that two lexical items $l$ and $l'$ should be marked as distinct for the computational system if they are formed by distinct

---

If it turns out to be true that chain links must be in a c-command relation, it is not obvious how Epstein's (1995) theory will be able to handle movement operations involving (noncyclic) adjunction. In the case of head movement, for instance, the moved element is only paired with its host and will not c-command its trace according to the definition in (47). Cases of adjunction to the head of the chain such as the one exemplified in (49c) also present problems. In the cyclic derivation of (49c), sketched in (50), the copied wh-phrase will not be able to c-command its trace according to (47), because it is only paired with the relative clause, which does not include the trace of the wh-phrase as one of its terms. One could say that c-command is possible in the noncyclic derivation of (49c), given that at least one point in this derivation the moved wh-phrase c-commands its trace, namely, before the adjunction of the relative clause. Although plausible, this "once and forever" notion of c-command is incompatible with the "First Law" of Epstein's (1995) theory, stated in (i) below (his emphasis).

(i) **First Law:**

$T_1$ can enter into a c-command (perhaps, more generally, syntactic) relation with $T_2$ only if there exists no derivational point at which:

(a) $T_1$ is a term of $K_1$ ($K_1 \not\in T_1$);

(b) $T_2$ is a term of $K_2$ ($K_2 \not\in T_1$); and

(c) There is no $K_3$ such that $K_1$ and $K_2$ are both terms of $K_3$.

Assume that $T_1$ is the moved wh-phrase, $T_2$ is the trace of the wh-phrase, $K_1$ is the segment dominating $T_1$ and the relative clause, $K_2$ is term immediately dominating $T_2$, and $K_3$ is the root term. In the noncyclic derivation of (49a), there is presumably a derivational step at which there is a term $K_3$ with terms $K_1$ and $K_2$, namely, after the relative clause adjoins to the wh-phrase in Spec of CP, forming $K_1$. 
applications of Select accessing the same lexical item of a numeration. Two occurrences of the pronoun *he*, for example, may have different properties at LF depending on whether they arose via two distinct applications of Select or via one application of Select and one application of Move. As Chomsky (1995:sec. 4.2.1) observes, this is nevertheless a departure from the inclusiveness condition on the mapping from N to λ.

I propose that a better way to approach this issue is to take a term of a given structure to be inherently distinct from all the other terms of that structure unless it is specified as being a copy. In other words, the output of a copying operation targeting a term T is a term T', which is interpreted at the C-I interface as nondistinct from T or other copies of T, and distinct from every other term by default. The copying operation can be informally described as follows: (i) if a term T has no index, Copy targets T, assigns an unused index i to it and creates a copy of the indexed term; or (ii) if a term T was already indexed by a previous copying operation, Copy simply creates a copy of the indexed term.

Applied to the term *the man* in the object position of *kissed* in (53), for instance, Copy assigns the index i to it and creates a copy of the indexed term, as shown in (53b). The two syntactic objects in (53) then merge, allowing T to have its strong D-feature checked. The two instances of *the man* in (53c) can then form the chain \( CH = ([DP \text{ the } man]_i, [DP \text{ the } man]_i) \).
(53) a. \[ TP \text{ was } [VP \text{ kissed } [DP \text{ the man }]] ] \\
    b. \[ TP \text{ was } [VP \text{ kissed } [DP \text{ the man }]] ] \\
       [DP \text{ the man }]] \\
    c. \[ TP \ [DP \text{ the man }])i \text{ was } [VP \text{ kissed } [DP \text{ the man }]] ] \\

The motivation for treating terms related by the Copy operation as nondistinct rather than identical has to do with feature checking, as will be discussed in detail in chapter III. Take the derivation in (53), for instance. In (53b), both instances of the term \textit{the man} have their Case-features unchecked; in (53c), on the other hand, the upper copy of \textit{the man} checks its Case-feature against \( T \), whereas the lower copy still has its Case-feature unchecked. Therefore, the two instances of \textit{the man} in (53c), although nondistinct, are not identical. Further evidence showing that nondistinctiveness is the relevant notion for chain formation will be adduced in chapters III and IV.

The reasons for switching from encoding distinctiveness through Select to encoding nondistinctiveness through Copy are twofold. First, it pinpoints the natural locus of departure from the inclusiveness condition on the mapping from \( N \) to \( \lambda \). In other words, the system is well-behaved in interpreting different terms as distinct; the departure occurs only when copies are involved.

Second, although still noninclusive, the nondistinctiveness approach
is computationally simpler. Under the approach proposed in Chomsky (1994, 1995), the system has to keep track of every application of Select throughout the derivation, which is almost equivalent to assigning indices to every lexical item pulled out from the numeration. If nondistinctiveness is encoded by the Copy operation, on the other hand, the system only has to keep track of the terms which are related by the copying operation.

Given that the Minimalist framework attempts to dispense with indexation (see fn. 26), the structural relation that indices annotate as a byproduct of the Copy operation must be made clear. Roughly speaking, they annotate the relation between a single term and multiple structural configurations. In other words, indexation under copying is meant to capture the fact that the Copy operation yields a discontinuous object.

It may be possible to determine whether or not a term is a copy of another term if we keep track of the history of the derivation (see Chomsky's (1994, 1995:chap. 4)) notion of reference set discussed in section II.15). If a term T is introduced into the derivation and the numeration corresponding to that derivational step is kept constant, T must have arisen through a Copy operation. Further computations would then be necessary in order to determine which term T is a copy of. Whether or not it is desirable that the recognition of copies by the computational system proceed along these lines remains to be determined. For concreteness, I will assume that part of the inner workings of the Copy operation is to ensure that its input
and output are made nondistinct, which I will encode by cosuperscripting, as illustrated in (53).

Assuming the requirement that chain links must be nondistinct to be a condition on the application of the operation Form Chain, (52) should be reformulated as in (54).

(54)  \textit{Form Chain}:

The syntactic objects $\alpha$ and $\beta$ can form the nontrivial chain $CH = (\alpha, \beta)$ only if:

(i) $\alpha$ c-commands $\beta$; and

(ii) $\alpha$ is nondistinct from $\beta$.

II.10.5. Last Resort

Chomsky (1993, 1994, 1995:chap. 4) proposes movement operations must be morphologically driven by the need to check some feature, the \textit{Last Resort} condition. Given that Merge and Copy are not independently subject to this condition, Last Resort should then be understood as a condition on Form Chain in the Copy + Merge theory of movement. (54) should then be revised as in (55).
Form Chain:

The syntactic objects $\alpha$ and $\beta$ can form the nontrivial chain $\text{CH} = (\alpha, \beta)$ only if:

(i) $\alpha$ c-commands $\beta$;
(ii) $\alpha$ is nondistinct from $\beta$; and
(iii) $\alpha$ participates in a checking relation.

I will return to a more refined definition of Last Resort in sections II.14 and III.6.2.5.

II.10.6. The Minimal Link Condition

Considerations similar to the ones in section II.10.5 extend to the Minimal Link Condition (see section II.7.3), repeated below in (56).

(56) Minimal Link Condition:

K attracts $\alpha$ only if there is no $\beta$, $\beta$ closer to K than $\alpha$, such that K attracts $\beta$.

Given that the legitimacy of the operations Merge and Copy does not
hinge on the notion of closeness, repeated below in (57) (where τ is the target of raising and K is the head with which the moved element enters into a checking relation), the Minimal Link Condition is to be reinterpreted as a condition on chain formation, along the lines of (58).51

(57)  Closeness:

α is closer to K than β is iff:

(i) α c-commands β; and

(ii) α is not in the same minimal domain as τ or β.

(58)  Minimal Link Condition:

Given the syntactic objects α and β such that α c-commands β and a certain feature F of α enters into a checking relation, α can form a chain with β if there is no element γ such that:

(i) γ has a feature F' which is of the same type as the feature F of α; and

(ii) γ is closer to α than β is.

51 The issue of when two features count as the same type will be addressed in detail in section II.14.14. For relevant discussion, see Ferguson and Groat (1994).
It should be noticed that conceptual reasons independent of the Copy + Merge theory of movement would also lead us to state the Minimal Link Condition as a condition on chain formation, as in (58), rather than a condition on the application of a movement operation, as in (56). (58) says that given a syntactic object, two of its terms can form a chain if they are sufficiently close to one another; in (56), on the other hand, positions which have not been created yet (the targets of movement) are taken into consideration in order to determine whether a movement operation can take place.

The latter approach is plausible within a richer framework assuming standard X'-Theory, which allows projected empty slots to be filled in the course of the derivation (see Chomsky (1993) and Kitahara (1995), for instance). From a bare X'-Theory perspective, however, the formulation of the Minimal Link Condition in (56) is rather dubious. A "position" is created in a phrase-structure only when an element is introduced in that position by Merge; hence, the computational system arguably cannot resort to a position which does not exist at a certain derivational step in order to allow or disallow a given operation, as is required in (56).

Similar considerations lead us to simplify the notion of closeness in (57) along the lines of (59):
(59) \textit{Closeness:}

Where $\gamma$ c-commands $\alpha$ and $\alpha$ c-commands $\beta$, $\alpha$ is closer to $\gamma$ than $\beta$ is iff:

(i) $\alpha$ c-commands $\beta$; and

(ii) $\alpha$ is not in the same minimal domain as $\gamma$ or $\beta$.

(60) below incorporates the version of the Minimal Link Condition given in (58) to the operation of Form Chain,

(60) \textit{Form Chain:}

The syntactic objects $\alpha$ and $\beta$ can form the nontrivial chain $\text{CH} = (\alpha, \beta)$ only if:

(i) $\alpha$ c-commands $\beta$;

(ii) $\alpha$ is nondistinct from $\beta$;

(iii) a given feature $F$ of $\alpha$ enters into a checking relation; and

(iv) there is no syntactic object $\gamma$ such that $\gamma$ has a feature $F'$ which is of the same type as the feature $F$ of $\alpha$, and $\gamma$ is closer to $\alpha$ than $\beta$ is.
As mentioned in fn. 44, still missing in this reformulation of chain formation is the reason why failure to form a chain may cancel the derivation. I will postpone the discussion of this issue until section III.6.2.5, where the subject will be discussed in detail.

II.11. The Label of Structures Formed by Movement

In pre-Minimalist versions of the Principles and Parameters Theory, the phrasal status of the objects participating in a given movement operation and the phrasal status of the resulting structure were determined by some primitives of X'-Theory combined with some version of Emonds's (1976) Structure Preservation Hypothesis. Since these resources are not available to the more parsimonious Minimalist versions of the Principles and Parameters Theory, it remains to be shown that the label of a structure formed by movement and the phrasal status of both the moved element and the target of movement can be established under Minimalist assumptions about bare phrase structures.

In this section, I present the essentials of Chomsky's (1994, 1995:sec. 4.4.2) analysis, where the label of structures formed by overt movement is predicted by the interaction between Last Resort and the additional condition on chains stated in (61) below (see Chomsky (1995:253)).^52 In

---

^52 I will use the term *Phrasal Uniformity Condition* rather than Chomsky's (1995:253) *Uniformity Condition*, because the latter has been used in the literature with
section II.13, I return to this issue, proposing an alternative analysis which does away with (61).

(61) *Phrasal Uniformity Condition:*

A chain is uniform with regard to phrase structure status.

Let us start with instances of movement targeting a root node, as sketched in (62), where a raises to target K:

(62) \[ L = \{?, \{\alpha, K\}\} \]

K in (62) is a root structure before \(\alpha\) raises and merges with it, forming L. Assume that this is an operation of "substitution". The question then is what the label of L is. Suppose that K projects, forming \(L = \{H(K), \{\alpha, K\}\}\); K then becomes an intermediate projection and \(\alpha\) becomes its specifier (see fn. 34). Since \(\alpha\) is in the checking domain of the head of K, these two elements can enter into a checking relation and eventually satisfy Last Resort. Assume

\[ \text{a different meaning (see Chomsky 1995:chap. 4, :fn. 75).} \]
that they do and that Last Resort is satisfied. We now need to consider the phrasal status of α before it moves in order to finally determine whether the structure L = \{H(K), \{α, K\}\} is a well formed object.

α can in principle be: (i) a maximal projection (with internal structure); (ii) a maximal head (with no internal structure); or (iii) a nonmaximal head (a head which has a complement). Take the first two possibilities. Since the phrasal status of α does not change after it moves to Spec of K, the Phrasal Uniformity Condition is satisfied. Now consider the third possibility. Before moving, α is a nonmaximal head; after K projects, the upper copy of α becomes a maximal head. Since the head of the chain and its trace do not have the same phrasal status, the Phrasal Uniformity Condition is violated, and L = \{H(K), \{α, K\}\} is not a licit structure. We thus derive the fact that maximal projections (with or without internal structure) can move to specifiers, but (nonmaximal) heads cannot.

Consider the above scenario with α projecting, forming L = \{H(α), \{α, K\}\}. If α was a nonminimal maximal projection before movement, the upper copy becomes an intermediate projection after projecting and K now becomes its specifier. Although it could in principle satisfy Last Resort, this derivation does not yield a well formed structure because it violates the Phrasal Uniformity Condition. Suppose, on the other hand, that α was a
head (maximal or nonmaximal) before movement. After the moved $\alpha$ projects, $K$ becomes its complement (see fn. 34). Given that no checking relation can obtain in head-complement configurations (see section II.7.1.2), Last Resort is violated.\(^{53}\) Hence, for substitution operations, if $\alpha$ raises and targets $K$ to form $L$, the target determines the label of $L$.\(^{54}\)

---

53 If $\alpha$ were a maximal head before movement in this case, the Phrasal Uniformity Condition should be violated, as well.

54 Consider how we can block instances of "self-attachment" such as the one illustrated in (i), where the verb moves targeting its own VP. As discussed above, if $V$ projects forming $L = \{V, \{V, K\}\}$, Last Resort is violated because $K$ becomes the complement of the upper $V$ and no checking relation obtains in this configuration. Chomsky (1995:320-321) observes that if $K$ projects, the object formed, namely $L = \{H(K), \{V, K\}\}$, is identical to the object formed when $V$ projects because $H(K) = V$. He then suggests that such an ambiguity should be interpreted as a crashed derivation. That, however, would be an unmotivated representational condition on phrase-structure building, which is taken to be strictly derivational.

(i) $L = \{?, \{V, K\}\}$

A simpler nonrepresentational approach is available, though. Suppose that $V$ was a nonmaximal head with some complement before movement. After $K$ projects in (i), the head of the verb chain becomes a maximal projection and the Phrasal Uniformity Condition is violated. Furthermore, Last Resort is also violated in (i) if $L = \{H(K), \{V, K\}\}$: given that $H(K) = V$, $H(K)$ and $V$ presumably do not have the appropriate feature specification to check each other's features.

For the sake of completeness, suppose now that before movement, $V$ is a maximal head, a verb without any complement (a possibility that may be excluded on independent grounds), as illustrated in (ii), where $V$ is the "moved" verb and $K$ is both its target and its trace. The structure in (ii) violates the Phrasal Uniformity Condition: $V$ is a maximal head, whereas its trace $K$ becomes a nonmaximal head after projecting. (ii) also violates Last Resort: $V$ becomes a complement of $K$ after $K$ projects (see fn. 34) and no checking configuration obtains; even if a checking configuration were possible, Last Resort would still be violated under the plausible assumption that no element has the appropriate features to check its own features. This last point also blocks self-attachment by adjunction, regardless of which element projects.
Take now the case where \( L \) in (62) is formed by an adjunction operation. If \( K \) projects, Last Resort will be violated regardless of the phrasal status of \( \alpha \). Since an element adjoined to a nonminimal projection of a head \( H \) is not in the checking domain of that head (see section II.7.1.2), no checking relation can be established between \( \alpha \) and the head of \( K \) in (62), if \( L = \{<H(K), H(K)>, \{\alpha, K\}\} \). The same violation obtains if \( \alpha \) is a maximal projection before moving and projects in (62) forming \( L = \{<H(\alpha), H(\alpha)>, \{\alpha, K\}\} \): no checking relation between \( K \) and the head of \( \alpha \) can be established.

If \( \alpha \) is a nonmaximal head, on the other hand, \( K \) will be in a checking domain of \( \alpha \). Chomsky (1995:319) proposes, however, that since Morphology deals only with \( X^0 \) elements and their features, "the morphological component gives no output (so the derivation crashes) if presented with an element that is not an \( X^0 \) or a feature."\(^{55}\) Presumably, this morphological requirement also rules out \( L = \{<\alpha_\alpha>, \{\alpha, K\}\} \) in (62), where \( \alpha_\alpha \) is a maximal head.

The discussion above shows that if \( \alpha \) moves to target a root structure \( K \),

\[
\begin{align*}
(ii) \quad L &= \{K, \{V, K\}\} \\
V &\quad K = V
\end{align*}
\]

\(^{55}\) If Morphology gives no output when presented with elements that are not \( X^0 \), as proposed by Chomsky (1995:319), the derivation should be canceled, rather than crash at PF, because presumably no PF object is formed (see section II.4).
forming a two-segment category, no licit structure is obtained regardless of which element projects. Under Minimalist assumptions, apparent instances of movement of a maximal projection by adjunction should thus be reanalyzed as movement by substitution to a specifier.

Let us now consider the only instance of movement by adjunction that takes place overtly: (noncyclic) head movement. Given the structure in (63) with K being a head, α adjoins to K forming the object N = \{<\gamma, \gamma>, \{\alpha, K\}\}, which then replaces K in M, forming M’ = \{K, \{N, L\}\}, as shown in (64).

The question then is whether the label of N in (64) is α or K. If N = \{<\alpha, \alpha>, \{\alpha, K\}\}, M’ is not a legitimate syntactic object because its label is distinct from the head of either of its constituents. Such a structure
presumably cancels the derivation.\textsuperscript{56} By contrast, if $N = \{<K, K>, \{\alpha, K\}\}$, $M'$ is a well formed object. Again, we reach the conclusion that only the target can project. Since $\alpha$ is in the checking domain of $K$ after the movement, Last Resort can be satisfied if they enter into a checking relation.

Consider now the phrasal status of $\alpha$. If $\alpha$ is a nonminimal maximal projection, the derivation will presumably be canceled (see fn. 55) because Morphology is only able to deal with $X^0$ elements.\textsuperscript{57} If $\alpha$ is a maximal head in (64), Morphology can operate with the $X^0$ element $N = \{<K, K>, \{\alpha, K\}\}$, and the derivation may eventually converge. That will be the case of a clitic adjoining to an inflectional head, for instance (see Kayne (1991), Martins (1994), and Uriagereka (1995a), among others).

If $\alpha$ is a nonmaximal head in (64), on the other hand, the Phrasal Uniformity Condition should be violated because the head of the chain would become a maximal head after the target projects. Were this the case,

\textsuperscript{56} Notice that covert movement in the framework of Chomsky (1994) would change the constituents but not the label of a category. Consider covert moment of the object to Spec of Agro in English, for instance. Before Spell-Out, AgroP is the category $K = \{\text{Agro, [Agro, VP]}\}$ and TP is the category $L = \{T, \{T, K\}\}$. At LF the object $O$ raises targeting $K$, forming the category $N = \{\text{Agro, [O, K]}\}$. $N$ should then replace $K$ in $L$, forming the category $L' = \{T, \{T, N\}\}$. Since the label of $L'$ is determined by one of its constituents, $L'$ is a licit syntactic object. This framework therefore assumes that noncyclic movement can change the constituents of a category, as long as its label remains the same (see the discussion in Chomsky (1994:15, 1995:250)).

\textsuperscript{57} In addition, if $\alpha$ is a nonminimal maximal projection and c-command is defined in terms of categories (cf. (51)), then the structure $N = \{<K, K>, \{\alpha, K\}\}$ in (64), cannot be linearized (cf. Kayne 1994:19 and section II.17 below). $\alpha$ will asymmetrically c-command the constituents of $L$ (the complement of $K$), and $L$ will asymmetrically c-command the constituents of $\alpha$, violating the Linear Correspondence Axiom.
however, every instance of head movement would be blocked. In order to circumvent this undesirable result, Chomsky (1995:322) assumes that "at LF, word-like elements are 'immune' to the algorithm that determines phrase structure status", and introduces the condition in (65):

(65) "At LF, $X^0$ is submitted to independent word-interpretation processes $WI$, where $WI$ ignores principles of $C_{HL}$ [the computational system; JMN], within $X^0$. (Chomsky (1995:322))

If $WI$ renders the phrasal status of $\alpha$ unavailable for inspection at LF, the Phrasal Uniformity Condition will be not violated and a derivation involving head movement may eventually converge.

Finally, consider instances in which a head $H$ moves to adjoin to a head $\alpha$, which is already adjoined to another head $K$. Suppose, for example, that a head $H$ moves from within $L$ in (64) and adjoins to $\alpha$, forming the object $O = \{<\gamma, \gamma>, \{H, \alpha\}\}$; $O$ then replaces $\alpha$ in $N$, forming $N' = \{<K, K>, \{O, K\}\}$, which in turn replaces $N$ in $M'$, yielding $M'' = \{K, \{N', L\}\}$, as shown in (66).
If H projects so that $O = \{<H, H>, [H, \alpha]\}$, $N'$ is a licit syntactic object, since its label is determined by one of its constituents. Nonetheless, the derivation forming $N'$ violates the morphological requirements of $K$. In order for the adjunction of $\alpha$ to $K$ to be a licit operation satisfying Last Resort, $K$ must require an affix of the type $\alpha$ (see Chomsky (1995:260)). On the other hand, this morphological requirement is satisfied in $N'$ of (66) if $O = \{<\lambda, \gamma>, [H, \alpha]\}$, because the label of $O$ is determined by $\alpha$ and not by $H$. Thus, if (66) satisfies Last Resort, then $O = \{<\alpha, \alpha>, [H, \alpha]\}$. Again, the conclusion is that the target of movement projects, as desired.

---

58 Notice that this reasoning entails that the requirement of an affix with the feature $F$ may be different from the requirement of a checking relation with the feature $F$. Were these requirements the same, $\alpha$ could have checked the $\alpha$-feature of $K$ in (64), allowing $H$ to project after moving in (66).

59 Implicit in Chomsky's (1995:260) analysis is the assumption that if $K$ requires an affix $\alpha$, this requirement may be satisfied if Morphology is provided with the object $M = \{K, [K, N]\}$, where $N$ is a segment built from $\alpha$ ($N = \{<\alpha, \alpha>, [\alpha, H]\}$).
II.12. Move-F and Covert Movement

Chomsky (1995:262-263) observes that if movement operations are triggered by feature checking, Minimalist considerations would lead us to expect Move to operate with features, rather than categories. Apparently, however, this expectation systematically fails to be met. A core property of human languages is that they place categories (lexical items and phrases) in positions different from the ones where they are interpreted. The question then is why the language faculty has such a departure from optimality.

Chomsky’s (1995:262-263) answer is that this departure is illusory. The operation Move does target features; however, properties of the phonological component require that when a feature of a lexical item or a phrase moves, all the other features of that category be pied-piped. Morphology presumably is not able to operate with isolated features or other scattered parts of words. Thus, overt movement of a feature F has the appearance of movement of a category containing F.

Consider, for instance, the movement of the wh-phrase in (67a) below, and assume that the wh-phrase has the structure in (67b) (see the discussion in Chomsky (1995:263)). Suppose that the null interrogative complementizer of (67a) has a strong wh-feature, which Chomsky (1995:289) takes to be a variant of a D-feature.60 If Move targets only the wh-feature of who,

\footnote{Recall that for Chomsky (1995:232), strong features are only checked by categorial features (see fn. 15).}
Morphology will not be able to deal with the structure formed and the derivation will be canceled (see fn. 55); thus, the remaining formal, semantic and phonological features of *who* must be pied-piped to ensure convergence at PF. This much of pied-piping is not enough, though. If only *who* moves, the affix 's will be stranded and the derivation will not converge; hence the affix must be pied-piped, as well.\(^{61}\) Move cannot target *who's*, however, because this is not a syntactic object.

\[(67) \quad \text{a. Whose book did you read?} \]

\[\text{b.} \quad \begin{array}{c}
\text{DP} \\
\text{who} \\
\text{D'} \\
\text{'s} \\
\text{book}
\end{array} \]

Therefore, the only way for the wh-feature of (67b) to check the strong feature of the complementizer in compliance with other morphological requirements is for the whole phrase *whose book* to move to Spec of CP, as shown in (67a).\(^{62}\) Thus, overt movement of categories can actually be

\(^{61}\) As Chomsky (1995:263) points out, if a language does not have an affixal D in structures such as (37b), the wh-element in Spec of DP will be allowed (in fact required) to move without further pied-piping. See Uriagereka (1988) for a correlation between left branch extraction of a wh-element and the nature of the head of DP.

\(^{62}\) Notice that in (67b), it is the DP *whose book* and not the wh-prase *who* that is in the checking domain of the null complementizer. Questions remain as to how the wh-prase *who* can then enter into a checking relation with the strong feature of the complementizer in such a configuration.
understood as the operation Move-F, where F is a feature (see section II.14 for further refinements), combined with generalized pied-piping triggered by morphological requirements of the phonological component.

Chomsky (1995:262) assumes that economy considerations require that only material needed for convergence be pied-piped. Under the Copy + Merge theory of movement, that would mean that the bigger the number of features or terms to be copied, the more derivationally costly the operation (see fn. 64 below). Thus, if the wh-phrase whose book is in an embedded clause and the complementizer of the matrix clause has a strong wh-feature, for example, it is not the case that the whole embedded clause is pied-piped to the Spec of the matrix CP, as shown in (68a); rather, only the minimal object which is able to check the strong feature without violating morphological requirements is moved, namely the wh-phrase whose book, as shown in (68b).

(68)  
   a. *That you read whose book did you say?  
   b. Whose book did you say that you read?

Since covert movement does not feed Morphology, it need not (therefore must not) resort to generalized pied-piping. Chomsky (1995:265) restricts pied-piping in the covert component only to formal features: if

---
Move targets a feature F of a lexical item LI, it automatically carries along the set of formal features of LI (FF(LI)). As evidence that covert movement involves FF(LI) rather than a given category containing FF(LI), Chomsky (1995:275) presents the contrast between (69a) and (69b) (see Lasnik and Saito (1992) and den Dikken (1995) for relevant discussion):

(69) a. They seemed to each other to have been angry.

b. *There seemed to each other to have been many linguists given good job offers.

---

64 Notice that pied-piping the remaining formal features of LI when a feature F of LI is moved at LF should in principle be excluded by the same economy considerations excluding pied-piping in (68a). If it is true that when a feature F moves, it automatically pied-pipes the remaining formal features of its set, this apparent departure of optimality needs to be accounted for.

It could be the case that the Copy operation just happens to deal with sets of features or sets of sets of features, but not with single features. Another possibility to consider is that movement in the covert component may actually target heads, which only have formal and semantic features after Spell-Out.

Yet another possibility is that the derivational cost with respect to feature copying may take into consideration three variables: number of features copied, number of applications of Copy, and number of checking relations made available by the features copied. The idea is that the most economical derivational step is the one which allows the greatest number of checking relations with fewest number of features in a single application of Copy. Compare the sentences in (68), for instance. Under the assumption that a single operation of Copy was employed and a single checking relation was established in each sentence, (68a) and (68b) differ in terms of the number of features copied. Since it has fewer features copied than (68a), the derivation of (68b) is more economical. Consider now verb movement at LF in a language such as English, for instance. The verb has to check its T-feature against T, and its φ-features against the subject in Spec of TP. We could in principle have multiple applications of Copy targeting each relevant feature of the verb, which would then merge with T and establish the required checking relation. On the other hand, if the set of formal features of the verb is copied and then merged with T, all the required checking relations will obtain with a single application of Copy, being therefore the more economical option according to the metric outlined above. I will leave the choice among these three options pending on further research.
Under the assumption that in both (69a) and (69b), the formal features of the anaphor adjoin to Infl in the covert component (see section III.2), the contrast between these two sentences should follow from the difference between moving the category they to the checking domain of T overtly in (69a) and moving the relevant FF(many linguists) covertly in (69b). Were the whole phrase many linguists to move covertly either replacing there or forming another specifier, no contrast between (69a) and (69b) should obtain. If only FF(many linguists) moves and adjoins to T for the reasons discussed below, an explanation for the unacceptability of (69b) may then be provided, under the assumption that neither structure in (70) is a legitimate binding theoretic configuration (see Chomsky (1995:275)).

(70)

a. \[ \tau [FF(each other)] [\tau [FF(many linguists)] [\tau T]]]  
b. \[ \tau [FF(many linguists)] [\tau [FF(each other)] [\tau T]]] 

This analysis raises questions as to (i) whether covert movement may involve both adjunction and substitution; and (ii) whether or not the target must project. In order to address these questions, assume that FF(LI) moves

---

65 Chomsky’s (1995) analysis of the contrast between (69a) and (69b) in terms of the impossibility of structures such as (70) predicts that in languages with subjects in situ (if such languages exist; see Jonas and Bobaljik (1993:74) and Chomsky (1995:chap. 4, fn. 80)), anaphors cannot be bound by local subjects. Since these languages will have structures similar to (70) after the formal features of both the anaphor and its local subject adjoin to T at LF, no local binding will be possible. Whether or not this prediction holds true, I do not know.
to check a feature of the head K of the category $M = \{K, \{K, L\}\}$ in (71).

\[
M = \{K, \{K, L\}\}
\]

(71)

Suppose that FF(LI) and K merge by substitution, forming the new category $N = \{\emptyset, \{FF(LI), K\}\}$. Regardless of the label of N in this case, FF(LI) is not in the checking domain of K, nor is K in the checking domain of FF(LI). Hence, no checking relation can be established, and Last Resort is violated. Last Resort will also be violated if FF(LI) adjoins to M in (71); since M is a nonminimal projection of K, FF(LI) will not be in the checking domain of K (see section II.7.1.2).66

Suppose then that FF(LI) and K are merged by adjunction and FF(LI) projects, yielding the segment $O = \{<FF(LI), FF(LI)>, \{FF(LI), K\}\}$; O then replaces K in M, forming $M' = \{K, \{O, L\}\}$. Although this derivation could license a checking relation between FF(LI) and K, it is canceled because $M' = \{K, \{O, L\}\}$ is an illegitimate syntactic object; its label is determined by neither of its constituents.

---

66 If the notion of projection is extended to FF(LI) (see below), merger of FF(LI) and the root object M in (71), yielding the object $N = \{<FF(LI), FF(LI)>, \{FF(LI), M\}\}$, in principle allows a checking relation to obtain. However, it is plausible to think that such structure would receive a deviant interpretation at the interface, since the root syntactic object would not encode clause type (interrogative, declarative, etc.).
Only two options are left: either FF(LI) merges with M in (71) becoming the specifier of K, or FF(LI) adjoins to K and K projects. Both possibilities could yield well formed syntactic objects and license checking relations between FF(LI) and K. The question now is whether or not the chain (FF(LI), t_i) is subject to the Phrasal Uniformity Condition, which in turn depends on whether or not the notion of projection should be extended to FF(LI).

Chomsky (1995:271) proposes that regardless of the final answer for these questions, only one of these options is possible. Assume that the notion of projection is extended to FF(LI). The chain (FF(LI), t_i) will then violate the Phrasal Uniformity Condition, if FF(LI) moves to the Spec of K: the moved FF(LI) is a maximal projection (it does not project any further), whereas its trace is not. By contrast, if FF(LI) adjoins to K, the Phrasal Uniformity Condition is inapplicable, because WI exempts instances of adjunction to a head from the Phrasal Uniformity Condition (cf. (65)). Therefore, only adjunction of FF(LI) to K is permitted. Suppose, on the other hand, that the notion of projection does not extend to FF(LI). Given that a specifier is a maximal projection by definition, movement of FF(LI) to the specifier of K is blocked. Again, the only possible option for a checking relation to obtain between FF(LI) and K is for FF(LI) to adjoin to K, which then projects.

To sum up, covert movement is restricted to movement of FF(LI) targeting a head K, forming the segment \( L = \{<K, K>, \{\text{FF(LI)}, K\}\} \).
II.13. Eliminating the Phrasal Uniformity Condition and WI

Let us reconsider the role that the Phrasal Uniformity Condition plays in Chomsky's (1994, 1995:chap. 4) system. Putting aside the instances where its effects are redundantly replicated by Last Resort, the Phrasal Uniformity Condition is responsible for the following:

(72)   a. It blocks movement of a nonmaximal head to a Spec position;
    b. it blocks movement of FF(LI) to a Spec position (if the notion of projection is extended to FF(LI)); and
    c. it prevents a moved nonminimal maximal projection from projecting after it merges with its target.

The Phrasal Uniformity Condition nevertheless has the undesirable consequence that it rules out every instance of head movement (and every instance of adjunction of FF(LI) to a head, if the notion of projection is extended to FF(LI)). This state of affairs leads Chomsky (1995:322) to introduce the condition in (65), repeated below in (73), which has the effect of exempting these cases from the Phrasal Uniformity Condition.
"At LF, X^0 is submitted to independent word-interpretation processes WI, where WI ignores principles of C_{HL} [the computational system; JMN], within X^0^\text{\textsuperscript{\textprime}}. (Chomsky (1995:322))

Conceptually, this is not an optimal design. Why should UG have a condition whose only effect is to void another condition? Below I offer an alternative analysis which will derive (72a)-(72c) from strictly Minimalist assumptions, therefore making it possible to eliminate both the Phrasal Uniformity Condition and the WI processes as described in (73).

II.13.1. Checking Domains Revisited

As seen in section II.7.1.2, the checking domain of a head H involves two different structural relations: the specifier(s) of H and the head(s) or features adjoined to H (see fn. 18). This constitutes a departure from optimality; Minimalist considerations would lead us to expect only one notion to be relevant. Between these two structural relations, the elements adjoined to H form the most natural configuration for the checking domain of H. After all, a given element moves to enter into a checking relation with the features of H, not with the projection formed by H and its complement. In other words, adjunction of a head or FF(LI) to another head should be the canonical rule, rather than the exception, as the conjunction of the Phrasal
Uniformity Condition and the condition in (73) implies.

I propose that this departure from optimality in the configurations encompassing the notion of a checking domain finds its roots in the morphological subcomponent of the phonological component. The need for a checking configuration other than the one expected under Minimalist assumptions arises from a morphological conflict. On the one hand, maximal projections may be pied-piped when a feature F is moved in the overt syntax in order for the morphological requirement banning features external to $X^0$ elements to be satisfied (see section II.12). On the other hand, maximal projections cannot adjoin to the head with which they enter into a checking relation (the optimal option), because Morphology presumably cannot operate with nonminimal maximal projections within $X^0$ elements (see Chomsky (1995:319) and section II.11). In order for overt checking relations to be obtained in compliance with both morphological requirements, UG must then resort to the Spec-head relation in addition to the optimal checking configuration established by adjunction to a head.

Given that this conflict only arises in the overt syntax, covert movement sticks to the optimal option: FF(LI) adjoins to the head with which it enters into a checking relation. Under this perspective, one need not resort to the Phrasal Uniformity Condition to exclude movement of a nonmaximal head or FF(LI) to a Spec position. Movement to the Spec of a head H is only triggered if the optimal option of adjunction to H is not
available. Notice also that movement to the Spec of H is always "longer" than adjunction to H, where "length" can be determined by the number of terms which dominate the trace but not the moved element (see Collins (1994:56) for a formulation of this idea). If the length between the links of a chain should be as short as possible (see Chomsky (1994:14)), adjunction to the head H will always outrank movement to the Spec of H, if both operations yield convergent derivations.\textsuperscript{67}

If the movements described in (72a) and (72b) can be prevented without the Phrasal Uniformity Condition, as proposed above, we have no need for the condition in (73). Minimalist considerations thus lead to a much simpler and conceptually natural system. If we could extend this analysis further and account for (72c) independently, the Phrase Uniformity Condition could be entirely dispensed with. This is the subject of the next section.

\textbf{II.13.2. The C-command Condition Revisited}

From the discussion above, the only remaining independent role of the Phrasal Uniformity Condition is to prevent maximal projections from projecting after they merge with their targets (cf. (72c)). This instance of movement can also be prohibited, however, if we make the standard

\textsuperscript{67} The same economy considerations would also block the possibility raised in fn. 66, where FF(LI) raises and merges with the root structure M, yielding the object N = \{<FF(I), FF(LI)>, \{FF(LI), M\}\}.
assumption that the links of a chain must be in a c-command relation (see section II.10.3).

If the maximal projection $\alpha$ projects after moving and merging with $K$, forming $L = \{\alpha, \{\alpha, K\}\}$ in (74) below, the upper copy of $\alpha$ will be an intermediate projection. As will be shown in II.17, in order for the linearization of a syntactic object to obtain in a bare $X'$-Theory framework, it must be the case that intermediate projections do not enter into c-command relations (see Chomsky (1995:336)). Assuming this to be true, the pair $(\alpha, \alpha)$ in (74) cannot form a chain because the upper link does not enter into c-command relations.

(74) 
\[
L = \{\alpha, \{\alpha, K\}\}
\]

\[\alpha \quad K
\]

\[\ldots \alpha \ldots\]

One could counterargue that on this account, we are merely replacing one condition by another one, without obtaining any theoretical gain. There are several reasons which indicate that this is not the case, however. The first one is methodological. C-command is a notion which appears to be the basis of almost every relation which takes place in the mapping from $N$ to $\lambda$. 
(see Epstein (1995) and Frank and Vijayashankar (1995), for instance). On the other hand, the Phrasal Uniformity Condition, would be specific to a grammatical domain; in fact, it would be only required to ensure (72c), if the considerations in section II.13.1 are on the right track. All things being equal, methodological considerations lead us to adopt a condition which is independently required in other domains, instead of postulating a domain-specific condition.

The second reason for an account in terms of c-command instead of the Phrasal Uniformity Condition is that the former allows us to eliminate a stipulation concerning the movement of intermediate projections, whereas the latter does not. In the system proposed in Chomsky (1994, 1995:chap. 4), intermediate projections are exceptional in that besides being unable to enter into c-command relations, they cannot undergo movement. This last exceptional property of intermediate projections can be shown to follow from the first one if the links of a chain must be in c-command relation (see Kayne (1994:17)), as I assume here.

Consider the structure in (75), for instance, where the intermediate projection K moves and merges with M:
We have already seen that if the movement of \( K \) in (75) proceeds by adjunction, Last Resort will be violated regardless of the label of \( N \) (see sections II.7.1 and II.11). Given that an element adjoined to a nonminimal projection of a certain head is not in the checking domain of that head, no checking relation can be established if the upper copy of \( K \) adjoins to \( M \) or \( M \) adjoins to the upper copy of \( K \). Let us then focus on movement by substitution, in which case either the upper copy of \( K \) will become the specifier of \( M \) or \( M \) will become the specifier of the upper copy of \( K \), depending on which category projects. If the head of \( M \) projects so that \( N = \{ \delta, \{ K, L \} \} \), the upper copy of \( K \) will be a maximal projection and, hence, able to enter into a c-command relation. The lower copy of \( K \) will not, however, because it is an intermediate projection. So the pair \((K, K)\) cannot form a chain in this instance. In turn, if the upper copy \( K \) projects forming \( N = \{ \alpha, \{ K, L \} \} \), both copies of \( K \) will be intermediate projections and, consequently,
neither of them will be able to enter into a c-command relation. Again, no chain \( CH = (K, K) \) can be formed.

Therefore, if we impose a c-command condition on chain formation (see section II.10.3), we can derive the fact that intermediate projections cannot undergo movement. The Phrasal Uniformity Condition has nothing to say on this issue. In fact, in the situation where the upper copy of \( K \) projects, the pair \( (K, K) \) satisfies the Phrasal Uniformity Condition, because both instances of \( K \) are intermediate projections. To conclude, the effects of (72c) should be captured in terms of the c-command condition on chain formation, and the Phrasal Uniformity Condition can be entirely dispensed with.

As a final remark, it should be noticed that in order for chains of the type \( (FF(LI), FF(LI)) \) formed by covert movement to be licit objects, \( FF(LI) \) should be able to enter into c-command relations. Chomsky (1995:274) argues that the licensing of control structures exemplified by the Italian sentence in (76a) and its French equivalent in (76b) provides independent evidence showing that this is indeed the case:
In Italian, the verbal complex agrees with its internal argument and PRO in the adjunct clause can be controlled by the internal argument; on the other hand, neither is true in French. This contrast can be captured if the difference in agreement correlates with covert raising of the relevant FF of the internal argument. If so, FF(tre uomini) in (76a) adjoins to T in the covert component and from this position, it will be able to c-command and therefore license PRO; in French no such raising takes place and PRO cannot be licensed.

To summarize, by sticking to strictly Minimalist considerations and adopting the standard c-command condition on chain links, we are able to eliminate the Phrasal Uniformity Condition from the theory of grammar. By doing so, the WI process in (73) is then rendered obsolete.
II.14. Interpretability, Checking Theory, and Last Resort

As discussed in section II.2, a derivation converges only if each object of the pair \((\pi, \lambda)\) satisfies Full Interpretation. Let us focus on the LF object \(\lambda\) formed from a given numeration \(N\) and see how it is able to comply with Full Interpretation (for relevant discussion with respect to \(\pi\), see sections II.5.1 and III.6).

Consider the lexical features present in \(N\). Surely, not all of them receive an interpretation at the C-I interface, for example. If they do not, they must then be eliminated somehow in the course of the derivation in order for \(\lambda\) to comply with Full Interpretation. Take semantic features. Since they are interpretable at the C-I interface, not only are they allowed to be part of \(\lambda\), but in fact they cannot be eliminated. No stipulation to that effect is necessary, though. A derivation which converges with an operation OP eliminating semantic features will always be less economical than a competing derivation which does not employ OP. By contrast, phonological features are not interpretable at the C-I interface and, by hypothesis, cannot be eliminated in the mapping from \(N\) to \(\lambda\) (see sections II.5 and II.6). However, these features do not cause the derivation to crash at LF if they are inserted overtly, because they are stripped away by Spell-Out (see section II.5).
Formal features, in turn, may or may not be assigned an interpretation at the C-I interface, depending on their type and on the category they are associated with. A Case-feature, for instance, never receives an interpretation, whereas [+plural] receives an interpretation if it is part of a noun, but not if it is part of a verb. [-interpretable] features should thus be eliminated by LF, in order for $\lambda$ to satisfy Full Interpretation. Chomsky (1995:280) proposes that elimination of [-interpretable] features through the checking operation follows the algorithm in (77).

\begin{equation}
\text{(77) a. A checked feature is deleted when possible.}
\text{b. Deleted $\alpha$ is erased when possible.}
\end{equation}

Deletion renders some object as "invisible at the interface but accessible to the computation", while "erasure is a 'stronger form' of deletion, eliminating the element entirely so that it is inaccessible to any operation, not just to interpretability at LF" (Chomsky (1995:280)). Application of either operation is taken to be governed by independently motivated grammatical considerations:\(^{68}\)

---

\(^{68}\) One wonders whether or not the interpretability of a given feature is already encoded as part of the lexical item makeup. If it is, the computational system is able to determine at any derivational point whether a given feature can or cannot be deleted after checked\(\chi\) if interpretability is not lexically encoded, on the other hand, the computational system can decide on whether or not to delete a checked feature only after "looking ahead" at LF. If the system should avoid computational complexity (see Chomsky 1995:227-228), the most optimal approach is to take the set of formal features to be comprised of a subset of [+interpretable] features and a subset of
"Possibility" in (77) [my numbering throughout; JMN] is understood relative to other principles. Thus, deletion is "impossible" if it violates principles of UG. Specifically, a checked feature cannot be deleted if it contradicts the overriding principle of recoverability of deletion, which should hold in some fashion for any reasonable system: Interpretable features cannot delete even if checked. The question of erasure, then, arises only for a -Interpretable feature F, which is erased by (77b) unless that operation is barred by some property P of F. One such property is parametric variation: F could be marked as not erased when checked (...). Erasure is also barred if it creates an illegitimate object, so that no derivation is generated. (Chomsky (1995:280-281))

Interpretability is taken to be closely related to accessibility to the computational system. Since [+interpretable] features are not deleted (and, therefore, not erased), they are always accessible to the computational system; [-interpretable] features, on the other hand, are accessible when unchecked, or when checked and deleted but not erased. Last Resort can then be defined as in (78) (see Chomsky (1995:280)), where F ranges over accessible features and a sublabel of K is either a feature of the head of K itself or a feature adjoined to the head of K (that is, a feature of the head [-interpretable features. Hence, at any derivational step, the computational system is able to identify which checked features are to be deleted. The principle of recoverability of deletion under this view can be derived from economy considerations: a derivation which converges without deleting a given [+interpretable] feature will always outrank a competing derivation which deletes it. For concreteness, I will be assuming in the following discussion that the interpretability of a given feature is lexically encoded, as suggested above.
K^{\text{max}}; see fn. 34):\footnote{In a movement operation which satisfies Last Resort, the target always has [-interpretable] features to be checked. Based on this, Chomsky (1995:297), following a suggestion by John Frampton, reinterprets Move-F as "feature attraction" by a target, $\text{Attract-F}$. Last Resort as defined in (78) replaces the different notions of Greed proposed in Chomsky (1993:33, 1994:14) and the notion of Enlightened Self-Interest of Lasnik (1993, 1995). For the differences between these notions and further discussion, see Lasnik (1995) and Chomsky (1995:sec. 4.3).}

(78) \textbf{Last Resort:}

Move F raises F to target K only if F enters into a checking relation with a sublabel of K.


It is not clear that the distinction between deletion and erasure, as developed in Chomsky (1995:sec. 4.5.2) and described above, is conceptually justified under Minimalist assumptions. The most visible problem concerns the reason why the system should bother to erase checked [-interpretable] features, given that, by assumption, they are already made "invisible" at the interface by the checking operation. If the whole purpose of feature checking is to render [-interpretable] features inert at LF so that Full Interpretation is satisfied, economy considerations should block extra operations wiping out these features. Of course, this innovation is an attempt to distinguish which features the computational system may or may not have access to. However,
the overall adjustments which are then made necessary compromise such an attempt.

The distinction between deletion and erasure appears to be relevant in only two sets of cases: transitive expletive constructions (cf. (23)) and successive movement of the expletive *there*. Let us consider whether each of these cases provides convincing empirical justification for the introduction of this distinction.

As mentioned in section II.7.3, Chomsky (1995:354-355, 373-375) proposes that languages can be parametrized in terms of the number of times their strong features can escape erasure after being checked and deleted or, alternatively, in terms of the number of "unforced" violations of Procrastinate that can be tolerated. Under this approach, the parametric property which permits transitive expletive constructions is conceived of as a property of the strong feature of T which allows the strong feature to escape erasure after being checked and deleted. In the derivation of a transitive expletive construction, the subject checks the strong feature of T, which is only deleted, thereby being accessible to the computational system; the expletive then enters into an additional checking relation with the strong feature, which is finally erased.

However, it is not crucial that erasure be invoked in order for this analysis of transitive expletive constructions to work. Chomsky's approach can be easily recast in terms of the number of checking relations a strong
feature may participate in, after being checked. For instance, one could say that the relevant parametric property which permits transitive expletive constructions is that the strong feature of T may escape deletion after being checked. I am not claiming that the latter is the proper way to analyze transitive expletive constructions (see fn. 24); I am only claiming that Chomsky's (1995:354-355, 373-375) analysis of transitive expletive constructions does not provide compelling evidence for the distinction between erasure and deletion.

Let us now consider the raising of there in constructions such as (79) below. As will be discussed in section II.14.3.2, Chomsky (1995:287) argues that the expletive there has a D-feature, but no Case or $\phi$-features; in addition, he assumes (p. 287) that the categorial feature of expletives is [-interpretable]. As a [-interpretable] feature, we should expect that the D-feature of there should be deleted and erased after checking the strong feature of the embedded T in (79b). Were that the case, however, there would be unable to raise and check the strong feature of the matrix T, yielding an incorrect result.

(79)  
   a. There seems to be a cat on the mat.
   b. [ there, seems [ t, to be a cat on the mat ] ]

Chomsky (1995:281) prevents this situation from arising by relying on
the condition in (80) below, which is based on the assumption that the syntactic object $K' = \{U, \{\beta\}\}$ formed by erasing $\alpha$ in $K = \{U, \{\alpha, \beta\}\}$ is an illicit object. In instances such as (79), deletion of the D-feature therefore cannot be followed by erasure, because this extra step would create an illegitimate syntactic object, canceling the derivation. Since the D-feature of the expletive is only deleted after being checked, the expletive is still accessible to the computational system and can therefore raise to the matrix subject position and check the strong feature of $T$.

(80) A term cannot erase.

If transitive expletive constructions do not provide compelling evidence for the introduction of the erasure operation, as claimed above, expletive raising constructions remain the only instances where erasure may be relevant. However, it is an odd result from a conceptual point of view that the only construction which we may take as evidence for erasure is one in which this operation must not apply. It should also be noted that raising of $there$ in (79) may require the postulated distinction between deletion and erasure only if its categorial feature is [-interpretable]. Although this is certainly a reasonable assumption to make, there is not much to be made from this, given that a worked-out analysis of the interpretation of categorial features at the C-I interface is still to be provided in the Minimalist
framework.

In section II.14.2 below, I offer an alternative analysis of (79) which does not resort to erasure. Before getting into that, let us now consider the empirical problem that a sentence such as (81a), under the derivation represented in (81b), poses for Chomsky’s (1995:chap. 4) checking theory.

(81)  
   a. *It seems that was told John that he was fired.
   b. *[ it seems that [TP t was told John that he was fired ] ]

As will be discussed in section II.14.3.2, Chomsky (1995:287) argues that expletives such as it have D-, Case-, and \( \phi \)-features. Assuming this to be the case, the expletive is inserted in the embedded subject position in (81b) and enters into a checking relation with the embedded T. According to Chomsky’s (1995:chap. 4) checking theory, the Case- and \( \phi \)-features of the expletive, which are [-interpretable], should then be deleted and erased after being checked. The categorial feature of the expletive, on the other hand, is deleted but not erased; otherwise, the whole term would be erased, violating the condition in (80) and canceling the derivation.

Since its categorial feature is not erased after being checked in (81b), the expletive is still accessible to the computational system and can raise to the matrix subject position and check the strong feature of the matrix T. In turn, FF(John) can raise in the covert component and check its own Case-feature
and the Case- and $\phi$-features of $[\text{FF}(\text{seems})+T]$. Crucially, movement of $\text{FF}(\text{John})$ over the trace of the expletive does not yield a Minimal Link Condition violation, because the Case- and $\phi$-features of the trace of $\text{it}$, which could block this movement, were erased after being checked. The D-feature of the trace of the expletive in (81b) should be as transparent for the movement of $\text{FF}(\text{John})$ as the D-feature of the trace of $\text{there}$ in (79) for the movement of $\text{FF}(\text{a cat})$. The conclusion appears to be that the unacceptability of (81a) cannot be accounted for because erasure has applied.\footnote{A more technical problem arises in the case of overt erasure of $[-\text{interpretable}]$ features, if Morphology is to check the compatibility between formal and phonological features (see fn. 9). Chomsky (1995:chap. 4, fn. 50) raises the possibility of interpreting "overt erasure of F as meaning conversion of F to phonological properties, hence stripped away at Spell-Out".}

Let us consider another approach.

II.14.2. An Alternative Approach

The discussion in the previous section has shown that the erasure operation is conceptually dubious and empirically faulty. Let us then assume that deletion, understood as an operation which renders a given element invisible to the interface (see Chomsky (1995:280)), is sufficient for the purposes of checking theory, and that there is no operation of erasure. All that is then necessary to account for the data discussed by Chomsky (1995:chap. 4) is to assume that (i) categorial features are $[+\text{interpretable}]$,
without exceptions; and (ii) [-interpretable] features cannot enter into a checking relation once checked (that is, a deleted feature cannot participate in a checking relation).

In the absence of a precise theory of the interpretation of categorial features, the assumption that categorial features are always [+interpretable] is in principle neither a good or a bad move to make. However, taking the categorial feature of *there* to be [+interpretable] has two welcome consequences. First, the C-I system may rely on the D-feature of the expletives *there* and *it* in order to make the appropriate distinction between them. And second, checking theory will be simplified, since the categorial features of expletives will pattern like any categorial feature in terms of their capability of entering into multiple checking relations. The expletive *there* in (79), for example, can check the strong feature of the matrix T even after having checked the strong feature of the embedded T; the D-feature of *there* is (by assumption) [+interpretable] and, as such, remains unaffected by participating of a checking relation.\footnote{In section III.6.2.5, I propose a slight modification regarding the effects on the phonological component of checking [+interpretable] features. The analysis developed in this chapter will remain unaltered, though.}

In turn, the unacceptability of (81a) is accounted for straightforwardly. When *it* is inserted in the embedded subject position in (81b), its D-, Case- and φ-features enter into checking relations with the features of the embedded T. The D-feature of *it* is unaffected by the checking relation that it
takes part in, because it is (by assumption) [+interpretable]; on the other hand, since the Case and $\phi$-features of *it* are [-interpretable], they are deleted but remain present in the structure. As [+interpretable], the D-feature of the expletive is still accessible to the computational system; the expletive can then raise and check the strong feature of the matrix T. Although the Case- and $\phi$-features are pied-pied to the subject position when the expletive raises, they cannot enter into checking relations with the relevant features of the matrix T, because they have been deleted in a previous checking relation.

The Case- and $\phi$-features of *John*, by contrast, are still accessible to the computational system. They could in principle raise in the covert component and check the Case- and $\phi$-features of the matrix T. However, the trace of the expletive prevents this movement from taking place: although the Case- and $\phi$-features of the trace of the expletive have been deleted, they remain present in the structure and can induce a Minimal Link Condition violation. Given that the Case-feature of *John* and the Case- and $\phi$-features of the matrix T are not checked, the derivation crashes at LF; hence the unacceptability of (81a).

In section III.4.1.5.1, I will return to the relevance of sentences such as (81a) for an erasure-based implementation of checking theory. Right now, let us reconsider Chomsky's (1995:sec. 4.5.2) definition of Last Resort stated in (78) and repeated below in (82), which Chomsky (1995:280) takes to be a
defining property of Move (see fn. 44).

(82)  

*Last Resort:*

Move F raises F to target K only if F enters into a checking relation with a sublabel of K.

Under the Copy + Merge theory of movement (see section II.10), (82) should be interpreted as a condition on chain formation along the lines of (83), since neither Merge nor Copy is independently subject to this constraint (see section II.10.1).

(83)  

*Last Resort:*

The chain $\text{CH} = (\alpha, \beta)$ can be formed if (at least) one feature of $\alpha$ enters into a checking relation with a sublabel of K, where K is the head (of the projection) with which $\alpha$ merges.

The main difference between (82) and (83) regards accessibility for movement/copying. As mentioned above, F in (82) ranges over accessible features, i.e., [+interpretable] features, unchecked [-interpretable] features, and checked [-interpretable] features which have been deleted but not erased. Under (83), any set of formal features can be copied regardless of the interpretability of its members. As in Chomsky's (1995:chap. 4) system,
[+interpretable] features will always be able to participate in a checking relation, regardless of whether or not they have been checked (but see fn. 71); [-interpretable] features, on the other hand, will be taken to be able to enter into a checking relation only if unchecked. As pointed out above, this is just a different technical implementation of Chomsky's (1995:sec. 4.5.2) proposal concerning the correlation between interpretability and checking capability. This version of Last Resort is nevertheless considerably simpler, because it does not make use of the distinction between deletion and erasure, thereby avoiding the conceptual and empirical problems associated with the postulation of the erasure operation.

Another difference between the two approaches relates to the status of Last Resort violations. Since Chomsky (1995:257) takes Last Resort to be part of the definition of Move, the derivation is canceled if it is violated. Under the version in (83), violation of Last Resort prevents a chain from being formed, but there is no obvious reason for why failure to form a chain should cancel the derivation (see fn. 44). In sections III.6.2.5 and IV.2.2, I argue that lack of chain formation may cancel the derivation, because the structure sent to PF cannot be linearized. Since the issue is orthogonal to the purposes of this chapter, I will simply assume that if a potential chain cannot be formed because it violates (83), the derivation is canceled.

Assuming the version of Last Resort given in (83), the formulation of Form Chain stated in (60) should be revised as in (84) (see section III.6.2.5 for
further refinement).

\[(84) \quad \text{Form Chain:}
\]

The syntactic objects \(\alpha\) and \(\beta\) can form the nontrivial chain \(\text{CH} = (\alpha, \beta)\) only if:

(i) \(\alpha\) c-commands \(\beta\);

(ii) \(\alpha\) is nondistinct from \(\beta\);

(iii) (at least) one given feature \(F\) of \(\alpha\) enters into a checking relation with a sublabel of \(K\), where \(K\) is the head (of the projection) with which \(\alpha\) merges; and

(iv) there is no syntactic object \(\gamma\) such that \(\gamma\) has a feature \(F'\) which is of the same type as the feature \(F\) of \(\alpha\), and \(\gamma\) is closer to \(\alpha\) than \(\beta\) is.

Let us now examine the effects of Last Resort as restated in (83) under the reformulations of Chomsky’s (1995:chap. 4) checking theory proposed above, namely: (i) categorial features are always [+interpretable]; (ii) [-interpretable] features cannot enter into a checking relation once checked; (iii) there is no erasure operation.
II.14.3. Analysis of Some Data

The data to be discussed in sections II.14.3.1, II.14.3.2, and II.14.3.3 below were analyzed in Chomsky’s (1995:sec. 4.5) under the erasure-based version of Last Resort given in (82). I will recast Chomsky’s analysis in terms of the version of Last Resort given in (83), which does not resort to the distinction between deletion and erasure. The two analyses make the same empirical predictions with respect to the data discussed below. They however differ in accounting for constructions such as (81a), as discussed in section II.14.2, and for the constructions to be discussed in section III.4.1.5. For the sake of presentation, I will not contrast the two different versions of Last Resort in cases where they make the same predictions.

II.14.3.1. NP Raising

Consider the derivations in (85) under the approach proposed in section II.14.2:

(85)  a. I believe [\TP John_{t_i} to be likely [\TP t_i to [\VP t_i v [\VP win ] ] ] ]

b. *John_{t_i} seems that [\TP t_i [\VP t_i v [\VP works hard ] ] ]
Both embedded clauses of (85a) have to satisfy the EPP. Therefore John raises from the Spec of the light verb and merges with the T projection of the most embedded clause; its D-feature enters into a checking relation with the strong D-feature of T, which is rendered invisible at both LF and PF (see section II.6.2). Since the D-feature of John is [+interpretable], it remains unaffected by this checking operation; it can thus raise again and check the strong D-feature of the intermediate T, in compliance with Last Resort. Finally, FF(John) raises in the covert component and adjoins to the matrix verb (see section III.4.1.4.3 for discussion), allowing the Case-features of John and believe to be checked. Given that every [-interpretable] feature is checked (therefore, rendered invisible at the interface), the derivation satisfies Full Interpretation.

In (85b), raising of John to the embedded Spec of TP to satisfy the EPP also permits a checking relation between the Case-features of John and T. John then raises to the matrix subject position to check the strong D-feature of the matrix T. As in (85a), this checking relation is possible because the D-feature of John is [+interpretable] and, therefore, was not affected by the previous checking relation in the embedded clause. Also as in (85a), the Case-feature of John is in a checking configuration with the Case-feature of T; however, the Case-feature of John was checked in the previous checking relation it participated in, and [-interpretable] features cannot enter into another checking relation once checked. Hence, although the chain formed
by raising John from the embedded to the matrix subject position satisfies Last Resort, the derivation in (85b) crashes at LF because the Case-feature of the matrix T has not been checked, violating Full Interpretation.

II.14.3.2. Expletive Constructions

Consider now the expletive constructions in (86) and assume that the contrast in acceptability between (86a) and (86b), on the one hand, and (86c) and (86d), on the other, results from convergence or lack thereof:

(86) a. [ it seems that [TP [ many people ]i are [PP ti in the room ] ] ]
    
    b. [ therei seems [TP ti to be [PP a man in the room ] ] ]
    
    c. *[ therei seems [TP ti to be [PP many people in the room ] ] ]
    
    d. *[ there seem that [TP [ many people ]i are [PP ti in the room ] ] ]

Under the assumption that the derivation of (86a) converges, all of its [-interpretable] features must have been checked by LF. In the embedded clause, the NP many people moves from a post-copula position to check the strong feature of T. This movement also allows two other checking relations: one between the Case-features of many people and T and another one between the φ-features of many people and [ are+T ]. So far, all the relevant [-interpretable] features of the embedded clause have been checked.
As for the [-interpretable] features of the matrix clause, we have at least the Case-feature and the strong D-feature of T, and the \( \phi \)-features of \textit{seems}. The expletive \textit{it} must then have D-, Case-, and \( \phi \)-features, enabling it to check all the [-interpretable] features of the other categories and allow Full Interpretation to be satisfied at LF.

Now consider (86b). \textit{There} is inserted in the derivation as the specifier of the embedded TP (see section II.15 for discussion) and then moves to the matrix Spec of TP; hence it must have at least a ([+interpretable]) D-feature in order to check the EPP of both matrix and embedded clauses (see section II.6). The contrast between (86b) and (86c), on the other hand, shows that the agreement of the matrix verb is determined by the associate of the expletive. This indicates that \textit{there} lacks \( \phi \)-features and that the formal features of the associate in (86b) raise covertly and adjoin to the matrix T, allowing a checking relation to be established between the \( \phi \)-features of FF(\textit{a man}) and the \( \phi \)-features of [ FF(\textit{seems})+T ].

Notice that covert movement of the associate in (86b) also places its Case-feature in a checking configuration with the Case-feature of T. The question then is whether T has its Case-feature checked by \textit{there} or by the associate. Suppose that it is checked by \textit{there}, and consider the derivation in (86d). In (86d), the associate checks the Case- and the strong feature of the embedded T and the \( \phi \)-features of [ are+T ]. We have seen above that \textit{there}
lacks $\phi$-features; thus, the associate raises covertly and checks the $\phi$-features of $[\text{FF(seem)}+T]$. Note that this movement yields a licit chain because the $\phi$-features of the associate are [+interpretable] and have not been affected by their previous checking relation in the embedded clause. Since *there* can check the strong $D$-feature and, by hypothesis, the Case-feature of the matrix $T$, all [-interpretable] features should have been checked by LF in (86d) and the derivation should converge, contrary to fact.

Suppose, on the other hand, that *there* lacks a Case-feature, as well. The unacceptability of (86d) then follows from a violation of Full Interpretation. Notice that although the formal features of the associate can check the $\phi$-features of $[\text{FF(seem)}+T]$, they cannot check the Case-feature of $T$ because the Case-feature of the associate has already been checked in the embedded clause and Case is a [-interpretable] feature. If $T$ cannot have its Case-feature checked by *there* either, the derivation will crash at LF, which is a welcome result.

If (86d) is to be so analyzed, we must conclude (together with Chomsky (1993:32, 1995:sec. 4.5.3), but contra Belletti (1988), Lasnik (1993), Chomsky (1994:36), among others) that the copula *be* in (86a)-(86d) does not have a Case-feature. Otherwise, the associate would check its Case-feature against the Case-feature of the copula in (86b) and would not be able to check the Case-feature of the matrix $T$ (see section II.14.3.2.1 below for further
Summarizing, the data in (86) indicate that the expletive *there* has only a D-feature (which is [+interpretable]), as any other categorial feature; see section II.14.1), as opposed to the expletive *it*, which has D-, Case-, and ϕ-features (see Chomsky (1995:sec. 4.5.3)).

72 Chomsky (1995:307) notes that if *there* has only a D-feature, the formal features of its associate must move across the expletive in constructions such as (i) in order to check its Case-feature, apparently yielding a violation of the Head Movement Constraint. This possibility seems to disappear, however, if we adopt Chomsky's (1995:364) suggestion (extending a proposal by Longobardi (1994)) that *there* has an N-feature to be checked. If so, it is possible that FF(*a book*) in (i) first adjoins to *there*, checking its N-feature, and then raises to the matrix clause to check its own Case-feature. It remains to be determined whether or not the second movement operation pied-pipes the D-feature of *there*.

(i) I expect there to be a book on the shelf.

Incidentally, two other things should be noted with respect to Chomsky's (1995:364) proposal that "[l]ike D that takes a complement, expletive D has a strong [nominal-] feature, which attracts [N] — a residue of the earlier a[junction-to-]expletive analysis". First, if *there* had a strong nominal feature, it should trigger overt movement (see section II.6), contrary to fact. Second, if the N-feature of the associate adjoins to *there*, we should expect the other formal features of the associate to be pied-piped (see section II.12 and fn. 64). If that happens, however, a potential problem may arise with respect to Chomsky's (1995:275) account of the contrast in (ii) based on the assumption that neither structure in (iii) is an appropriate binding configuration (see section II.12). If FF(*many linguists*) adjoins to the expletive in (ii) to check its nominal feature, a binding configuration similar to the one in (iia) will be formed, and the sentence in (ii) should thus be acceptable, again contrary to fact.

(ii) a. They seem to each other to have been angry.
   b. *There seem to each other to have been many linguists given good job offers.

(iii) a. [τ [FF(each other)] [τ [FF(many linguists)] [; T] ] ]
   b. [τ [FF(many linguists)] [τ [FF(each other)] [; T] ] ]
II.14.3.2.1. Expletives and Auxiliary Selection

Chomsky (1995:288) observes that agreement with the associate correlates with the feature composition of the expletive: the relevant verb agrees with the associate if the expletive lacks Case- and \( \phi \)-features (e.g. English *there*, German *es*, and Italian *pro*), as illustrated in (87), but not if the expletive is fully specified (e.g. French *il*), as illustrated in (88). If the expletive does not have \( \phi \)-features, the \( \phi \)-features of the relevant verbal complex adjoined to T will have to be checked by the \( \phi \)-features of the associate.

(87) There are three children at the party.

(88) Il y a trois enfants à la soirée.

EXPL CLITIC have three children at the party

'There are three children at the party.'

At first sight, there appears to be a problem with this reasoning. If the expletive *il* in (88) is fully specified for categorial, Case-, and \( \phi \)-features, its Case-feature should enter into a checking relation with the Case-feature of T; if this is true, however, it seems that the associate cannot be able to check its
Case-feature and the derivation should crash, contrary to what we expect.

Rather than being a problem for Chomsky's (1995:288) correlation, the sentence in (88) may actually lead us to another correlation: languages which have a fully specified expletive in existential constructions (e.g. French *il*, Brazilian Portuguese *pro*, Spanish *pro*) employ an auxiliary verb other than a *be*-like copula (e.g. French *avoir*, Brazilian Portuguese *ter*, Spanish *haber*). Let us reconsider the English and the French existential constructions in (87) and (88), for instance.

In (87), the expletive *there* has just a D-feature; the Case- and $\phi$-features of the complex [ *are+T* ] are then checked in the covert component after $\text{FF}(*\text{three children}*)$ raises. In (88), on the other hand, the expletive *il* is like English expletive *it* in having D-, Case-, and $\phi$-features; *il* can then check all of the [-interpretable] features of the complex [ *a+T* ], making it impossible for the associate to check its Case-feature against T. If the derivation of (88) is to converge, it must be the case that the associate checks its Case-feature against an element other than T. Assuming that *have*-like auxiliaries differ from *be*-like auxiliaries in having a Case-feature (see Roberts (1987), Nunes (1990, 1993, 1994c), Freeze (1992), Kayne (1993), among others), the auxiliary *a* in (88) is able to check the Case-feature of the associate, allowing the derivation to converge.
II.14.3.3. Complementizers and Movement Operations

The analysis of wh-movement in terms of Last Resort offered in Chomsky (1995:280) is considerably different from previous analyses in terms of Greed (see Chomsky (1993:33, 1994:14)). Under a Greed approach, movement of a given category C is allowed only if some morphological requirement of C is satisfied. Wh-movement is thus triggered by the need of a wh-phrase to check its wh-feature. Whether wh-movement takes place overtly or covertly depends on whether or not the category against which the wh-phrase checks its wh-feature is strong or not. From this perspective, all languages look alike at LF as far as the position of wh-elements is concerned.

Under the approach based on Last Resort, on the other hand, movement is triggered in order to eliminate [-interpretable] features. The wh-feature of a wh-expression is interpretable at the C-I interface; therefore it cannot be eliminated or trigger a movement operation per se. Since the formal features of a complementizer which determine clause type (interrogative, declarative, etc.) also have some effect at the C-I interface, they are [+interpretable] as well, and need not be checked, unless they contain a strong feature. Languages will thus differ with respect to the

---

73 Chomsky (1995:289) takes clause type to be determined by formal features. It does not seem implausible, though, that the "force" of a clause is determined by the semantic features of its complementizer.
position of wh-features at LF. A language such as Chinese, for instance, whose (null) interrogative complementizer Q does not have a strong feature, will employ no wh-movement at all (overtly or covertly): since the relevant features of Q and the wh-expression are [+interpretable], they cannot be eliminated and therefore do not trigger movement (see Uriagereka (forthcoming: chap. 5 for discussion). In instances like this, Chomsky (1995:291) proposes that C-I must employ an interpretive strategy involving something like unselective binding of the wh-phrase in situ by an interrogative complementizer Q (see Baker (1970), for instance).74

The (null) interrogative complementizer Q of English, in contrast, has a strong wh-feature, therefore requiring the insertion of a feature FQ in its checking domain. In (89a) below this strong feature is checked by merging whether in the specifier of Q, whereas in (89b) it is checked by adjoining if to Q. Checking the strong feature of Q via movement can also obtain by both substitution and adjunction: in (90a) the wh-feature of which book raises to Spec of Q, while in (90b) the V-features of did in Infl adjoins to Q (in both cases, a whole category is pied-piped for the reasons discussed in section II.12).75

74 This interpretative strategy is different from the two QR-like interpretative strategies proposed in Chomsky (1993), which will be discussed in section III.2. See Nunes (1995a) for the proposal that the QR-like strategies should be subsumed under the strategy involving unselective binding by an interrogative complementizer.

75 Recall that strong features are supposed to be checked only by categorical features (see Chomsky (1995:232), and fn. 15 above). Hence, Chomsky (1995:289-290)
a. [I wonder [CP whether Q [he left yet]]]

b. [I wonder [CP [Q if Q] [he left yet]]]

a. [guess [CP [which book]] Q [the man gave Q to Mary]]

b. [CP [Q did, +Q] [John give a book to Mary]]

Under the Last Resort approach, the derivations of the sentences (91b) and (92b) below from the structures in (91a) and (92a) are both convergent. In (91a), Infl raises to Comp, checking the strong V-feature of Q; since the wh-feature of which book is [+interpretable], it need not (therefore does not) move to Spec of CP to be checked.\textsuperscript{76} In turn, (92a) has two Q complementizers with strong features, which are checked by the wh-feature of which book. As a [+interpretable] feature, a wh-feature is allowed to enter into more than one checking relation. Chomsky (1995:290-291) then ascribes

Takes Q to have a strong V-feature in (90b) and a strong wh-feature which would be a variant of a D-feature, in (89a), (89b) and (90a). Since Chomsky (1995:291 and chap. 4, fn. 64) abstracts away from the contrast between matrix and embedded clauses with respect to Infl-to-Comp movement, suggesting that this asymmetry might have its origins in the phonological component, the strong feature of Q appears to be satisfied by either a wh- or a V-feature, yielding a different reading in each case.

Incidentally, although it is reasonable that the wh-feature of a wh-phrase is a kind of D-feature, the came cannot be said with respect to if in (89b). A more plausible approach is to take if in instances such as (89b) to be an interrogative complementizer with no strong feature.

\textsuperscript{76} In a multiple question such as (i), the wh-phrase what moves overtly in order to check the strong wh-feature of the interrogative complementizer Q. No covert movement of the wh-phrase in situ to the Spec of CP is required because its wh-feature is [+interpretable]. Interesting questions concerning strong feature checking are raised by languages which involve fronting of more than one wh-phrase (see fn. 18).

(i) What did John give to whom?
the unacceptability of the sentences in (91b) and (92b) to some deviant interpretation that C-I ascribes to the structures in (91a) and (92b).

(91)  a. \([_{CP \ Q \text{ did} \ i \ CP} \ [ \text{John t} \ i \ \text{give which book to Mary}\ ]]\]

b. *Did John give which book to Mary?

(92)  a. Guess \([_{CP \ [ \text{which book}]} \ [ \text{Q [ they remember [_{CP \ t} \ Q \ [ \text{[John gave t} \ i \ \text{to Mary}]]]}]}]\]

b. *Guess which book they remember John gave to Mary.

Since Chomsky (1995:232) allows lexical access to elements without phonological features in the covert component (see section II.5), questions arise regarding the insertion of a Q complementizer with a strong feature after Spell-Out in structures such as (93). Were this option available, FF(\text{will}) in (93a) and FF(\text{who}) in (93b) would adjoin to Q in the covert component, checking its strong feature.

(93)  a. \([_{Q \ [ \text{IP Mary will visit John}]}]\]

b. \([_{Q \ [ \text{IP who visited John}]}]\]

Chomsky excludes this possibility for (93a) based on the fact that the variant of Q which is checked by a V-feature has phonological properties
that determine rising intonation. Thus, if this variant of Q is inserted covertly, its phonological features will cause the derivation to crash at LF. As for the possibility of inserting a strong Q covertly in (93b), which is taken to have no phonological properties, Chomsky rules it out based on the economy principle in (3), repeated below in (94).

\[(94) \quad \alpha \text{ enters the numeration only if it has an effect on output.}\]

As was discussed in section II.6.1, (94) suffers from serious conceptual and empirical problems. A uniform account of the impossibility of inserting a null Q with a strong feature covertly in structures such as (93a) and (93b) can however be obtained, if a strong feature is a hybrid formal-phonological feature, as suggested in section II.6.2. If the strong features of (93) are inserted covertly, their phonological part will cause the derivation to crash regardless of whether or not their formal part is checked (see section II.6.2 for discussion). Thus, in a convergent derivation with Q being inserted covertly in (93b), Q cannot have a strong feature and FF(who) moves neither overtly nor covertly.\(^{77}\) Its interpretation at the C-I interface must be obtained through the in-situ strategy alluded to above.\(^{78}\)

\(^{77}\) On whether weak Q is inserted overtly or covertly, see below.

\(^{78}\) The analysis of the impossibility of inserting a strong Q covertly in terms of the hybrid nature of strong features proposed in II.6.2 inherits the problems that exist in Chomsky’s (1995:sec. 4.5.4) analysis. For instance, one should ask whether Q in (93b) can have a strong feature and be inserted overtly. If possible, then the next question is
For the sake of completeness, consider now a matrix clause such as (95) below in English. Assuming that the "force" of a clause is determined by the nature of its complementizer (see fn. 73), (95) must have a null declarative complementizer C. Based on the discussion above (see also section II.6.2), we must conclude that this complementizer does not have a strong feature, because it does not trigger overt movement. The question then is whether it is inserted overtly or covertly. Chomsky (1995:292) proposes that a complementizer C with no strong feature be inserted covertly "on grounds of economy, if we assume that Procrastinate holds of Merge as well as Move".

(95) John left.

why movement of a subject wh-phrase does not trigger Infl-to-Comp movement (see fn. 74), as illustrated in (i) with unstressed *did. Also, one needs to explain why the in-situ interpretive strategy is available for subjects as in (93b) but does not seem to be available for objects, for instance, as illustrated in (ii) (without the intonation of focus of an echo-question).

(i) *Who did come?
(ii) *Mary gave what to John?

79 As Chomsky (1995:292) observes, the null complementizer that appears in matrix clauses is different in nature from the overt complementizer that in English: the former carries declarative force, whereas the latter does not. Thus, (ii) is an appropriate answer for the question in (ia), but not for the one in (ib):

(i) a. What did Mary say?
   b. What happened?

(ii) That John left.
If the effects of Procrastinate should be derived along the lines of the proposal in section II.10.2, where Copy and Spell-Out are compared in terms of derivational cost, Chomsky (1995:292) claim that Merge is subject to Procrastinate would amount to saying that Merge and Spell-Out should be compared for purposes of economy and that the former is more costly. However, there appears to be no principled reason for taking Merge to be inherently more costly than Spell-Out. In the absence of empirical evidence to the contrary, Merge and Spell-Out should be analyzed as equally economical, given that a pair \((\pi, \lambda)\) can only be formed with applications of these operations (see sections II.4, II.10.1, and II.10.2). Comparing Spell-Out with the sequence of derivational steps involving Select and Merge (i.e. lexical insertion) is not illuminating either, because Select is also derivationally costless (see section II.4). It is possible that this is an instance in which the grammar allows true optionality: if Merge and Spell-Out are equally economical, a matrix complementizer with no phonological or strong features can be inserted either before or after Spell-Out.

Another possibility, which is consistent with Chomsky's (1995:292) suggestion, is that the sequence of derivational steps involving Select and Merge (i.e. lexical insertion) can be rendered indirectly more costly than Spell-Out by requiring further computations in the phonological component. For instance, suppose that (i) later stages in the derivation can be taken into consideration for the choice between two equally economical
steps; (ii) null categories are also to be linearized in accordance with Kayne's (1994) Linear Correspondence Axiom (see sections II.17); and (iii) the number of elements to be linearized correlates with the cost of the linearization operation.

Consider now a given derivational step where a TP is formed and the only remaining element in the numeration is a complementizer with no phonological or strong features. If the computational system selects this lexical item and merges it with the assembled TP, it will be introducing more elements to be linearized in the phonological component than if it had applied Spell-Out. Under this view, spelling out objects without phonological features would require unnecessary extra work in the phonological component as far as linearization is concerned.

I will leave the choice between these options open.

II.14.4. Feature Mismatch

In an attempt to reduce computational complexity, Chomsky (1995:309) proposes that when a feature moves, if any of the pied-piped formal features mismatches a sublabel of the target, the derivation is canceled:
"Suppose that $f$ is the Case-assigning feature of $K$, $\alpha$ and $\beta$ have the unchecked Case features $F_{\alpha}$ and $F_{\beta}$ (respectively), and $F_{\alpha}$ but not $F_{\beta}$ matches $f$. Suppose that $\beta$ is closer to $K$ than $\alpha$. Does $\beta$ prevent $K$ from attracting $\alpha$? The Case feature $F_{\beta}$ of $\beta$ does not do so; it is not attracted by $K$, and is therefore no more relevant than some semantic feature of $\beta$. Suppose, however, that $\beta$ has some other feature $F'_{\beta}$ that can enter into a checking relation with a sublabel of $K$. Then $\beta$ is attracted by $K$, which cannot "see" the remote element $\alpha$. A mismatching relation is created, and the derivation is canceled: $\alpha$ cannot be attracted." (Chomsky (1995:310))

Below I examine some instances of potential feature mismatch or violations of the Minimal Link Condition under the system proposed in Chomsky (1995:chap. 4), leaving the discussion of some problems of this approach for section II.14.4.1. Let us start by considering the wh-movement in (96).

(96)  [ they wonder [CP [ which book ]], Q [TP the man gave t_i to whom ] ]

The interrogative complementizer $Q$ in (96) has a strong wh-feature, which is taken to be a type of D-feature (see fn. 75). Since the phrase the man, which also has a D-feature, intervenes between the moved wh-phrase and its trace, one wonders whether the chain ([ which book ]), $t_i$) satisfies the
Minimal Link Condition. It does, according to Chomsky’s reasoning cited above. The D-feature of the man is not the appropriate type of feature to check the strong feature of Q; hence, it could not enter into a checking relation with the strong feature of Q and does not yield a Minimal Link Condition violation. If Q had another feature that could be checked by the man, then this phrase would raise and its D-feature and the wh-feature of Q would mismatch, canceling the derivation. Since this is not the case, (96) converges and is assigned a sound interpretation by the C-I interface. 

Now consider the derivation of the sentence (97) in English, where a nominative object moves to Spec of T crossing an accusative subject in the Spec of vP, as represented in (98):

(97)  *He her saw.

'She saw him.'

---

Under this approach, the sentences resulting from the structures in (i) and (ii) are unacceptable for different reasons. (i) violates the Minimal Link Condition, because the wh-phrase which book has the appropriate feature to check the strong feature of Q' and is closer to to whom than ti; hence, the derivation is canceled. (ii), on the other hand, coverages, but presumably receives a deviant interpretation at the C-I interface (see Chomsky (1995:sec. 4.5.4.) and section II.14.33).

(i) [CP [to whom], did+Q' [TP they remember [CP [which book], Q [TP John gave t, ti]]]]

(ii) [CP [which book], did+Q' [TP they remember [CP t, Q [TP John gave t, to whom]]]]
As far as Case-features are concerned, movement of the object over the subject in (98) does not violate the Minimal Link Condition according to the discussion above, because the subject could not have its accusative Case-feature checked in Spec of TP. However, overt movement to the Spec of TP in English is triggered by the strong D-feature of T. Since the subject also has a D-feature and is closer to *he* than the trace of *he*, the chain (*he*, *ti*) violates the Minimal Link Condition; hence, the unacceptability of (97).

If, on the other hand, an accusative subject moves to Spec of TP to check the strong feature of T, as illustrated in (100), the Case-features of the subject and T will mismatch and the derivation will be canceled; hence the unacceptability of (99).
II.14.4.1. Conceptual and Empirical Problems with Feature Mismatch

Consider the existential construction involving a pronoun illustrated in (101).

(101)  There's Mary, there's Sue, and there's him/*he.

Putting aside the special conditions that allow a name or a pronoun to appear in an existential construction, what is relevant for our purposes is that a nominative pronoun is barred in constructions such as (101). This is rather
unexpected from a feature mismatch approach. After all, the Case-features of the T head of the last conjunct and the pronoun are presumably checked against each other in (101) after the formal features of the pronoun attach to T at LF (see section II.14.3.2).

Suppose that nominal elements (including pronouns) in English are underspecified with respect to the type of Case they bear, and that a default morphological rule realizes pronouns with unchecked Case-features as accusative. If so, the pronoun in (101) must be realized as accusative and not nominative, because it has not been checked. If a pronoun overtly checks its Case-feature against a finite T head or a possessive determiner (see Abney (1987)), it will be realized as nominative or genitive, respectively. Since this difference between nominative and accusative pronouns in English is established in the phonological component, no problem will arise when the formal features of him raise to check its Case-feature against the relevant T in (101): in the mapping from N to λ, the Case-feature of the pronoun behaves like the Case-feature of any standard associate.81

Evidence for this default realization of accusative Case in English is provided by answers to questions involving a wh-phrase in subject position,
as exemplified in (102), by coordinate NPs in subject position as in (103), and by topicalization constructions such as (104):

(102)  
A: — Who left?  
B: — him/*he

(103) Me and him went to the movies.

(104) Him/*he, I like his poems.

If English nominals are underspecified for the type of Case they bear, as the evidence above indicates, the unacceptability of (99), repeated below in (105a), cannot be due to feature mismatch. According to the morphological realization rules discussed above, the pronouns in (105b) should be realized as she and him. Thus, the unacceptability of (105a) is due to the illicit instances of morphological realization.

(105)  
a. *Her saw he.  
'She saw him.'

b. [TP [TP heri [VP ti saw he]]]

---

82 Thanks to Ellen Thompson (p.c.) and Juan Uriagereka (p.c.), Who brought the relevance of constructions such as (103) and (104) to my attention.
In turn, the unacceptability of (97), repeated below in (106a), is due to the fact that both the D-feature and the Case-feature of her in (106b) induce a Minimal Link Condition violation; since the Case-feature of the pronoun in Spec of vP is underspecified, it can enter into a checking relation with T, thus preventing movement of the object pronoun.

(106) a. *He her saw.
   'She saw him.'
   
   b. [TP hei [vP her saw ti ] ]

The English data discussed above, therefore, cannot be used as empirical evidence for the proposal that feature mismatch cancels the derivation. Notice furthermore that Chomsky's (1995:310) account of (106) in terms of a Minimal Link Condition induced by the D-feature of her makes the unlikely prediction that in languages where the EPP does not hold and both the subject and the object remain in situ, the sentences corresponding to the English glosses in (107a) and (107b) below should be synonymous. Recall that the only reason why the derivation in (106b) is not possible under Chomsky's analysis is that the strong D-feature of T must be checked by the accusative subject, yielding a Case mismatch. If T has no feature other than Case to check, movement of the formal features of the object across the subject in (107b) should be parallel to the wh-movement of which book across
the man in (96) and should not violate the Minimal Link Condition.

(107)  
   a. he saw her  
   b. him saw she  
      'He saw her.'

A similar problem would arise in languages in which light verbs have a strong D-feature but T heads do not. An accusative subject could move to the outer Spec of the light verb, and the formal features of an object with nominative Case could adjoin to T to establish a Case-checking relation, without yielding a violation of the Minimal Link Condition. In this scenario, the sentences corresponding to the English glosses in (108a) and (108b) should also be synonymous, which is unlikely to be true.

(108)  
   a. her he saw  
   b. him saw she  
      'He saw her.'

It may be the case that the problems related to (107) and (108) do not actually arise, because T heads (for some reason) universally have a strong feature (see Jonas and Bobaljik (1993:74) and Chomsky (1995:chap 4, fn. 80)). However, if the strong feature of T is a parametric option, in order for us to
prevent the situation in which the pairs in (107) and (108) receive the same interpretation, the Case-feature of the subject must block adjunction of the formal features of the object to T for Case-checking, even if it does not establish a successful checking relation with the Case-feature of T. In other words, the specific value of a Case-feature is irrelevant for the computation of the Minimal Link Condition. That something along these lines may be on the right track is independently argued for by Uriagereka (1995b), based on conceptual grounds and some empirical evidence concerning the irrelevance of Case in obviation effects.

If it turns out to be true that a Case-feature prevents a movement operation for Case-checking reasons regardless of its value, the assumption that feature mismatch cancels the derivation should be maintained only on the conceptual grounds that it reduces computational complexity (see Chomsky (1995:309)). However, this assumption has the undesirable consequence of requiring the stipulation of [-interpretable] features with no PF reflex in some instances, only to prevent feature mismatch. Consider the derivation of transitive sentences in languages such as German, with overt object movement and subject agreement, for instance. Under the assumption that movement proceeds cyclically, after the object moves to the outer Spec of vP to check the strong D-feature of the light verb, we have the structure in (109):
The configuration in (109), in addition to allowing the strong D-feature of $v$ to be checked, also permits two other checking relations: (i) between the Case-features of OB and $[ V+v ]$; and (ii) between the $\phi$-features of OB and $[ V+v ]$. Although this is a welcome result with respect to Case-feature checking, problems arise regarding $\phi$-feature checking. If the checking relation between the $\phi$-features of the object and the $\phi$-features of the verbal complex were successful, the verb should agree with the object, yielding an incorrect result for the languages under consideration. If the $\phi$-features of the verbal complex are "Agrs-features" (i.e., agreement features associated with nominative or ergative Case), the derivation should be canceled because these features and the "Agro-features" of the object (i.e., agreement features associated with accusative or absolutive Case) mismatch; if that were the case, however, no language should have overt object movement.83

---

83 Similar considerations may apply to covert object movement for Case reasons in a language without overt verb movement to T. If in languages like English, the formal features of the object adjoin to the complex verbal head formed by adjunction of the main to the light verb before the formal features of this complex head adjoin to T (see section III.4.1.2.3), the $\phi$-features of the object the $\phi$-features of the verbal complex should either establish a successful checking relation, in which case English should exhibit object agreement, or mismatch, in which case English sentences with transitive verbs could not be derived. This particular problem does not arise in Chomsky's (1995:sec. 4.10.2) system because he assumes that in English-type languages the formal features of the verbal complex always raise to T before the formal features of the object also adjoin to T; hence the $\phi$-features of the verbal complex are erased after being checked against the subject, and no issue of feature mismatch arises. For some
This problem would not arise if we assumed the system in Chomsky (1993), according to which transitive verbs always have a set of Agro-features, that is, a set of φ-features associated with accusative Case. If so, the φ-features of the DP in the Spec of vP in (109) would enter into a checking relation with the Agro-features of the verbal complex, allowing the derivation to converge. This solution is dubious, however. To postulate [-interpretable] φ-features which have no reflex at PF and are motivated only by theory internal reasons is comparable to the postulation of an Agr projection.

Suppose, on the other hand, that we drop that assumption that feature mismatch cancels the derivation. The derivation of transitive constructions in languages with subject agreement and overt object shift or in languages with subject agreement and verbs and objects in situ (see fn. 83) can then be accounted for, if we take the specific choice of a set of φ-features to be somehow associated with a particular type of Case realization (which in the discussion above was referred to as Agrs- and Agro-features). If so, the Agro-features of an accusative object will not be able to enter into a checking relation with the Agrs-features of the verbal complex; thus, neither German nor English will exhibit object agreement, and no postulation of Agro-features for these languages will be required.

problems regarding this assumption, see section III.4.1.2.3 below.
To summarize, the empirical motivation for feature mismatch is not compelling, if we assume that the value of a Case-feature is irrelevant for the computation of the Minimal Link Condition.\textsuperscript{84} Furthermore, the assumptions regarding feature mismatch require that the theory be enriched with [-interpretable] features which have no role other than preventing feature mismatch. In this dissertation, I will therefore not assume that feature mismatch cancels the derivation.

\textbf{II.15. Computing Economy}

As mentioned in section II.2, the set of admissible derivations $A$ is the subset of the set of convergent derivations $C$ which satisfy economy criteria. One such economy criterion is the Procrastinate principle of Chomsky (1993:30), which states that covert movement is less costly than overt movement (see section II.10.2 for a deduction of Procrastinate). In this section, I review Chomsky's (1994, 1995:chap. 4) proposal for how economy criteria are to be computed.

Consider the structures in (110), for instance, which both converge. (110b) violates Procrastinate, where (110a) does not. Were the computation

\textsuperscript{84} However, I will have nothing to say about why Case-features should differ from categorical and $\phi$-features regarding the Minimal Link Condition. This is perhaps related to the fact that both the moved element and the head of the target have a [-interpretable] feature in a checking relation involving Case, whereas only the head of the target has a [-interpretable] feature in a checking relation involving categorical or $\phi$-features.
of economy just based on the number of violations of economy principles, (110a) should rule out (110b), yielding a wrong result.

(110) a. [ there is [ someone in the room ] ]

b. [ someone i is [ t i in the room ] ]

Chomsky (1994:7, 1995:227) proposes that the reference set to compare derivations for purposes of economy should be based on the same initial numeration. If so, the two sentences in (110) are not comparable and the fact that (110b) violates Procrastinate is irrelevant; if someone does not raise in (110b) before Spell-Out, the derivation will crash at PF because the phonological component will not be able to eliminate the formal part of the strong feature of T (see section II.6.2).85

Now consider the structures in (111) below. In (111a), there is inserted in the embedded subject position and moves to the matrix subject position, checking the strong feature of both T heads (see section II.14.3.2); in (111b), on the other hand, someone moves to the embedded subject position, checking the strong feature of the embedded T, and there is inserted in the matrix subject position, checking the strong feature of the matrix T. Both constructions are based on the same initial numeration and involve a single

85 Interesting questions arise with respect to do-support in English, if only derivations which share the same numeration can be compared. See Arnold (1994, 1995) for a proposal that the relevant comparison is in terms of lexical insertion at Spell-Out.
violation of Procrastinate; hence they should pattern alike, contrary to fact.

(111)  a. [ there\(_1\) seems \([\text{TP} t_i]\) to be someone in the room ]

   b. *[ there seems \([\text{TP} \text{ someone}_i]\) to be \(t_i\) in the room ]

Chomsky (1994, 1995: chap. 4) proposes that not only should competing derivations be based on the same initial numeration, but they should also be evaluated at every derivational step:

Suppose that the derivation \(D\) with the initial numeration \(N\) has reached stage \(\Sigma\). The reference set within which relative economy is evaluated is determined by \((N, \Sigma)\): it is the set \(R(N, \Sigma)\) of convergent extensions of the derivation \(N \rightarrow \Sigma\), using what remains of \(N\). At \(\Sigma\), the operation \(\text{OP}\) is blocked if \(\text{OP}'\) yields a more economical derivation in \(R(N, \Sigma)\). (Chomsky (1995:297))

According to this proposal, the reference set at the point where (111a) and (111b) diverge is comprised of the numeration \(N\) in (112a) and the structure \(\Sigma\) in (112b):

(112)  a. \(N = \{\text{there}_1, \text{T}_1, \text{seems}_1, \text{to}_0, \text{be}_0, \text{someone}_0, \text{in}_0, \text{the}_0, \text{room}_0\}\)

   b. \(\Sigma = [\text{TP} \text{ to be someone in the room }]\)
There are two options available at this point which could yield convergent derivations: either *there* is selected from N and merges with $\Sigma$, eventually deriving (111a); or *someone* is copied and merges with T, eventually deriving (111b). As seen in section II.10.1, the sequence of derivational steps involving Copy and Merge (i.e. movement) is more costly than the combination of Select and Merge (i.e. lexical insertion), because Copy is associated with convergence requirements, whereas Select is a defining property of a derivation. The derivation involving the insertion of *there* is therefore more economical than the one raising *someone*; hence (111a) blocks (111b).\(^{86,87}\)

\(^{86}\) The pair of sentences in (i), which were pointed out to me by Juan Uriagereka (p.c.), Who in turn credited Juam Romero for the observation, appears to posit a problem for this analysis. Assuming that both sentences involve the same initial numeration, at the point where the constituent *is a man in the room* is assembled, insertion of *there* should block movement of *a man*, wrongly entailing that the sentence in (1B) should be unacceptable.

(i)  
\(\begin{align*}
    a. & \text{ The fact is that there is a man in the room.} \\
    b. & \text{ There is the fact that a man is in the room.}
\end{align*}\)

The problem Will not arise, however, IF the assumption that the sentences of (i) share the same initial numeration is wrong. It could be the case, for instance, that the matrix equative *be* in (1a) has different features than the matrix existentical *be* in (1b). That something along these lines is correct is shown by the contrast in (ii), brought to my attention by Norbert Hornstein (p.c.).

(ii)  
\(\begin{align*}
    a. & \text{ In Otello, a soprano is always Desdemona.} \\
    b. & \text{ *In Otello, there is a soprano always Desdemona.}
\end{align*}\)

\(^{87}\) Notice that under the Copy + Merge theory of movement, the same results are obtained IF instead of merging *there* with $\Sigma$ after selecting it from N in (1127), the derivation proceeds to select *seems* and T, yielding the numeration N' in (1a) and the set of syntactic objects in (1b). In (i), it is still less costly to merge *there* with $\Sigma$ than to make a copy of *someone* and merge it with $\Sigma$. 
Incidentally, the analysis of (111a) in terms of raising of *there* appears to be inconsistent with Chomsky’s (1995:366) account of why a pure expletive cannot be inserted in the outer Spec of the light verb in languages which allow overt object shift. If that were possible, the expletive would check the strong feature of the light verb and then move to Spec of TP to check the strong feature of T, yielding a transitive expletive construction as illustrated in (113).  

(113) \[ \text{TP Expl}_I T [\text{VP} \text{t}_I [\text{v} \text{SU} \text{v} [\text{VP} \text{V OB } ] ] ] \]

Chomsky (1995:366) rules out (113) by resorting to the economy principle in (94), repeated below in (114). The strong feature of the light verb in (113) has no effect on the LF output because it is [-interpretable], and no effect on the PF output, because the PF sequence resulting from inserting the expletive directly in Spec of TP would be the same. Hence, according to (114), the PF output expletive-verb-subject-object cannot be associated with a numeration in which the light verb has a strong feature.

---

(i) a. \( N' = \{\text{there}, \text{T}, \text{seems}, \text{too}, \text{be}, \text{someone}, \text{in}, \text{the}, \text{room}\} \)

b. \( \Sigma = \{\text{TP to be someone in the room}\} \)

K = there

L = T

M = seems

---

88 See Chomsky (1995:sec. 4.10.2 and 4.10.3) for a discussion of economy issues in languages with transitive expletive constructions.
(114) \( \alpha \) enters the numeration only if it has an effect on output.

This account of the impossibility of (113) entails that (114) would allow a strong feature to be assigned to the embedded T of (111b), repeated below in (115b), when its numeration was formed, but not to the embedded T of (111a), repeated below in (115a). Whereas the strong feature does have an effect at PF in (115b), it does not in (115a): the sequence associated with (115a) would also be obtained if the expletive were inserted in the matrix subject position.\(^{89}\) If (115a) and (115b) differ in whether or not their embedded T has a strong feature, their numerations are actually different, which make them unsuitable for comparison for economy purposes. Thus, the unacceptability of (115b) is not accounted for.

(115)  
\begin{align*}
\text{a. } & \text{[ there seems } [_{TP} t_i \text{ to be someone in the room }] \\
\text{b. } & \text{*[ there seems } [_{TP} \text{ someone}_i \text{ to be } t_i \text{ in the room }] ] \\
\end{align*}

Given that the economy principle in (114) is problematic in several other respects, as discussed in section II.6.1, we may dispense with it entirely and keep Chomsky's (1994) analysis of the contrast in (115) in terms of economy. A principled account of the ungrammaticality of the structure in

\(^{89}\) The strong feature in (115b) would also have an effect at LF, given that movement of a DP is able to license anaphor binding, whereas movement of FF(DP) is not (see section II.12).
(113), as well as many other properties of transitive expletive constructions remains to be developed, however.

Let us now consider the contrast in (116) below. In (116a), someone moves from its postcopular position to check the strong feature of the embedded T and John moves from the Spec of vP to the matrix subject position, where it checks the strong feature of the matrix T. In (116b), John is inserted in the embedded subject position, checking the strong feature of the infinitival T, and raises to the matrix subject position, checking the strong feature of the matrix T head. Given the analysis of (115), if the two derivations in (116) were convergent, we should then expect (116b) to outrank (116a), which does not correspond to the facts. Chomsky (1994:39, 1995:347) solves this puzzle by proposing that (116b) actually crashes due to a θ-Criterion violation. If so, (116b) cannot block (116a), yielding the correct results.90

90 Suppose that at LF, FF(someone) in (116b) adjoins to the matrix verb to check its Case-feature. According to the discussion in section II.14.4, the chain (FF(someone), t) violates the Minimal Link Condition because the trace of John, which also has a Case feature, is closer to FF(someone) than the trace of FF(someone). In Chomsky’s (1995:chap. 4) system, raising of FF(someone) over the trace of John in (116b) does not induce a Minimal Link Condition violation because traces of A-movement are taken to be unable to block movement. As Will be discussed in section III.4.3, however, the assumption that traces of A-movement do not induce violations of the Minimal Link Condition has undesirable empirical consequences.
Although I will follow Chomsky (1995:347) in taking the θ-Criterion to be a convergence condition and taking (116b) to involve a θ-Criterion violation (in addition to a Minimal Link Condition violation; see fn. 90), I will depart from him with respect to the specific interpretation assigned to the θ-Criterion, as discussed in the next section.

II.16. The θ-Criterion as a Condition on θ-role Assignment

The θ-Criterion violation in (116b) is subject to three interpretations: (i) the argument chain (John, t₁) does not receive a θ-role; (ii) the light verb fails to assign the external θ-role; and (iii) both possibilities in (i) and (ii). Minimalist considerations should eliminate the third possibility, given its overdetermination. Between the options in (i) and (ii), Chomsky (1995:347) takes the contrast in (116) to indicate that an argument without a θ-role violates Full Interpretation at LF, leaving open the status of failure to assign a θ-role.

I will instead interpret the θ-Criterion as a condition on θ-role
assignment for three reasons.\footnote{I am thus returning to the version of the $\theta$-Criterion assumed in Chomsky (1993:fn. 23).} First, it is an unavoidable property of the system that $\theta$-roles are inherently associated with the heads that assign them and not with the expressions which can bear them.\footnote{As pointed out by Jaeggli (1986:fn. 4), \textit{[t]he term argument is used ambiguously in Chomsky (1981). In one sense, (...) it is understood as a relational or functional notion, like direct-object-of or subject-of (...)}. In the other, it is used in an absolute sense, and it is closer in meaning to 'referential expression' as in the original Pisa lectures. The sense intended in the statement of the \(\theta\)-Criterion in Chomsky (1981) is clearly the latter.} Thus, stating the $\theta$-Criterion in terms of $\theta$-role assignment is conceptually more natural than stating it in terms $\theta$-role bearing: while the former resorts to properties which are independently specified in the lexicon, the latter introduces new assumptions in the system.

Second, if the $\theta$-Criterion is a condition on $\theta$-role assignment, we have an account of the unacceptability of the sentence in (117a) under the derivation of (117b):

\begin{align*}
(117) & \quad \text{a. It kissed Mary.} \\
& \quad \text{b. } [\text{TP } T_{\text{exp}} \text{ } T [\text{VP kissed Mary }]]
\end{align*}

In (117b), the expletive \textit{it} is inserted in Spec of TP, checking the Case- and the strong feature of T, and the $\phi$-features of \textit{kissed} after it adjoins to T at LF (see...
section II.14.3.2). After FF(\textit{Mary}) raises at LF and checks its Case-feature against the verbal complex, every [-interactable] feature will have been checked. If the 0-Criterion is to be understood as a condition on argument chains, the derivation of (117a) partially sketched in (117b) should then converge and receive an interpretation along the lines of the interpretation of a passive sentence such as (118). This clearly is an undesirable result.

(118) Mary was kissed.

Notice that given that the expletive of (117b) is not an argument (see fn. 93), it vacuously satisfies the version of the 0-Criterion according to which argument chains must receive a 0-role. Moreover, we cannot rule out this derivation based on an incompatibility between expletives and 0-roles (presumably, 0-roles can only be assigned to elements which have semantic features). Since the expletive in (117b) is not inserted in the Spec of vP, no incompatibility issue arises.

The unacceptability of (117a) under the derivation of (117b) should therefore be traced to the fact that the light verb does not assign the external 0-role. The question now is whether unacceptability in this case follows from a crashed derivation, or from a convergent derivation with a deviant interpretation. Notice that in the case of the derivation sketched in (116b),
the answer is clear: the unacceptability of the corresponding sentence must be due to lack of convergence; as discussed in section II.15, if (116b) converged, it would rule out (116a), yielding an incorrect result. Therefore, if we take the unacceptability of (117a) under the relevant reading to also follow from lack of convergence, we can account for both (116b) and (117b) with the same condition. If on the other hand, we attribute the unacceptability of the sentence resulting from (117b) to the interpretation it receives at the C-I interface, (116b) and (117b) will require two distinct explanations. In the absence of a theory of how θ-roles are interpreted at the C-I interface, methodological considerations lead us to assume a uniform account of (116b) and (117b) and take the θ-Criterion to be a condition on θ-role assignment.

Assuming this to be true, let us consider the final contrast in (119).

(119) a. *Johni [vp t' expected [ t' to be someone in the room ] ] ]}

b. Someonei was expected [ t' to be t in the room ] ]

(119a) John is inserted in the embedded subject position, satisfying the EPP, and then raises to Spec of vp. Recall that θ-role is not a formal feature and that feature checking is in complementary distribution with θ-role assignment (see section II.7.5). Thus, if the light verb assigns the external θ-
role to John in its Spec, satisfying the $\theta$-Criterion, no checking relation is established between these two elements. The chain $\text{CH} = (\text{John}, t_i)$ then violates Last Resort and the derivation is canceled (see fn. 44).

In (119b), by contrast, the two instances of movement allow the moved element to enter into a checking relation, thus satisfying Last Resort. The question of how the light verb assigns its external $\theta$-role in compliance with the $\theta$-Criterion now reduces to the question of how the apparent elimination of the external $\theta$-role in passive constructions in general is to be handled in a Minimalist framework.\(^{93}\)

### II.17. Linear Order

It has been standardly assumed within the framework of the Principles and Parameters Theory that a hierarchical structure may be associated with more than one linear order.\(^{94}\) A syntactic object such as $K = \{D, \{D, E\}\}$, for instance, was taken to surface with $D$ preceding or following $E$. Kayne (1994)

\(^{93}\) For an analysis of passives within the Minimalist framework, see Nunes (1994b).

\(^{94}\) A strict linear order is a relation with the properties listed in (i) (see Partee, ter Meulen and Wall 1990:sec. 3.5).

(i) Properties of a Linear Order:
   a. transitivity: if $xRy$ and $yRz$, then $xRz$;
   b. asymmetry: if $xRy$, then it must not be the case that $yRx$;
   c. irreflexivity: if $xRy$, then $x$ is different from $y$; and
   d. totality: for any $x$ and $y$ in a given set, $xRy$ or $yRx$.  

radically breaks with this tradition, by proposing a more restrictive theory of grammar in which the left-to-right linearization of the terminal symbols of a phrase-marker is determined by the notion of asymmetric c-command, as defined in (120) and (121) (see Kayne (1994:4, 18)).

(120)  *Asymmetric C-command:*

\[ \alpha \text{ asymmetrically c-commands } \beta \iff \alpha \text{ c-commands } \beta \text{ and } \beta \text{ does not c-command } \alpha. \]

(121)  *C-command:*

\[ \alpha \text{ c-commands } \beta \iff \alpha \text{ and } \beta \text{ are categories and } \alpha \text{ excludes } \beta \text{ and every category that dominates } \alpha \text{ dominates } \beta. \]

The gist of Kayne's theory involves two related but distinct proposals, as summarized as in (122) below. The proposal in (122a) is intended to derive the major properties of standard X'-Theory, whereas the one in (122b) is intended to provide the basis for a more restrictive theory of word order. In this section I will focus on Kayne's proposal in (122b), which will be crucial in chapters III and IV.\(^{95}\)

---

\(^{95}\) See Nunes (1995a) for a discussion of the relevance of (122a) for syntactic objects at LF.
a. A subset of the asymmetric c-command relations among the nonterminal nodes of a phrase-marker should form a linear order; and

b. Such a linear order should be mapped onto another linear order, this time involving terminal nodes in a precedence relation.

The mapping from a phrase-marker onto a linear order of terminal symbols is referred to as the *Linear Correspondence Axiom* (LCA), as defined in (123) (from Kayne (1994:33)):

\[(123) \quad \text{Linear Correspondence Axiom (LCA):} \]

\[\text{Let } X, Y \text{ be nonterminals and } x, y \text{ terminals such that } X \text{ dominates } x \text{ and } Y \text{ dominates } y. \text{ Then if } X \text{ asymmetrically c-commands } Y, x \text{ precedes } y.\]

Consider the linear order that the LCA determines for the terminal symbols of the phrase-marker in (124) below, where capital and lowercase letters stand for nonterminal and terminal symbols, respectively.
Since C asymmetrically c-commands E, the LCA establishes that c must precede e; in turn, since F asymmetrically c-commands H, I and J, f must precede h and j; by the same reasoning, h must precede j because H asymmetrically c-commands J. Given that D, G and I in (124) correspond to the complements of the heads C, F and H, respectively, Kayne (1994:35) concludes that heads universally precede their complements; languages exhibiting the order complement-head must involve movement of the complement to a position from which it asymmetrically c-commands the head.

Now consider the order between B and A₂ in (124) under the assumption that A₁ and A₂ are different categories. Since B asymmetrically c-commands F, G, H, I and J, the terminals c and e should precede f, h and j; on the other hand, given that A₂ asymmetrically c-commands C, D and E, the terminals f, h and j should precede c and e. The conjunction of these two
precedence relations however constitutes a violation of the asymmetry condition on a linear order (if $\alpha$ precedes $\beta$, then it must be the case that $\beta$ does not precede $\alpha$; see (ic) in fn. 94). This amounts to saying that if $A_1$ and $A_2$ are distinct categories, no linear order can be established between the head $F$ and its specifier $B$, for instance.

In order to prevent this undesirable result, Kayne (1994:16) proposes that specifiers are actually adjuncts. Given the definition of c-command in terms of categories and exclusion in (121), adjuncts will then be able to c-command the elements they are adjoined to, but not vice versa. Thus, if $A_1$ and $A_2$ in (124) are segments of the two-segment category $[A_1, A_2]$, $B$ will asymmetrically c-command $[A_1, A_2]$, $F$, $G$, $H$, $I$ and $J$, but neither $A_2$ nor $[A_1, A_2]$ will be able to c-command $C$, $D$ or $F$; hence the terminals $c$ and $e$ precede the terminals $f$, $h$, and $j$. To put it in general terms, adjuncts/specifiers precede the categories they are adjoined to.

Combining the two results discussed above, Kayne (1994:47) proposes that there cannot be any directionality parameter, because UG imposes the order specifier-head-complement. Word order variation must then be understood in terms of different combinations of movement.

Chomsky (1994:25) incorporates the LCA into the Minimalist framework, reformulating it in terms of bare $X'$-Theory. As discussed in section II.8), the bare $X'$-Theory system keeps the traditional distinction between specifiers and adjuncts and makes no distinction between terminals
and nonterminals. Thus, the structure in (124) reduces to (125), in a bare X'-Theory representation.

\[(125)\]
\[
A_1 = \{f, \{B, A_2\}\}
\]
\[
B = \{c, \{c, e\}\}
\]
\[
A_2 = \{f, \{f, G\}\}
\]
\[
f \quad G = \{h, \{h, j\}\},
\]

As in (124), if B and A_2 in (125) are allowed to enter into a mutual c-command relation, no order can be established between the constituents of B and the constituents of A_2. In order to prevent mutual c-command between a specifier and its sister, Chomsky (1994:27) proposes that intermediate projections do not enter into c-command relations (see section II.13.2). Hence, B asymmetrically c-commands the constituents of A_2 in (125), and A_2 does not enter into c-command relations because it is an intermediate projection, which yields the linear order Spec-head.

Under these assumptions, the LCA can be stated as in (126) (see Uriagereka (forthcoming: chap. 3) for discussion), based on the notion of c-command in (127) (from Chomsky (1994:30)), where intermediate projections are not accessible to the computational system:
(126) *Linear Correspondence Axiom (LCA):*

A category α precedes a category β iff:

(i) α asymmetrically c-commands β; or

(ii) γ precedes β, and γ dominates α.

(127) *C-command:*

Where α and β are accessible to the computational system, α c-commands β iff:

(i) α does not dominate β;

(ii) α ≠ β; and

(iii) every category dominating α dominates β.

Another difference between the system proposed by Kayne (1994) and its reformulation in Chomsky (1994) regards the order between heads and complements. When a complement has a complex internal structure such as G in (124) and (125), both systems predict that the head precedes its complements because the head asymmetrically c-commands the constituents of its complement. These systems nonetheless differ in the treatment of nonbranching complements such as D in (124) and e in (125).

For Kayne, the vacuous projection E in (124) is required to ensure that a
linear order is specified for the terminals $c$ and $e$; lack of $E$ (or $J$) would rule out the phrase-marker in (124), because no linear order among some of its terminals could be established, violating the totality condition on a linear order (if $\alpha$ and $\beta$ are in a linear order, either $\alpha$ precedes $\beta$ or $\beta$ precedes $\alpha$; see (id) in fn. 94).

Given that vacuous projections are banned in the bare $X'$-Theory system, Chomsky (1994:28) embraces the conclusion that no linear order can be established between a head and a complement without structure, proposing two possible ways to circumvent this problem: (i) Morphology converts two heads in mutual c-command relation into a "phonological word", and the LCA is taken to apply after Morphology and not have access to the internal structure of words (contra Kayne (1994:sec.4.5)); and (ii) if the complement has moved, "there is no reason for LCA to order an element that will disappear at PF, for example, a trace" (Chomsky (1994:28)). Here I will assume the proposal in (i), leaving discussion of (ii) for section III.4.2.97

96 A problem for Chomsky's (1994) adaptation of Kayne's (1994) theory appears to arise with respect to the linear order between the subject of a given clause and an adverb preceding an auxiliary, such as in (i), for instance. Chomsky (1995:235) suggests that in sentences such as (i), the adverb $\text{probably}$ merges with TP, forming $K = \{<T, T>, \{\text{probably}, \text{TP}\}\}$, which then merges with $he$, forming $L = \{T, \{he, K\}\}$. If c-command is to be stated in terms of categories, as in (127), He adverb and the subject in (i) should then be in a mutual c-command relation and no order between them could be established in accordance with the LCA. Notice in addition that since neither element is a sister of a head, they cannot be reanalyzed by Morphology as part of a word.

(i) He probably will leave tomorrow.

97 An interesting proposal concerning cyclicity of overt movement operations is made by Kawashima and Kitahara (1995). Assuming Epstein's (1995) definition of c-command given in (47), they observe that noncyclic merger or movement operations
II.18. Summary

Below I present a brief summary of some of the proposals I have advanced in this chapter:

(i) Chomsky’s (1995:226) proposes that operations which are required to define what is a possible derivation are derivationally costless, whereas derivations which are required for a pair \((\pi, \lambda)\) to be legitimate are derivationally costly. Applying this criterion to the three versions of Spell-Out outlined in Chomsky (1993, 1994, 1995:chap. 4), I adopted the version formulated in Chomsky (1994), according to which Spell-Out ships semantic features to the covert component, phonological features to the phonological component, and formal features to both components. Compared to the others, this version entails less derivational cost in that it requires that fewer features be eliminated in the phonological component (see section II.5.2).

---

Introducing an overt category \(C\) yield a phrase-marker violating the LCA: \(C\) will not enter into c-command relations with the material “higher” in the structure and therefore no linear order can be established. Under the assumption that the LCA is not relevant for LF, noncyclic operations are then permitted in the covert component. To account for overt head movement, which is noncyclic, the authors rely on Chomsky’s (1994) proposal that the LCA does not look inside \(X^0\) elements. Although their emphasis is on accounting for the asymmetry between covert and overt movement with respect to cyclicity, one wonders how their system blocks presumably unwanted derivations in which lexical insertion of items without phonological features in the covert component does not proceed cyclically. For arguments that the LCA may hold at LF as well, see Nunes (1995a).
(ii) In section II.6.2, I proposed that a strong feature is a hybrid formal-phonological feature which is a member of the set of the phonological features of a given lexical item. If a lexical item with a strong feature is introduced in the covert component, the derivation crashes at LF: by assumption, the covert component cannot eliminate phonological features. If a strong feature has not been checked when Spell-Out applies, it will be shipped to the phonological component together with the other phonological features. It will then cause the derivation to crash at PF: the rules which apply to the set of phonological features cannot eliminate the formal part of the strong feature and the rules which eliminate unchecked formal features apply only to the set of formal features. Thus, a derivation involving strong features converges only if they are checked overtly. By being checked before Spell-Out, a strong feature is rendered invisible at both interface levels, allowing Full Interpretation to be met.

(iii) I adopted Nunes and Thompson's (forthcoming) definition of domination in set-theoretic terms, which is actually an adaptation of Chomsky's (1994) definition of term (see section II.9.2). Accordingly, the notion of reflexive domination was not resorted to in the revision of the definitions of minimal, internal, and checking domains (see section II.7.1.2).

(iv) I proposed that in order to capture c-command relations involving
adjunction structures, syntactic objects of the type $K = \{<\gamma, \gamma>, \{\alpha, \beta]\}$ should be taken as a segment of a two-segment category, rather than the two-segment category itself (see section II.9.3).

(v) I proposed that Move is to be taken as a mere description of the interaction of the operations Copy, Merge, Form Chain, and Chain Reduction, rather than being a primitive operation of the computational system. I refer to this approach to movement operations as the Copy + Merge theory of movement (see section II.10).

(vi) A framework which takes Move to be a primitive operation such as the one in Chomsky (1994, 1995:chap. 4) has the conceptual problem that Merge is taken to be a suboperation of Move and an operation in its own right, despite the fact that Merge and Move are compared in terms of derivational cost. In the Copy + Merge theory of movement, Merge is always analyzed as an operation and, as such, is compared with the other operations of the computational system (see section II.10).

(vii) Under the Copy + Merge theory of movement, the sequence of derivational steps involving Select and Merge (i.e. lexical insertion) is always less costly than the sequence of derivational steps involving Copy and Merge (i.e. movement), because Select is a defining property of a derivation, whereas
Copy is associated with convergence conditions on the pair \((\pi, \lambda)\) (see section II.10.1).

(viii) Procrastinate is not a primitive of the grammar. Rather, it merely describes the result of comparing Spell-Out with the operation Copy at a given derivational step where the numeration has been exhausted. Spell-Out is less costly than Copy, because it is a defining property of a derivation, whereas Copy is related to the legitimacy of the pair \((\pi, \lambda)\) (see section II.10.2).

(ix) The Copy + Merge theory of movement allows instances of sideward movement, where a term of syntactic object \(K\) is copied and merges with the syntactic object \(L\), unconnected to \(K\). By allowing instances of sideward movement, the Copy + Merge theory of movement is able to provide an alternative cyclic derivation to apparent instances of noncyclic adjunction of relative clauses (see section II.10.3).

(x) I proposed that a term of a given structure is to be taken as distinct from all the other terms of that structure by default, unless it is specified as being a copy. The copying operation works as follows: (a) if a term \(T\) has no index, Copy targets \(T\), assigns an unused index \(i\) to it and creates a copy of the indexed term; or (b) if a term \(T\) was already indexed by a previous copying operation, Copy simply creates a copy of the indexed term (see section II.10.4).
(xi) C-command, Last Resort, and the Minimal Link Condition are to be taken as conditions on the application of Form Chain, as stated in (128) (see sections II.10.3, II.10.5, and II.10.6).

(128) *Form Chain:*

The syntactic objects $\alpha$ and $\beta$ can form the nontrivial chain $CH = (\alpha, \beta)$ only if:

(i) $\alpha$ c-commands $\beta$;

(ii) $\alpha$ is nondistinct from $\beta$;

(iii) (at least) one given feature $F$ of $\alpha$ enters into a checking relation with a sublabel of $K$, where $K$ is the head (of the projection) with which $\alpha$ merges; and

(iv) there is no syntactic object $\gamma$ such that $\gamma$ has a feature $F'$ which is of the same type as the feature $F$ of $\alpha$, and $\gamma$ is closer to $\alpha$ than $\beta$ is.

(xii) The elements adjoined to a head $H$ are in the optimal checking domain of $H$. However, maximal projections cannot adjoin to the head with which they enter into a checking relation (the optimal option), because Morphology presumably cannot operate with nonminimal maximal projections within $X^0$. 
elements. Thus, UG must resort to the Spec-head relation in addition to the optimal checking configuration established by adjunction to a head (see section II.13.1).

(xiii) By taking head adjunction to be the optimal checking configuration and by assuming the standard c-command condition on chain links (cf. (128i)), the label of a structure formed by a movement operation can be determined straightforwardly, and the Phrasal Uniformity Condition and the WI processes proposed in Chomsky (1995:chap. 4) can be dispensed with (see section II.13.1 and II.13.2).

(xiv) [-interpretable] features are rendered invisible at the interface by a checking operation and cannot participate in another checking relation after being checked. No further process of erasure of a checked feature is required (see section II.14).

(xv) Chomsky's (1995) proposal that feature mismatch cancels the derivation should be abandoned on conceptual and empirical grounds (see section II.14.4.1).

(xvi) Any type of intervening Case-feature (in English) counts for purposes of the Minimal Link Condition (see II.14.4.1).
(xvii) The $\theta$-Criterion is to be taken as a convergence condition on $\theta$-assignment (see section II.16).
CHAPTER III

LINEARIZATION OF NONTRIVIAL CHAINS AT PF

(or Why Traces Are Not Phonetically Realized)

III.1. Introduction

In the system proposed in Chomsky (1994, 1995:chap. 4), the inner workings of the operation Move encompass four suboperations: (i) a suboperation of copying; (ii) a suboperation of merger; (iii) a suboperation forming a chain between two copies; and (iv) a suboperation deleting traces for PF purposes. As discussed in section II.10, this approach to movement operations has the conceptual problem that Move is to be compared with the operation Merge for purposes of economy, even though Merge is also a suboperation of Move. In addition, this system assumes an asymmetry

---

1 This chapter is an extensive expansion and revision of Nunes (1994a). Part of this material has been presented at the Fifth Student Conference of the University of Maryland (May, 1994), at the Universidade Federal de Minas Gerais (June, 1994), and the Universidade Estadual de Campinas (June, 1994), at the Universidade Federal de Santa Catarina (July, 1994), at the I Congresso Internacional da Associação Brasileira de Linguística (Universidade Federal da Bahia, September, 1994), at the 19th Penn Linguistics Colloquium (University of Pennsylvania, February, 1995), at NELS 26 (MIT/Harvard University, October, 1995), at the International Conference of Interfaces in Linguistics (Universidade do Porto, November, 1995), and at the University of Southern California (Fall, 1995). I am thankful to these audiences. I am also indebted to Mark Arnold, Norbert Hornstein, Richard Kayne, Ellen Thompson and Juan Uriagereka for valuable suggestions and comments.
between overt and covert applications of Move in terms of derivational cost (the Procrastinate principle of Chomsky (1993:30)), which violates the uniformity condition on the mapping from N to λ (see section II.3).

In section II.10, I proposed an alternative to this analysis, which I referred to as the Copy + Merge theory of movement, in which Move is not an operation of the computational system, but simply the description of the effects of the interaction of the operations Copy, Merge, Form Chain, and Chain Reduction, which correspond to the suboperations in (i)-(iv) above.

The motivation for this alternative view comes from both conceptual and empirical considerations. On the conceptual side, the comparison between lexical insertion (the sequence of derivational steps involving Select and Merge) and movement becomes more natural, given that Merge is always analyzed as an operation. Under the Copy + Merge theory of movement, the derivational cost ascribed to movement operations follows from the status of the operation Copy (see section II.10.1). Since Copy is an operation which is required by convergence considerations, it is derivationally costly, whereas the operations Select and Merge, which are required as defining properties of a derivation, are derivationally costless (see Chomsky (1995:226) and section II.4). In turn, the effects of Procrastinate are derived in compliance with the uniformity condition on the mapping from N to λ by comparing Copy with Spell-Out at the point where the numeration has been exhausted. Since Spell-Out is required as a defining
part of a derivation, it is derivationally costless and therefore is to be preferred over Copy (see section II.6), deriving the fact that overt movement is only triggered in order to allow the derivation to converge at PF by eliminating strong features (see section II.10.2).

The Move-based approach and the Copy + Merge theory of movement differ in their empirical predictions in that the latter permits instances of sideward movement, where a term of a syntactic object K is copied and merges with a syntactic object L, unconnected with K. In these cases, the sequence of derivational steps involving Copy and Merge do not yield a chain, overcoming the redundancy between movement operations and chain formation noted by Brody (1995), for instance. To the extent that parasitic gap constructions and across-the-board extraction can be successfully analyzed in terms of sideward movement, as I argue to be the case in chapter IV, the Copy + Merge theory of movement is to be preferred over a Move-based theory on empirical grounds as well.

Under the Copy + Merge theory of movement, conditions which are taken to apply to Move such as c-command, Last Resort and the Minimal Link Condition are reinterpreted as holding of the operation Form Chain (see sections II.10.3, II.10.6 and II.14.2):
(1) Form Chain:

The syntactic objects $\alpha$ and $\beta$ can form the nontrivial chain $CH = (\alpha, \beta)$ only if:

(i) $\alpha$ c-commands $\beta$;

(ii) $\alpha$ is nondistinct from $\beta$;

(iii) (at least) one given feature $F$ of $\alpha$ enters into a checking relation with a sublabel of $K$, where $K$ is the head (of the projection) with which $\alpha$ merges; and

(iv) there is no syntactic object $\gamma$ such that $\gamma$ has a feature $F'$ which is of the same type as the feature $F$ of $\alpha$, and $\gamma$ is closer to $\alpha$ than $\beta$ is.

So far, I have delayed a discussion of why traces cannot be phonetically realized. It should be noted that the lack of phonetic realization of traces is an issue which is independent from the specifics of the Copy + Merge theory of movement and arises in any version of the copy theory of movement. If traces are taken to be copies, they should in principle have the same properties as the elements they are copied from, including phonetic realization. Hence, the optimal version of the copy theory of movement will be the one which accounts for differences between heads of chains and traces.
without postulating intrinsic differences between them. The goal of this chapter is to pursue such an optimal approach.

The chapter is organized as follows. In section III.2, I review the Binding-theoretic reasons which led Chomsky (1993) to revive the copy theory of movement in the Minimalist Program; some advantages of treating traces as copies are then discussed in section III.3. In section III.4, I discuss apparent differences between heads of chains and traces regarding accessibility to the computational system, linearization, and phonetic realization. In section III.5, I show that the phonological component recognizes the difference between nondistinct copies and terms with identical sets of features which are distinct in the initial numeration. Section III.6 is the core of this chapter; I propose that the lack of phonetic realization of traces follows from the interaction of the fact that traces are subject to Kayne’s (1994) LCA with economy considerations concerning the number of applications of deletion of unchecked features in the phonological component. In section III.7, I discuss some apparent counterexamples to the analysis proposed in section III.6. The conclusions of the chapter are then presented in section III.8. Finally, I suggest in the appendix that the analysis developed in this chapter may be extended to null operator constructions, accounting for why null operators are not phonetically realized.
As discussed in section II.2, one of the main goals of the Minimalist Program is to show that all principles and parameters should be stated in terms of the interface levels LF and PF (see chapter II:fn. 3). One of the challenges for such an approach is Binding Theory. Consider a structure such as (2) below, for instance, which is ambiguous in that the anaphor can be coreferential with the embedded or the matrix subject. The reading under which the anaphor is bound by the embedded subject appears to be a counterexample to the claim that Principle A only applies at LF, because the anaphor seems to be bound prior to LF, before movement of the wh-phrase takes place.2

(2) [ Johni wondered [ which picture of himselfi,j ]i k Billj saw t_k ]

One possible way to analyze (2), which is consistent with Principle A being an LF condition, is to reconstruct the wh-phrase to the position of the trace at LF, allowing the anaphor to be bound by the embedded subject at that level (see Chomsky (1976), Hornstein (1984), and Barss (1986), among others).

---

2 Whether Binding Theory is to be computed at LF or at the C-I interface is a controversial issue. Given that the main purpose of this section is to discuss the relevance of the copy theory of movement for the elimination of non-interface levels (see section III.3.1), I will simply follow Chomsky (1993), according to whom the Binding Theory holds at LF. Chomsky’s (1993) results nevertheless remain untouched, if we take them to show that Binding Theory cannot apply prior to LF.
If such an analysis is to be pursued, one needs to explain why reconstruction appears to be optional in (2), hence the ambiguity; obligatory in (3) and (4), yielding Principle B and C effects; and impossible in (5), allowing coreference between John and him.³

(3)  [ John wondered [ which picture of him\_] | Bill\_ saw ti ]

(4)  [ John wondered [ which picture of Bill\_] | he\_ saw ti ]

(5)  [ [ the claim that John\_ was asleep \_] \_seems to him\_ [ ti to be correct ] ]

³ Lebeaux (1988) argues that in constructions such as (ia) below, the noun complement clause must be present at D-Structure to satisfy the 0-Criterion, yielding a Principle C effect; on the other hand, since the relative clause in constructions such as (ib) is an adjunct, it need not be present at D-Structure and may be inserted by a generalized transformation after the wh-phrase moves, bleeding Principle C. Chomsky (1993:24) captures Lebeaux’s insight without reference to D-Structure conditions, by assuming that overt substitution operations must extend their target, while adjunction operations need not. In the system developed in Chomsky (1995: chap. 4), the cyclic property of head-complement relations presumably follows from the interaction of two assumptions: (i) conceptual simplicity requires that only root structures merge (see Chomsky (1995:248), and section II.10.3); and (ii) a nontrivial chain cannot assign a T-role (see Chomsky (1995:313)). Throughout this discussion, I will take the of-phrase of (2)-(4) to be a complement; thus, it cannot merge with picture after the wh-phrase moves to the embedded Spec of CP, behaving like the noun complement clause in (ia) and not like the adjunct relative clause in (ib). The contrast between (iia) and (iib), discussed in Chomsky (1994:22), remains unexplained, however.

(i)  a. *[ [ which claim [ that John\_ was asleep ] ] was he\_ willing to discuss ]
    b. [ [ [ which claim] [ that John\_ made ] ] was he\_ willing to discuss ]

(ii) a. Which picture of John’s brother did he\_ expect that you would buy?
    b. *Pictures of John’s brother, he\_ never expected that you would buy.
As Chomsky (1993:34) observes, reconstruction is an odd operation from a Minimalist perspective, because it undoes a movement which is required for convergence. Chomsky (1993:35) accounts for the ambiguity of (2) without resorting to reconstruction, by adopting the copy theory of movement. According to the copy theory, a trace is a copy of the moved element which gets deleted in the phonological component but remains present at LF, thus being able to be computed for Binding Theory purposes. Therefore, (2) is to be seen as an abbreviation of (6).

(6)  [ John wondered [wh- which picture of himself ] Bill saw 
    [wh- which picture of himself ] ]

The LF component then converts the wh-phrases of (6) into (7a) or (7b) "by an operation akin to QR" (Chomsky (1993:35)), as shown in (8a) and (8b), respectively.

(7)  a.  [ [ which picture of himself ]i [wh- ti ] ]
    b.  [ [ which ]i [wh- ti picture of himself ] ]
Deletion of the nonquantificational material in the operator position, and of quantificational material in the variable position in (8) yields the structures in (9a) and (9b), which receive the interpretation given in (10a) and (10b) respectively. 4 (9a) receives the reading in which the anaphor is bound by the matrix subject, and (9b) receives the reading in which the

4 Notice that in (8), the same option of raising is chosen for both the operation and the variable. Chomsky (1993:36) suggests that mixed options cause the derivation to crash. Within Chomsky’s (1993) system, it is more plausible to assume that the derivation converges but receives a deviant interpretation at the C-I interface, because no proper operator-variable relation can be constructed. The mixed options represented in (i) and (ii), for instance, should receive the interpretations in (iii), after deletion takes place in (ii). In (iiia), the restriction of the operator is vacuous, whereas in (iiib) the quantifier does not have a restriction. See Nunes (1995a), where I show that all the data discussed in this section can be accounted for, if we extend to regular wh-phrases in situ by an interrogative complementizer (see Baker (1970) and section II.14.3.3).
anaphor is bound by the embedded subject.⁵

(9)  a. [John wondered [ which picture of himself ]ₗ Bill saw [wh- tᵢ ]ₗ ]
    b. [John wondered [ which ]ₗ Bill saw [wh- tᵢ picture of himself ]ₗ ]

(10)  a. [John wondered [ which x, x a picture of himself ] Bill saw x ]
    b. [John wondered [ which x ] Bill saw [ x picture of himself ] ]

The question now is to explain why the two options exemplified in (7) are not available in (3) and (4). In other words, were (3) and (4) able to undergo the option shown in (7a), the pronoun and Bill could be coreferent without giving rise to Principle B (cf. (11)) or Principle C (cf. (12)) effects:

(11)  [John wondered [ which picture of himj ]ₗ Billₗ saw [wh- tᵢ ]ₗ ]

(12)  [John wondered [ which picture of Billj ]ₗ heₗ saw [wh- tᵢ ]ₗ ]

⁵ Even putting a Move-F approach aside (see Chomsky (1995:sec. 4.4.4) and section II.12), such a QR-type movement is not unproblematic. First, as in standard QR, it is not clear what morphological feature triggers such movement, i.e. how this movement complies with Last Resort (see Hornstein (1995:chap. 8)). Second, this operation does not seem to leave copies, contrary to standard movement under the copy theory. If the trace in (9b), for instance, were a copy of the quantifier that was raised and deleted (see (8b)/(9b)), we would have a quantifier in a variable position, which would undermine the whole purpose of this process of raising and deleting. See Nunes (1995a) for an alternative based on Chomsky’s (1995:291) analysis of wh-phrases in situ (see section II.14.3.3).
Chomsky (1993:41) accounts for the contrast between (2), on the one hand, and (3) and (4), on the other, by postulating a preference for the option in (7b) over the one in (7a), if possible. That is, between two convergent derivations, the system chooses the one which minimizes the restriction in the operator position. Assuming that the pairs (11) and (13), and (12) and (14) converge, the second derivation of each pair blocks the first one, because it minimizes the restriction of the operator. Thus, coreference between the pronoun and Bill is prohibited by Principle B in (13) and Principle C in (14).6

(13) [ John wondered [ which ]; Bill saw [wh- t; picture of him; ] ]

(14) [ John wondered [ which ]; he saw [wh- t; picture of Bill; ] ]

This preference principle will be inapplicable in constructions such as (2) because the options in (7) may or may not yield convergent derivations.

---

6 According to Chomsky (1993:34), the wh-phrase of the sentence in (i) can be assigned a DP variable ranging over houses (the answer could be The old one) or a D variable ranging over entities (the answer could be That (house)). The structures corresponding to these readings are represented in (ii). To the extent that these readings are interpretively distinct, it is not obvious why the DP variable reading should be available in (i), given that its derivations should be blocked in favor of the D variable reading by the preference principle employed to exclude (11) and (12). See Nunes (1995a) for a proposal that economy considerations regarding the number of applications of deletion in the mapping from LF to the C-I system derives the effects of the preference principle.

(i) Guess in which house John lived.

(ii) a. [ guess [ which house ]; John lived in [wh- t; ] ]
    b. [ guess [ which ]; John lived in [wh- t; house ] ]
depending on a previous operation. Chomsky (1993:40) adopts an LF raising approach to anaphors (see Lebeaux (1983)), which he calls cliticization_{LF}, according to which the anaphor or part of it raises by an operation similar to cliticization. (15) below shows the structure in (6) after cliticization_{LF} applies to the upper or to the lower wh-phrase.\(^7\)

\[(15)\] 
\[\begin{align*}
\text{a. [ John self-wondered } \text{[wh- which picture of } t_{self} ] \text{ Bill saw} \hfill \\
\text{[wh- which picture of himself ] ]}
\end{align*}\]

\[\begin{align*}
\text{b. [ John wondered } \text{[wh- which picture of himself ] Bill self-saw} \hfill \\
\text{[wh- which picture of } t_{self} ] ]
\end{align*}\]

If the anaphor of the upper copy undergoes cliticization_{LF}, as in (15a), only option (7a) can yield a convergent derivation, being interpreted as (10a). Option (7b) would require that picture of } t_{self} \text{ in (15a) be deleted, which would "break" the chain (} self, t_{self}).\(^8\)

\(^7\) Chomsky (1993:43) conjectures that Principle A itself could follow from locality restrictions on movement operations, if the cliticization_{LF} approach proves to be correct. Under a Move-F approach, cliticization_{LF} is to be understood as raising of the formal features of anaphors at LF. Still to be addressed in the framework of Chomsky (1995:chap. 4) is the kind of [-interpretable] feature that this operation checks in order for Last Resort to be satisfied. Also unclear is the checking relation between declarative complementizers and the trace of a wh-phrase in instances of successive cyclic movement such as the one in (i):

\[(i) \text{ [CP [ which picture of himself ] did John say [CP [ which picture of himself ] that Sue liked [ which picture of himself ] ] ]]}

\(^8\) Chomsky (1993:41) assumes that option (7b) cannot apply to (15a) because the reflexive would lack a } \theta\text{-role at LF. If the } \theta\text{-Criterion is a condition on } \theta\text{-role assignment, as argued in section II.16, that cannot be the reason for the unacceptability}
lower copy, as in (15b), on the other hand, only option (7b) is possible. Besides breaking the chain (\textit{self}, \textit{tself}) in the embedded clause, option (7a) would end up with an unmoved anaphor in the upper wh-phrase, going against Chomsky's (1993:40) assumption that the operations in (7) necessarily follow cliticization_{LF}.\footnote{One could ask what prevents cliticization_{LF} from applying to both copies in (15). In constructions in which only two copies are available, the answer within Chomsky's (1993) system is trivial. Regardless of which option of (7) the computational system chooses, one clitic chain will be broken (see fn. 8). But the question still arises when more than two copies of the wh-phrase are involved. The structure in (i), for instance, could yield (ii), with cliticization_{LF} applying twice and option (7b) being taken. In (ii), both anaphor chains are kept intact, which should yield an interpretation in which each copy of the anaphor is bound by a different subject. See Nunes (1995a) for an account of the impossible reading in (ii) in terms of linearization of chains at LF.}

This analysis explains why the potential antecedent for the reflexive in a sentence such as (16) below depends on the literal or idiomatic interpretation of \textit{take picture}: under the literal meaning either subject can be the antecedent for \textit{himself}, but under the idiomatic reading only the embedded subject can. Under the assumption that an idiomatic expression must be a unit at LF (see Chomsky (1993:39)), only the representation in (17b), in which the anaphor must be interpreted as coreferential with the embedded subject, allows an idiomatic interpretation.

\begin{itemize}
\item[(i)] \text{[ [ which picture of himself ] did John say [ [ which picture of himself ] Peter liked [ [ which picture of himself ] ] ] ]}
\item[(ii)] \text{[ [ which ], did John self-say [ [ t, picture of tself ] Peter self-liked [ t, picture of tself ] ] ]}
\end{itemize}
(16) John wondered which picture of himself John took.

(17) a. [ John self-wondered \[ wh \text{-} \text{which picture of } t_{\text{self}} \] \text{, Bill took} \[ wh \text{-} t_i \] ]
   
b. [ John wondered \[ wh \text{-} \text{which } \] \text{, Bill self-took} \[ wh \text{-} t_i \text{ picture of } t_{\text{self}} \] ]

Finally, Chomsky (1993:42) accounts for the absence of Principle C effects in raising constructions such as (5), repeated below in (18), by assuming that the reconstruction process illustrated in (7) applies to operator-variable constructions but not to A-chains. Noting that raising constructions such as (19) do allow an idiomatic reading, Chomsky (1993:42) suggests that the head of an A-chain is interpreted with respect to scope, while the lowest trace enters into idiom interpretation and θ-marking more generally (the intermediate traces being deleted or ignored).

(18) [ [ the claim that John₁ was asleep ]₁ seems to him₁ [ t₁ to be correct ] ]

(19) Several pictures were taken.

Notice that Chomsky's conclusion that reconstruction is prohibited in
A-chains is not necessary.\textsuperscript{10} If reconstruction is optional for A-movement, the reading of (18) under which the subject does not reconstruct will allow coreference between $\textit{John}$ and $\textit{him}$. Evidence that reconstruction in A-chains is not only optional, but that it can make use of any intermediate link, is provided by the sentences in (20), respectively represented in (21) (from Lebeaux (1991:234)).

\begin{enumerate}
\item[(20)]
\begin{enumerate}
\item His$_i$ mother$_j$'s bread seems to every man$_i$ to be known by her$_j$
\quad to be the best there is.
\item *His$_i$ mother$_j$'s bread seems to her$_j$ to be known by every
\quad man$_i$ to be the best there is.
\end{enumerate}
\end{enumerate}

\begin{enumerate}
\item[(21)]
\begin{enumerate}
\item [ [ his mother's bread ] seems to every man [ [ his mother's 
\quad bread ] to be known by her [ [ his mother's bread ] to be the
\quad best there is ] ] ]
\item [ [ his mother's bread ] seems to her [ [ his mother's bread ] to 
\quad be known by every man [ [ his mother's bread ] to be the best
\quad there is ] ] ]
\end{enumerate}
\end{enumerate}

Assuming that reconstruction under A-movement is optional, any of

\textsuperscript{10} For reasons of familiarity, I continue to use the term $\textit{reconstruction}$ in the discussion of (18)-(22). It should be kept in mind, however, that this term is meant to refer to the interpretation of only one of the links of the chain (see Hornstein (1995:chap. 8) and Nunes (1995a) for further discussion).
the copies of the phrase *his mother's bread* in (21) can in principle be used for interpretation. However, in order for either sentence in (20) to receive the interpretation indicated by the indices, two requirements regarding the reconstruction of *his mother's bread* must be met: (i) it must be reconstructed outside of the c-command domain of the pronoun *her* in order to circumvent Principle C effects; and (ii) it must be reconstructed within the c-command domain of the universally quantified phrase in order for the pronoun *his* to be bound.

The only position which can satisfy both requirements in (21a) is the intermediate subject position. Thus, if (21a) is converted into (22) by deleting the topmost and the lowest copies of the A-chain (see Hornstein (1995:chap. 8) and Nunes (1995a) for further discussion), the relevant reading of (20a) is correctly derived:

(22) seems to every man, [ his mother's bread ] to be know by her to be the best there is

On the other hand, no reconstruction of the chain in (21b) allows the two requirements to be satisfied: if *his mother's bread* is reconstructed in the matrix or the intermediate subject position, the pronoun *his* cannot be bound; if it reconstructs in the most embedded subject position, the pronoun can be bound, but coreference between *her* and *his mother* induces a
Principle C violation. Hence, the contrast in (20) argues against the lack of reconstruction in A-chains.\footnote{See Hornstein (1995:chap. 8) for further arguments. Hornstein's (1995) analysis of scope ambiguity in terms of reconstruction in A-chains is reinterpreted in Nunes (1995a) in terms of linearization of chains at LF. Still to be accounted for is the absence of reconstruction under ECM verbs, as illustrated by (i), which induces a Principle B effect (see Chomsky 1995:327). For lack of reconstruction effects under negation, see Hornstein (1995:chap. 8).}

III.3. Some Advantages of the Copy Theory

III.3.1. Elimination of Non-interface Levels

Chomsky's (1993) analysis of Binding Theory phenomena reviewed in section III.2 makes a strong case for the elimination of D- and S-Structure as syntactic levels where principles of Binding Theory can apply. As pointed out by Chomsky (1993:25), conceptual reasons dictate that an analysis which is able to handle Binding Theory data solely in LF terms is to be preferred over one which resorts to D- or S-Structure conditions; whereas LF is a conceptually necessary level of representation, D- and S-Structure are not (see section II.2). As seen in section III.2, it is the copy theory of movement which provides the basis for such a conceptually sound analysis (see fn. 2).

(i) John expects him, to seem to me, to be intelligent
III.3.2. Interpretation of Idioms at LF

Another advantage of the copy theory is that it provides the basis for idiom interpretation at LF. Recall that in the Minimalist Program, the notions of D-structure, the Projection Principle and reconstruction are abandoned for conceptual and/or empirical reasons (see Chomsky (1993:sec. 4.3)). Given that there is empirical evidence showing that displaced elements may be understood idiomatically as if they had not moved from their original position (cf. (19), and (16) under the reading in (17b)), there must be some way to allow two constituents separated by a movement operation to be interpreted as an idiomatic expression. The copy theory of movement provides such means at no cost: the same copies left behind by movement which are computed for Binding Theory purposes may be used in the interpretation of idioms, as well.12

III.3.3. Elimination of Reconstruction

Still another advantage of adopting the copy theory of movement regards the elimination of reconstruction as an additional operation in the

---

12 In the absence of D-Structure and the Projection Principle, the copy theory of movement could in principle allow the \( \theta \)-Criterion, now understood as a condition on \( \theta \)-role assignment (see section II.16), to be ensured at LF. However, if nontrivian chains cannot assign \( \theta \)-roles (see Chomsky (1995:313) and Uriagereka (forthcoming:chap. 6)), at least some \( \theta \)-role assignment will have to take place prior to the movement of the \( \theta \)-role assigners.
theory of grammar. As observed in section III.2, an analysis based on
reconstruction is dubious on conceptual grounds, because it undoes a
movement that is necessary for convergence. Under the copy theory, on the
other hand, the copies left by movement operations can be used for the
purposes of Binding Theory at LF (see fn. 2).

Furthermore, reconstruction may lead to wrong empirical predictions.
Take, for instance, an approach that assumes reconstruction at LF in order to
allow the idiomatic reading of *take pictures* in (19), repeated below in (23),
and takes Principle A to be checked in the course of the derivation in order
to account for the ambiguity of sentences such as (16), repeated in (24). As
noted by Chomsky (1993:39), under such an approach, the anaphor of (24)
could be bound by the matrix subject prior to LF and the wh-phrase could be
reconstructed at LF, making the idiomatic reading possible. However, as
discussed in section III.2, this reading is not available in (24); the idiomatic
interpretation of *take picture* in (24) is only possible if the anaphor is bound
by the embedded subject.

(23) Several pictures were taken.

(24) John wondered which picture of himself Bill took.

Thus, the analysis combining the copy theory of movement with the
claim that Binding Theory does not apply prior to LF (see fn. 2 and section III.3.1) accounts for the correlation between the idiomatic reading and anaphor interpretation in sentences such as (24), whereas an analysis which assumes reconstruction and takes (some principles of) Binding Theory to apply prior to LF does not.

III.3.4. Elimination of Traces as Grammatical Primitives

Within the Principles and Parameters Theory, the term *trace* actually encompasses several types of objects which have in common the properties of being produced by a movement operation and being phonetically null. Regardless of whether traces have intrinsic features or acquire some feature specification in the course of the derivation (see Chomsky (1982), for instance), they end up being different entities. A trace of an NP occupying an A-position, for instance, is subject to the Principle A of Binding Theory, a trace of an NP occupying an operator position is a variable subject to Principle C, and a trace of a verb is subject to neither (see Chomsky (1973, 1981, 1982), Fiengo (1977), Aoun, Hornstein, Lightfoot and Weinberg (1987), Rizzi (1990), Cinque (1990), and Epstein (1991), among others).

In this traditional approach, the distinct nature of each type of trace is captured by means of the coindexation between the trace(s) and the moved element. This notation encoding the relation between links of a chain was
often reified as a theoretical primitive subject to formal principles. Given that indices should not be taken as theoretical entities within the Minimalist framework (see chapter II:fn. 26), an alternative analysis of traces is called for.

Another conceptual reason for the elimination of traces as theoretical primitives within the Minimalist framework is that the introduction of traces in the course of the derivation should violate the inclusiveness condition on the mapping from N to \( \lambda \) (see section II.3).

A very important advantage of the copy theory of movement is that it provides the means for treating the notion of trace itself as epiphenomenal. If a trace is actually a copy of a given moved element, it need not be specified as a primitive in UG; it is either a lexical item (here taken to include functional heads) or an \( X' \)-theoretic object built from lexical items. From a Minimalist perspective, the properties of different traces should then follow either from the content of the copies themselves or from the movement operation. For instance, the locality restriction on the distribution of traces may follow from the Minimal Link Condition (see section II.10.6), while the interpretation of a wh-trace as a variable may be due to its own features (see Chierchia (1991), Hornstein (1995:chap. 6), and Nunes (1995a), for instance).

In this chapter, I show that traces are not grammatical primitives. Rather, the term *trace* should be taken as a mere label for a chain link whose deletion for purposes of linearization is more economical than the deletion
of another chain link. Before I present the details of my proposal, I will first discuss properties which have been taken to be inherent to traces such as inaccessibility to the computational system (see Chomsky (1995:chap. 4)), irrelevance for the LCA (see Kayne (1994) and Chomsky (1994, 1995:chap. 4)), and the lack of phonetic realization.

III.4. Differences between Heads of Chains and Traces

The simplest — and therefore most desirable — version of the copy theory of movement would take heads of chains and traces to be subject to the same constraints. This appears to be unfeasible, given that one of the fundamental properties of traces is that they cannot be phonetically realized. In addition to this difference between heads of chains and traces regarding phonetic realization, two other asymmetries have been proposed: (i) traces (of A-movement) are not accessible to the computational system (see Chomsky 1995:301, 304); and (ii) traces are not subject to the LCA (see (Kayne 1994:chap. 2, fn. 3; and Chomsky (1994:28)).

If traces are to be taken as descriptive notions, as suggested in section III.3.4, these and any other differences between the head of a chain and its traces must be shown to follow from independently motivated considerations which do not rely on properties intrinsically assigned to traces. Below I argue that there is no such accessibility or linearization
difference between traces and heads of chains; moreover, I propose that the fact that traces are subject to the LCA is one of the factors that contributes to their lack of phonetic realization. It will thus be shown that the simplest copy theory is attainable, despite apparent evidence to the contrary.

III.4.1. Accessibility to the Computational System

In pre-Minimalism versions of the Principles and Parameters Theory, A-chains were subject to the Chain Condition (see Chomsky (1986a:63) and Chomsky and Lasnik (1993:523-524)), according to which every argument chain must be headed by a Case-marked position and must terminate in a θ-position. If every movement operation forms a new chain (but see section II.10.3 and chapter IV), a problem for the Chain Condition is raised by constructions involving successive NP raising such as the one illustrated in (25) below, with indices taken to indicate different positions. The question is how each of the chains in (26), which are formed by raising *John* in (25), satisfies the Chain Condition. The chain CH₂ in (26), for instance, does not satisfy either of the requirements of the Chain Condition.

(25) \[ \text{John}_4 \text{ seems } [ \text{John}_3 \text{ to be likely } [ \text{John}_2 \text{ to } \text{be kissed John}_1 ] ] ] ]
Chomsky and Lasnik (1993:563) propose that in cases of successive cyclic movement, the Chain Condition should hold of the linked chain, which is the chain formed by linking two chains $CH_1$ and $CH_2$, where the tail of $CH_1$ is the head of $CH_2$. According to this proposal, although the chains $CH_1$, $CH_2$ and $CH_3$ in (26) do not satisfy the Chain Condition, the linked chain $CH_4 = (John_4, John_3, John_2, John_1)$ obtained by linking $CH_1$, $CH_2$ and $CH_3$ does.

In the system proposed in Chomsky (1995:chap. 4), the problems posed by successive A-movement arise with respect to both the Chain Condition, which Chomsky (1995:300) incorporates into the Minimalist framework, and Full Interpretation. Let us see why. Consider first an A-chain involving only two links, as illustrated in (27).

\begin{enumerate}
  \item \textbf{(26)}
    \begin{enumerate}
      \item $CH_1 = (John_2, John_1)$
      \item $CH_2 = (John_3, John_2)$
      \item $CH_3 = (John_4, John_3)$
    \end{enumerate}
\end{enumerate}

In Chomsky's (1995:chap. 4) theory, movement operations are triggered by the need to eliminate [-interpretable] features (see section II.14). Once the copy theory of movement is assumed, it must be ensured that in a
convergent derivation, no "traces" have [-interpretable] features. Chomsky (1995:chap. 4, fn. 12) assumes that "the features of a chain are considered a unit: if one is affected by an operation, all are".\(^{13}\) For instance, since the Case-feature of the head of the chain CH = \((John, John)\) in (27b) is deleted and erased after entering into a checking relation with the Case-feature of the T head, the correspondent feature of the tail of CH is also deleted and erased. Therefore, CH in (27b) satisfies both the Chain Condition and Full Interpretation.

As for successive A-movement, Chomsky (1995:300) assumes that the tail of a chain formed by raising the head of a nontrivial chain CH = \((α, t)\) is the trace \(t\) rather than \(α\). According to this assumption, the chains formed by raising \(John\) in (25) are the ones given in (28) below, rather than the ones in (26).

\[(28)\]

\[\text{a. CH}_1 = (John_2, John_1)\]
\[\text{b. CH}_2 = (John_3, John_1)\]
\[\text{c. CH}_3 = (John_4, John_1)\]

In (28), only CH\(_3\) satisfies the Chain Condition and has checked its Case-feature; the other two chains, on the other hand, violate the Chain

---

Condition and their unchecked Case-features should induce a violation of Full Interpretation at LF. In order to prevent these results, Chomsky (1995:301) admittedly stipulates that if \( \alpha \) heading the chain \( \text{CH} = (\alpha, t) \) raises, the formal features of the trace created by this operation are deleted and erased. Intermediate traces are thus invisible at LF and the only chain subject to interpretation is the pair \( (\alpha, t) \), where \( \alpha \) is in the highest position of raising and \( t \) is in the position of lexical insertion.\(^{14}\) Erasure of the formal features of an intermediate trace \( t \) of an argument renders \( t \) inaccessible to the computational system; intermediate traces therefore are taken to be unable to enter into checking relations and do not induce Minimal Link Condition violations (see Chomsky (1995:301)).

Noting that in instances such as (29), the formal features of the trace of wh-movement may be required for further checking relations (see sections III.4.1.3 and III.4.1.4 for discussion), Chomsky (1995:303) restricts to cases of A-movement the stipulation that the formal features of traces must be deleted and erased.

\(^{14}\) The semantic features of intermediate traces are not deleted, though. Chomsky (1995:chap. 4, fn. 75) suggests that "[d]epending on exactly how interface operations are understood, the semantic features of intermediate traces could be accessible to interpretive operations before they become invisible for further interpretation, possibly allowing implementation of ideas about the interpretive role of intermediate traces".
There are two different issues at stake here. The first one concerns the feature composition of traces at LF. It is clear that if traces contain unchecked [-interpretable] features, Full Interpretation will not be satisfied and the derivation will crash at LF. Therefore, it must be the case that these features are somehow eliminated; furthermore, we expect the elimination of these features to follow Minimalist guidelines.

The second issue relates to accessibility of traces for operations of the computational system. Under the copy theory of movement, we should expect either that there is no difference between heads of chains and traces with respect to accessibility, or that the differences eventually found must not be dependent on properties intrinsically attributed to traces. From this point of view, Chomsky's (1995:301) proposal that traces of A-chains are inaccessible to the computational system is conceptually dubious. First, it admittedly rests on the stipulation that the formal features of traces of A-movement are erased when the head of a nontrivial chain is raised. Second, it crucially relies on the distinction between A and A'-movement, whereas one of the goals within the Minimalist Program is to eliminate any such distinctions (see Chomsky 1995:317-318).
We have already seen (see section II.14.1) that there is no compelling evidence for postulating an erasure operation. Below I show that the claim that erasure renders traces of A-movement inaccessible to the computational system in fact leads to wrong empirical predictions.

The discussion is organized as follows. In section III.4.1.1, I point out some problems for Chomsky's (1995:chap. 4) general account of the elimination of [-interpretable] features of traces, focusing on his claim that traces of A-movement are inaccessible to the computational system. In section III.4.1.2, I propose an alternative which takes the elimination of some formal features of traces to be related to the uniformity of chains with respect to the feature composition of their links. Sections III.4.1.3 and III.1.4.4 show how instances of A'-movement such as (29) are treated under this alternative approach. Finally, section III.4.1.5 presents empirical evidence against Chomsky's proposal that traces of A-movement are inaccessible to the computational system.15

III.4.1.1. Inaccessibility of Traces: Some Problems

As mentioned above, Chomsky (1995:chap. 4, fn. 12) assumes that if a feature of a chain link is affected by an operation, the corresponding feature

15 Here I will not discuss Chomsky's (1995:365) conjecture that "α can be attracted by K only if it contains no trace", given that he dismisses this possibility later, because it would wrongly allow adjunction of the formal features of the object to the expletive of transitive expletive constructions.
of the other links of the same chain is also affected. Thus, the Case-feature of the tail of the chain CH = \((\textit{John}, \textit{John})\) in (27), repeated below in (30), is deleted and erased because the Case-feature of the head of CH is deleted and erased after entering into a checking relation with the Case-feature of T.

\[(30) \quad \text{a. John was kissed.} \]
\[(30) \quad \text{b. [ John T [ was kissed John ] ]} \]

Although this proposal provides a technical way for the [-interpretable] feature of the trace in (30) to be invisible for purposes of Full Interpretation at LF, the result achieved does not come for free. First, the Case-feature of the T head is in a sense participating in two checking relations, which otherwise is not a possibility for a [-interpretable] feature. Second, the Case-feature of the trace of John is being affected by a checking relation with T even though it is not in the checking domain of T. Third, as pointed out to me by José Camacho (p.c.), the copy of John in object position is being affected by a checking relation even though it is in a position where it participates in a 0-relation; this, however, is at odds with the assumption about the complementarity between 0- and checking relations (see Chomsky (1995:312) and section II.7.5).

In section III.6, I will show that if traces are unaffected by the checking relations that heads of chains participate in, the phonological component
has an independent criterion to choose which chain links to delete. Before we get into this discussion, let us consider the paradigm in (31).

(31)  

a. John expected [ a book ]i to be given to Mary  
b. John expected that [ a book ]i was given to Mary  
c. *John expected that [ a book ]i was given to Mary

Given that Case is a [-interpretable] feature, the verb *expect in (31a) and (31b) should not be understood as a single verb which may or may not assign its Case-feature; rather, (31a) and (31b) should involve two different lexical entries of *expect: a Case-assigning entry in (31a) and a Caseless entry in (31b) (see Chomsky (1995:308)). Assuming this to be true, let us consider the derivation of (31c). The DP a book raises from the object position to check the strong feature of the embedded T; this movement also allows the head of chain CH = ([ a book ], t) to check its Case-feature, which is then deleted and erased, according to Chomsky (1995:chap. 4). As a side effect, the Case-feature of the tail of CH is also deleted and erased.

The unacceptability of (31c) should then follow from the fact that Mary does not have its Case-feature checked and the derivation crashes at LF. Notice, however, that this reasoning holds true only under the assumption that the verb *expect in (31c) is not a Case-assigner. Suppose, on the other hand, that in (31c) we have the Case-assigner expect. If so, the formal
features of *Mary* could raise and check its Case-feature against the Case-feature of *expected*, and the derivation should converge. Crucially, although the embedded subject and the direct object c-command *Mary*, their Case-features have been erased and cannot prevent movement of the formal features of *Mary*. Therefore, the unacceptability of the sentence resulting from (31c) under this derivation is not accounted for.

On the other hand, if traces are not affected by operations affecting the head of their chain, the derivation of (31c) described above is ruled out straightforwardly. Under the standard assumption that in constructions such as (31), the direct object c-commands the indirect object (see Barss and Lasnik (1986)), movement of the formal features of *Mary* across the direct object in (31c) for Case purposes yields a Minimal Link Condition violation: the direct object has an unchecked Case-feature and it is closer to *expected* than *Mary*. Conceptual and empirical considerations thus appear to converge against Chomsky's claim that when a given link of a chain CH is affected by an operation, all the other links of CH are affected in the same way.

At any rate, this proposal about checking features of a chain cannot handle instances of multiple raising such as (25), repeated here in (32), without further assumptions.
(32)  a. John seems to be likely to be kissed.

    b. [ John₂ to [ be kissed John₁ ] ]

    c. [ John₃ to [ be likely [ John₂ to [ be kissed John₁ ] ] ] ]

    d. [ John₄ seems [ John₃ to be likely [ John₂ to [ be kissed John₁ ] ] ] ]

When John moves from the object to the embedded subject position, forming the structure in (32b), the Case-feature of neither the head nor the tail of the chain CH₁ = (John₂, John₁) is checked, because only the D-feature of John₂ enters into a checking relation with the infinitival T head. The same holds of the chain formed by the next movement, which Chomsky (1995:300) takes to be CH₂ = (John₃, John₁). By contrast, the chain CH₃ = (John₄, John₁) allows John₁ to have its Case-feature checked because the Case-feature of John₄ enters into a checking relation with the Case-feature of the matrix T. Since the copies John₂ and John₃ still have an unchecked Case-feature, the derivation should violate Full Interpretation at LF, yielding an undesirable result.

As mentioned above, in order to prevent the chains CH₁ = (John₂, John₁) and CH₂ = (John₃, John₁) in structures such as (32) from violating the Chain Condition, Chomsky (1995:301) suggests that when the head α of a nontrivial A-chain raises, the formal features of the trace of α are erased. Thus, the formal features of John₂ and John₃ are erased when the structures
in (32c) and (32d), respectively, are formed, which presumably allows the chains $CH_1 = (John_2, John_1)$ and $CH_2 = (John_3, John_1)$ to (vacuously) satisfy the Chain Condition. Erasure of the formal features of $John_2$ and $John_3$ then allows the derivation to comply with Full Interpretation at LF by eliminating the [-interpretable] features of these copies.

There are several problems with this approach, besides introducing the special checking operation for the trace in constructions such as (30b) and the stipulated erasure of the formal features of intermediate traces employed in (32). The first one regards "reconstruction" in intermediate traces. As discussed in section III.2, intermediate traces may be used for the computation of scope and binding (see Barss (1986), Lebeaux (1991), Aoun and Li (1993), and Hornstein (1995), among others). Under the assumption that formal features are relevant for this computation, erasure of the formal features of intermediate traces would wrongly rule out the interpretation obtained through reconstruction in intermediate traces.

The second problem regards the generality of this approach, given that it takes the Chain Condition to be the motivation for the required elimination of ([-interpretable]) formal features. The Chain Condition arguably does not extend to head movement; movement of T to Comp, for instance, yields the chain $CH = (T, T)$, whose head is not involved in Case checking and whose tail is not involved in 0-role assignment. Bearing in mind that the Chain Condition holds only of A-chains, let us consider a
structure such as (33), for instance, which is formed after the main verb adjoins to the light verb, yielding the chain \( \text{CH}_1 = (V, V) \), and the complex verbal head adjoins to \( T \), forming the chain \( \text{CH}_2 = ([V+V], [V+V]) \).

\[
(33) \quad [\text{TP SU} \left[ \left[ V+V \right]+T \right] \left[ \nu P, \left[ V+V \right] \left[ V P \ V O B \right] \right] ]
\]

Under the assumption that the main verb is the verbal element which carries \( \phi \)-features (see chapter II:fn. 22), verb movement constructions such as the one represented in (33) pose a problem similar to the one posed by (32). Although in Chomsky’s (1995:chap. 4) theory both links of the chain \( \text{CH}_2 = ([V+V], [V+V]) \) can have the \( \phi \)-features checked because the upper link enters into a checking relation with the \( \phi \)-features of the subject, the \( \phi \)-features of the lower link of \( \text{CH}_1 = (V, V) \) remain unchecked. Given that these features are \([-\text{-interpretable} \), they should cause the derivation to crash at LF if they are not eliminated. Since the Chain Condition only applies to \( A \)-chains, elimination of the unchecked \( \phi \)-features of the lowest copy of \( V \) in (33) must then follow from something other than the Chain Condition.

Notice in addition that the tail of the chain \( \text{CH}_2 = ([V+V], [V+V]) \) in (33) is different from the head of the chain \( \text{CH}_1 = (V, V) \); hence, the formal features of the intermediate copy of \( V \) cannot be erased when \( \text{CH}_2 \) is formed,
and no chain between the highest and the lowest copies of V can be formed along the lines of $\text{CH}_3 = (\text{John}_4, \text{John}_1)$ in (32). Thus, whatever the process is which eliminates the unchecked $\phi$-features of the lowest copy of V in (33), it must be allowed to do so even in the absence of a linked chain in the sense of Chomsky and Lasnik (1993:563) between the copy that has unchecked $\phi$-features and the copy which has them checked.

It is also not obvious that when the head of a nontrivial chain CH raises, the new chain involves the newly created position and the tail of CH, as proposed by Chomsky (1995:300). Some evidence regarding the anti-c-command restriction on parasitic gap licensing, which will be discussed in section IV.2.3, suggests that the Minimal Link Condition prevents one copy $C_1$ from forming a chain with a copy $C_3$ over a copy $C_2$ such that $C_2$ c-commands $C_3$ (see also section III.6.2.5 below). With this in mind, I will assume Chomsky and Lasnik's (1993:563) notions of chain formation and linked chains.

Finally, there appears to be no theoretical gain in incorporating the Chain Condition into the Minimalist framework. Other than creating the problem which the stipulated erasure of intermediate traces is intended to solve, the Chain Condition is largely redundant with Last Resort.

Let us consider an alternative approach to these issues.
III.4.1.2. An Alternative Approach: The Feature Uniformity Condition

In order for Full Interpretation to be satisfied, all that is required as far as chains are concerned is that the \([-\text{interpretable}\)] features of their links be invisible at LF. In principle, the same reasons which led Chomsky (1995:chap. 4, fn. 75) not to eliminate the semantic features of intermediate traces (see fn. 14) should also prevent the elimination of their \([+\text{interpretable}\)] formal features. Suppose then that at LF, all links of a chain must be uniform in terms of feature composition, as stated in (34):

\[
(34) \quad \text{Feature Uniformity Condition:}
\]

Given a chain \(\text{CH} = (\alpha_1, ..., \alpha_n)\), every \(\alpha_i \ (1 \leq i \leq n)\) must have the same set of features visible at LF.

Let us reconsider the chain \(\text{CH} = (\text{John, John})\) of the structure in (30b) repeated below in (35), under the assumption that traces are unaffected by the operations affecting heads of chains. In the covert component, both links of \(\text{CH}\) have the same semantic features and no phonological features. They differ, however, with respect to formal features: the lower link has categorial, Case- and \(\phi\)-features visible, whereas the upper link only has categorial and \(\phi\)-features visible, because its Case-feature was made invisible
by the checking operation it participated in. As it stands, the chain CH in (35) violates the Feature Uniformity Condition.

(35) [ John T [ was kissed John ] ]

There are two possible ways for the chain in (35) to become uniform: either the head of the chain somehow "gains" an unchecked Case-feature, or the Case-feature of the trace is eliminated. The first possibility can be easily discarded. Besides being noninclusive (see section II.3), gaining a [-interpretable] feature is at odds with one of the main purposes of the mapping from N to \( \lambda \), which is to eliminate [-interpretable] features (see section II.14). We can formalize the second possibility along the lines of (36), where deletion targets a single feature per application.

(36) Chain Uniformization:

Delete the minimal number of features of a nontrivial chain CH in order to allow its links to satisfy the Feature Uniformity Condition.

If Chain Uniformization deletes the Case-feature of the tail of the chain \( CH = (John, John) \) in (35), CH complies with the Feature Uniformity Condition. Furthermore, since no [-interpretable] feature survives at LF, Full
Interpretation is satisfied and the derivation converges. Notice that if Chain Uniformization had deleted all the semantic features of CH in addition to the unchecked Case-feature, CH would also satisfy the Feature Uniformity Condition. This undesirable result does not arise, however, because, as stated in (36), Chain Uniformization resorts to the fewest possible applications of deletion. If a single application of deletion allows CH to comply with the Feature Uniformity Condition, economy considerations block further applications.\(^{16}\)

Let us now reconsider the chain \(CH_4 = (\text{John}_4, \text{John}_3, \text{John}_2, \text{John}_1)\) of (32d), repeated below in (37), assuming Chomsky and Lasnik's (1993:563) notion of linked chain. The links of \(CH_4\) differ in that the head of the chain does not have a Case-feature visible at LF, whereas the other three links do. Thus, in order for \(CH_4\) to satisfy the Feature Uniformity Condition, Chain Uniformization employs three applications of deletion, each targeting an unchecked Case-feature. After these operations, \(CH_4\) satisfies not only the Feature Uniformity Condition, but also Full Interpretation, because no [-interpretable] feature is visible at LF; hence, the derivation in (37) converges.

\(^{16}\) Actually, it is not necessary to specify that Chain Uniformization deletes as few features as possible. We may simply specify that Chain Uniformization deletes features in order for the Feature Uniformity Condition to be complied with (each application of deletion targeting a single feature), and leave the number of features to be deleted to be established by economy considerations concerning the number of applications of the suboperation of deletion. For purposes of exposition, I will make use of the formulation of Chain Uniformization given in (36).
\[ (37) \quad [\text{John}_4 \ T \ \text{seems} \ [\text{John}_3 \ \text{to be likely} \ [\text{John}_2 \ \text{to be kissed} \ [\text{John}_1 ]]]] ] \]

Consider now the chain \( CH = (\text{Bill}, \text{Bill}) \) in (38b). In this instance, both links have the same feature composition at LF. The D-feature of the upper copy enters into a checking relation with the infinitival \( T \), but since this feature remains unaffected by the checking operation, the two copies remain identical at LF. Although satisfying the Feature Uniformity Condition, this derivation does not satisfy Full Interpretation and crashes at LF, because both copies in (38b) have an unchecked Case-feature.\(^{17}\)

\[ (38) \]

a. *Bill to be kissed is surprising.

b. [ [ Bill to [ be kissed Bill ] ] is surprising ]

If Chain Uniformization had deleted the unchecked Case-features, \( CH \) would also satisfy the Feature Uniformity Condition, but the derivation in (38b) would be incorrectly allowed to converge, because Full Interpretation would be met. Again, this incorrect result does not arise, because Chain Uniformization does not apply to chains which are already uniform with respect to feature composition. The important thing to keep in mind is that,\(^{17}\)

---

\(^{17}\) The derivation in (38b) should also crash because the infinitival \( T \) did not check its null Case-feature, which can only enter into a checking relation with the Case-feature of PRO (see Chomsky and Lasnik (1993:561) and Martin(1992)).
as stated in (36), deletion of [\-interpretable]) features is triggered by the Feature Uniformity Condition, not by Full Interpretation at LF. This is a natural assumption to make; if Full Interpretation at LF could trigger deletion of [\-interpretable] features, no movement operation would ever be necessary.

Given this analysis of (38b), it seems that we cannot account for the required elimination of the $\phi$-features of the lowest copy of the main verb in (33), repeated in (38a). Notice that a linked chain like CH$_3$ in (38c) cannot be formed, because the tail of CH$_2$ is different from the head of CH$_1$. Suppose that CH$_1$ is inspected with respect the Feature Uniformity Condition before CH$_2$. CH$_1$ meets this condition trivially: since no [\-interpretable] feature of the head of the chain has been checked, the two copies are identical in terms of feature composition.$^{18}$ CH$_2$, on the other hand, does not satisfy the Feature Uniformity Condition: whereas the $\phi$-features of the upper copy are made invisible by being checked against the subject in the Spec of T, the $\phi$-features of the lower copy remain unchecked and, therefore, visible at LF. Chain Uniformization is then required to delete the $\phi$-features of the lower copy, rendering CH$_2$ uniform. This derivation will not converge, however, because the $\phi$-features of the tail of CH$_1$ will survive at LF, violating Full

---

$^{18}$ The light verb presumably has a strong V-feature (see chapter II:fn. 20), which is checked by the categorical feature of the head of CH$_1$. Since categorical features are [+interpretable], the upper copy remains unaffected by this checking relation. See sections III.6.2.4 and III.6.2.5 for further discussion.
Interpretation.

(39)  a. $[\text{TP} \text{ SU}_i [V+\nu+T] [\nu \text{ t}_i [\nu [V+\nu] [\nu \text{ VP} \text{ OB}]]]]$

b. $\text{CH}_1 = (V, V)$

c. $\text{CH}_2 = ([V+\nu], [V+\nu])$

d. $^*\text{CH}_3 = ([V+\nu], [V+\nu], V)$

If the derivation discussed above were the only one available for (39a), we would not be able to account for why a presumable violation of Full Interpretation would consistently yield well-formed results. However, there is an alternative convergent derivation of (39a) yielding the expected results. Suppose that $\text{CH}_2$ is inspected with respect to the Feature Uniformity Condition prior to $\text{CH}_1$. As before, Chain Uniformization will delete the $\phi$-features of the tail of $\text{CH}_2$, which are actually features of $V$ (see chapter II:fn. 22). After this operation, $\text{CH}_1$ is no longer uniform: its tail still has $\phi$-features visible at LF, whereas its head does not. Hence, Chain Uniformization deletes the $\phi$-features of the tail of $\text{CH}_1$, allowing it to satisfy the Feature Uniformity Condition and Full Interpretation to be met.\(^{19}\)

\(^{19}\) Notice that we need not impose a specific order for chains to be inspected with respect to feature composition. In instances such as (39), at least one sequence of applications of Chain Uniformization will allow the derivation to converge, which is enough for the C-I interface to assign the relevant interpretation to the structure obtained. If the range of possibilities for examining chain uniformity should be
The approach outlined above has several advantages over the one proposed by Chomsky (1995:chap. 4). First, it attempts to follow Minimalist guidelines as closely as possible by deriving the elimination of [-interpretable] features of traces from Full Interpretation and some plausible conjectures about the feature composition of chains. Hence, no stipulation requiring the elimination of [+interpretable] features of intermediate traces is resorted to. This in turn has the welcome result of being consistent with "reconstruction" in intermediate traces (see Barss (1986), Lebeaux (1991), Aoun and Li (1993), and Hornstein (1995), among others). Second, it does not rely on the Chain Condition, which has problems of its own (see section III.4.1.1). Finally, it keeps checking theory as simple as possible: (i) [-interpretable] features are not allowed to enter into more than one checking relation; and (ii) a checking relation can only be established by elements in a checking configuration; hence, both conditions would prevent the matrix T and the object in (35), for example, from entering into a checking relation.20

Sections III.4.1.3 and III.4.1.4 show how the proposed elimination of [-interpretable] features of chain links traces based on the Feature Uniformity restricted due to considerations regarding computational complexity, all that is necessary is that chains be inspected in a top-down fashion. For concreteness, I will assume that chains may be inspected in any order.

20 In sections III.6.2.3 and III.6.3 below, I will discuss two more considerations, having to do with linearization and strong features, which argue against Chomsky's (1995:chap. 4, fn. 12) proposal that when the head of a two-membered chain is affected by a checking operation, its trace is affected as well.
Condition accounts for wh-movement. In turn, section III.1.4.1.5 provides empirical evidence against Chomsky’s (1995:chap. 4) proposal that intermediate traces of A-movement are inaccessible to the computational system because their formal features are erased.

### III.4.1.3. Checking Features Against Traces

Consider the sentence in (40a), whose structure is represented in (40b).

\[\begin{align*}
\text{(40)} & \quad \text{a. Who do you think works hard?} \\
& \quad \text{b. } [\text{CP who}_3 \text{ do you think } [\text{TP who}_2 T [\nu P \text{ who}_1 \text{ works hard } ] ] ]
\end{align*}\]

In (40), who moves from the Spec of the light verb to the embedded subject position, allowing the Case- and the strong D-feature of T to be checked; from this position, it moves to Spec of CP to check the strong wh-feature of the interrogative complementizer. The $\phi$-features of the traces of who, being [+interpretable], are not deleted by Chain Uniformization. Thus, FF(works) in (40) can adjoin to T and check its $\phi$-features against the trace of who left in the Spec of TP, allowing the derivation to converge at LF.

---

21 I disregard the potential trace in the Spec of the embedded CP, which is irrelevant for the present discussion.
The approach developed here may also derive the standard assumption that in cases of successive cyclic movement such as (40b), it is the head of the chain which moves (see section III.6.2.5 for further discussion). Take the linked chains CH₁ = (who₁, who₂, who₁) and CH₂ = (who₂, who₂, who₁), which are formed by copying either the instance of who in the Spec of the embedded light verb (who₃ = who₁), or the one in the Spec of the embedded T (who₃ = who₂), as indicated by the subscripts. As they stand, neither chain satisfies the Feature Uniformity Condition: the highest and lowest links of CH₁, and the lowest link of CH₂, have unchecked Case-features, while the remaining links do not. Deletion is thus required to apply twice in CH₁ and only once in CH₂. Thus, if these two derivations are to be compared for economy purposes, the derivation forming the chain CH₂ rules out the one forming CH₁ because it employs fewer applications of deletion (see sections IV.2.1 and IV.2.3 for further discussion).

III.4.1.4. Movement of Traces

Consider now constructions such as (41) in English, where a wh-phrase moves from the object position to check the strong wh-feature of the complementizer Q, forming the chain CH₁ = (what, t).

(41)  [what, did you buy t₁]
Assuming that traces are accessible to the computational system, which is the null hypothesis under the copy theory of movement, the least stipulative analysis of (41) must then involve raising of the formal features of the copy of what in object position for Case checking, as represented in (42).^{22,23}

(42) \[ \text{[ what did you, } _\nu \text{ t; } _\nu \text{ FF(what)+V+} _\nu \text{ V what/FF(what) ] ] ] ]

In order for (42) to satisfy Full Interpretation, no [-interpretable] feature can be visible at LF. Let us focus on the chains CH\(_1\) = \(\text{(what, what)}\) and CH\(_2\) = (FF(what), FF(what)) and see how their unchecked Case-features are eliminated. In order for CH\(_2\) to satisfy the Feature Uniformity Condition, Chain Uniformization deletes the Case-feature of the lower copy of FF(what), which is one of the three sets of features which form the lexical item what (see section II.3 and fn. 22). After this operation takes place, CH\(_1\) now becomes non-uniform, because its upper link has a Case-feature visible at LF, whereas its lower link does not. Chain Uniformization then applies to CH\(_1\) and deletes the Case-feature of its upper link. Since both chains now satisfy the Feature Uniformity Condition and Full Interpretation, the

^{22} The representation what/FF(what) is meant to show that the link of CH\(_1\) have both semantic and formal features (the phonological features have been stripped away by Spell-Out), whereas the links of CH\(_2\) have only formal features. The lower link of CH\(_2\) is a subset of the set of features of the lower link of CH\(_1\).

^{23} See the discussion of covert movement of objects in section III.4.1.5.3.
derivation converges.

III.4.1.5. Traces and the Minimal Link Condition

In this section I examine the constructions which Chomsky (1995:chap. 4) takes to show that traces of A-movement do not count for the purposes of the Minimal Link Condition, repeated below in (43) (see discussion in sections II.7.4 and II.10.6):

(43)  Minimal Link Condition:

Given the syntactic objects \( \alpha \) and \( \beta \) such that \( \alpha \) c-commands \( \beta \) and a certain feature \( F \) of \( \alpha \) enters into a checking relation, \( \alpha \) can form a chain with \( \beta \) if there is no element \( \gamma \) such that:

(i) \( \gamma \) has a feature \( F' \) which is of the same type as the feature \( F \) of \( \alpha \); and

(ii) \( \gamma \) is closer to \( \alpha \) than \( \beta \) is.
(44) *Closeness:*

Where $\gamma$ c-commands $\alpha$ and $\alpha$ c-commands $\beta$, $\alpha$ is closer to $\gamma$ than $\beta$ is iff:

(i) $\alpha$ c-commands $\beta$; and

(ii) $\alpha$ is not in the same minimal domain as $\gamma$ or $\beta$.

III.4.1.5.1. Expletive Constructions

Let us consider the structure in (45) below. The expletive has been inserted in the most embedded clause, checking the strong D-feature of $T$ (see section II.15), and raises twice, checking the strong D-feature of the intermediate and the matrix $T$ heads. At LF, FF(*many books*) adjoins to \[ FF(seem)+T \], checking its Case against $T$, and checking the $\phi$-features of FF(*seem*) (see section II.14.3.2). Notice that neither trace of *there* prevents movement of FF(*many books*); they only have a D-feature, and no checking relation between FF(*many books*) and \[ FF(seem)+T \] involves D-features. Thus, (45) cannot tell us anything about the possibility of traces being accessible to the computational system, given that the traces of *there* could not enter into a checking relation with a sublabel of the matrix $T$ head.
(45) \[
\text{[ there, seem [t, to be believed [ t, to be many books out of place ]]]}
\]

Chomsky (1995:302) takes this to be the only instance in which a trace of an expletive could induce a Minimal Link Condition violation:

This is (...) the only kind of construction in which the problem of attracting expletive trace could arise. We therefore conclude that the trace of an expletive does not enter into the operation Attract/Move; it is immobile and cannot bar raising. (Chomsky (1995:302))

There are, however, other constructions in which not only does the issue arise, but a trace of an expletive is shown to induce a Minimal Link Condition violation (see section II.14.1). Consider, for instance, a derivational step in which a structure such as (46) has been assembled. Discussing a structure such as (46), Chomsky (1995:295) claims that no convergent derivation arises from inserting the expletive it in the subject position of the embedded clause. Raising of John to the matrix clause to check the strong D-feature of T, as shown in (47a), violates the Minimal Link Condition, because the expletive is a closer element which can check this feature. Recall that in Chomsky's (1995:chap. 4) system, the categorial feature of an expletive is deleted, but not erased (see section II.14.1). Hence,
the sentence in (47b) is unacceptable.

(46) \([TP \; T \text{ seems } [CP \; \text{ that it was told John } [CP \; \text{ that he was fired } ] ]]\)

(47) a. \([TP \; \text{ John } \; T \text{ seems } [CP \; \text{ that it was told } t_i \; [CP \; \text{ that he was fired } ] ]]\)

b. \(*\text{John seems that it was told that he was fired.}\)

If the expletive of (46) raises to the matrix subject position in compliance with the Minimal Link Condition, as illustrated in (48a) below, it will be able to check the strong feature of T. However, it will not be able to check the Case-feature of T or the $\phi$-features of $[FF(seems)+T]$; since the Case and $\phi$-features of $it$ are [-interpretable], they were erased in the previous checking relation in the embedded subject position. Given that the Case-features of T and John remain unchecked at LF, the derivation violates Full Interpretation and crashes; hence the unacceptability of (48b).

(48) a. \([TP \; it_i \; T \text{ seems } [CP \; \text{ that } t_i \; \text{ was told John } [CP \; \text{ that he was fired } ] ]]\)

b. \(*\text{It seems that was told John that he was fired.}\)

The only convergent alternative for the derivation before the expletive is inserted in the embedded clause, therefore, is to move John to the embedded subject position and insert the expletive in the subject position, as
illustrated in (49a), yielding the acceptable sentence in (49b). Notice that the option of moving John to the embedded subject position, which violates Procrastinate, is not to be compared with the option of inserting it for purposes of economy, because the latter does not lead to a convergent derivation (see section II.15).

(49) a. [TP it T seems [CP that John was told it [CP that he was fired ] ] ]
    
    b. It seems that John was told that he was fired.

There is one problem with Chomsky's analysis of the derivation of (48b), however. As mentioned above, when the expletive enters into a checking relation with the embedded T in (46), its Case- and $\phi$-features are deleted and erased, in accordance with Chomsky's (1995:chap. 4) system. Its D-feature, on the other hand, is just deleted; if it were erased, the whole category would be erased, violating the condition that terms cannot erase (see Chomsky (1995:281)). Although invisible at LF, the deleted D-feature of the expletive is accessible to the computational system; it then raises and checks the strong D-feature of the matrix T, as illustrated in (48a). At LF, FF(John) should now be able to raise and adjoin to [ FF(seems)$+T$ ] checking the Case-feature of T and the $\phi$-features of FF(seems). The trace of the expletive cannot block this movement because its Case- and $\phi$-features were
erased. According to this analysis, the derivation in (48a) converges and therefore should be compared with (49a) for purposes of economy. At the point where the strong feature of the intermediate T is checked, (49a) violates Procrastinate, but (48a) does not; hence (48a) should block (49a) (see section II.15).

Chomsky’s (1995:chap. 4) analysis of checking in terms of deletion and erasure therefore wrongly predicts that (49b) should be unacceptable, and that (48b) should be acceptable. A plausible way to prevent this undesirable result is to allow the trace of it in (48a) to block raising of FF(John) at LF. If such an approach proves successful in accounting for the unacceptability of (47b) and (48b) and for the acceptability of (49b), we will have evidence against Chomsky’s (1995:chap. 4) proposal that traces are inaccessible to the computational system.

The system I have adopted here provides a straightforward account of the sentences in (47)-(49). Recall that under the system assumed here, a checking operation merely renders a [-interpretable] feature invisible at LF; no further process of erasure is invoked by the checking procedure. Thus, if the expletive is inserted in the subject position of the structure in (46), repeated below in (50a), its Case- and ϕ-features will be rendered invisible at LF by entering into a checking relation with [ was+T ], but they remain present in the structure. Its D-feature however remains unaffected after checking the strong feature of the intermediate T head and is still able to
check the strong feature of the matrix T, when the expletive moves to the matrix subject position, as shown in (48a), repeated below in (50b).\(^{24}\)

\[
\begin{align*}
(50) & \quad \text{a. [TP T seems [CP that it was told John [CP that he was fired ] ]]}
\end{align*}
\]

\[
\begin{align*}
& \quad \text{b. [TP it T seems [CP that it was told John [CP that he was fired ] ]]}
\end{align*}
\]

In the covert component, the Case- and \(\phi\)-features of [ FF(seems)+T ] need to be checked in order for the derivation to converge. Movement of FF(John) to check them violates the Minimal Link Condition, because the trace of the expletive, which also has Case- and \(\phi\)-features, is closer to [ FF(seems)+T ] than John (see sections II.7.4 and II.10.6). Movement of the formal features of the expletive, on the other hand, violate Last Resort because their relevant features are already checked and can establish no checking relation with [ FF(seems)+T ]. Hence, no convergent derivation can obtain if the expletive is inserted in the intermediate subject position; this in turn prevents (50b) from being compared for economy purposes with (49a), repeated below in (51), yielding the correct results.

\[
\begin{align*}
(51) & \quad \text{[TP it T seems [CP that John was told it [CP that he was fired ] ]]}
\end{align*}
\]

\(^{24}\) Recall that categorical features were taken to be always [+interpretable] (see sections II.14 and III.6.2.4).
We have seen that upon close inspection, Chomsky's (1995:chap. 4) checking theory in terms of deletion and erasure makes wrong predictions about the status of the sentences in (48b) and (49b). Furthermore, I argued that the contrast between these two sentences can be accounted for if traces are accessible to the computational system and therefore are able to block movement across them, yielding Minimal Link Condition violations. As pointed out at the outset, this is a sound result from a conceptual point of view. Once the copy theory of movement is adopted, there is no principled reason for us to expect lower copies of a chain to be defective in not being computed by the Minimal Link Condition.

III.4.1.5.2. Raising Constructions

Another piece of data which Chomsky (1995:305) takes as evidence that traces do not block movement involves raising constructions such as the ones illustrated by the French sentences in (52).
Raising of Jean to check the strong feature of the matrix T in (52a) is presumably ruled out by the Minimal Link Condition: Marie could check the strong feature of T and it c-commands the embedded subject. Assuming that the clitic in (52b) adjoins to the T head, Jean is allowed to cross both the clitic and its trace, without yielding a Minimal Link Condition violation: the clitic does not prevent this movement because it is in the same minimal domain as the moved subject (see sections II.7.4 and II.10.6), and the trace of the clitic does not block this movement because its formal features have presumably been erased.

Before examining Chomsky's analysis of the contrast in (52), let us consider the English counterpart of (52a) given in (53). The same reasons which rule out (52a) should in principle apply to (53). Chomsky (1994, MIT Fall lectures) suggests that raising of the embedded subject in instances such as (53) may be accounted for if inherently Case-marked elements do not count for the purposes of the Minimal Link Condition. Thus, if Mary
receives an inherent Case from *seems* in (53), it will not be able to block raising of an embedded element to the subject position and the derivation can converge, as desired.

\[(53) \quad [ \text{John}_{i} \text{ seems to Mary } [ t_{i} \text{ to be talented } ] ]\]

If (53) can be analyzed along these lines, we are then not forced to take (52b) as evidence that traces do not block movement. Suppose that raising verbs in French assign inherent Case to dative clitics. If so, the trace of the clitic would not prevent movement in (52b) for the same reasons that Mary does not in (53). Evidence that this hypothesis might be on the right track comes from the unacceptability of the sentence in (54). In (54), à Marie moves to the subject position to check the strong feature of T; its Case-feature is not of the type which can check the Case-feature of T, but it can check the Case-feature of the preposition. So, the matrix T and Jean in (54) still have their Case-features to be checked. If traces do not block movement due to the erasure of their formal features, as proposed by Chomsky (1995:chap. 4), the trace of à Marie in (54) should not prevent movement of FF(Jean) to check its Case-feature against T, yielding a convergent derivation. Hence, the contrast between (52b) and (54) cannot be explained under Chomsky’s approach.
Under the checking theory I have adopted here, the Case-feature of the trace of à Marie in (54), even if checked, prevents movement of FF(Jean) for Case reasons, in the same way the D-feature of à Marie in (52a) prevents Jean from checking the strong feature of the matrix T head. The absence of a Minimal Link Condition violation in (52b) must then be a reflex of the properties of inherent Case-marked elements, rather than a reflex of the properties of traces.

III.4.1.5.3. Covert Movement

The last piece of evidence Chomsky (1995:chap. 4) presents in favor of the idea that traces of A-movement do not block movement is related to how the T head should look at LF if we assume that "interpretive operations at the interface should be as simple as possible", which in turn demands that "forms that reach the LF level must be as similar as typological variation permits — unique, if that is possible" (Chomsky 1995:359).25

25 However, as pointed out in chapter II:fn. 3, Chomsky (1995:291) proposes that structures involving wh-phrases are different at LF in languages with and without overt wh-movement, requiring different interpretive devices at the C-I system. For a unified analysis of these two types of languages, see Nunes (1995a).
Chomsky begins the discussion of the range of possibilities for the structure of a T head at LF by examining two types of verb movement languages: the ones in which both subject and object may move overtly, such as Icelandic, and the ones in which only the subject moves overtly, such as French. The T head of the former is as simple as (55), where VB is the verbal complex formed by adjoining the main verb to the light verb in a transitive construction.

\[(55)\]

```
T
  
VB    T
```

In languages such as French, FF(OB) in principle has the option of adjoining to the tail or the head of the VB chain, or adjoining to T itself, since the Case-feature of VB is also a sublabel of T (see section II.14). Chomsky (1995:304) excludes adjunction of FF(OB) to the tail of the VB chain, because traces supposedly do not enter into checking relations (but see section III.4.1.3). The choice between adjunction to the head of the VB chain or to T is then determined by the "poverty-of-interpretation conjecture", according to which the syntax of the T head should vary as little as possible from language to language. By comparison with the structure in (55), the structure of the T head in French is then taken to be as in (56). Notice that FF(OB) crosses the trace of the subject in the Spec of the light verb when it
moves to T. Since traces are said to have no formal features, this movement is taken to comply with the Minimal Link Condition.

\[ (56) \]

\[
\begin{array}{c}
T \\
\frac{FF(OB)}{\text{VB}} \\
\frac{T}{T}
\end{array}
\]

Given the structure in (56), the poverty-of-interpretation conjecture could in principle predict that in languages in which both subject and object remain in situ (if such languages exist; see Jonas and Bobaljik (1993:74) and Chomsky (1995:chap. 4, fn. 80)), FF(OB) adjoins to T as in (56). This movement would however cross the subject in the Spec of the light verb, yielding a Minimal Link Condition violation. If FF(SU) adjoins to T first, on the other hand, movement of FF(OB) is taken to comply with the Minimal Link Condition: FF(SU) is in the same minimal domain as FF(OB) and the trace of FF(SU) is to be disregarded. At LF, the T head of this type of language will thus be as in (57).
The discussion above is based on the premise that traces do not have formal features and, therefore, are irrelevant for the Minimal Link Condition. As we saw in sections III.4.1.5.1 and III.4.1.5.2, expletive and raising constructions strongly suggest that this is not the case. If so, the Case-feature of the trace of the subject in the Spec of the light verb may be taken to block movement of FF(OB) to T in both (56) and (57).

The analysis pursued here, which takes traces to be as accessible to the computational system as heads of chains, however, is not incompatible with Chomsky's poverty-of-interpretation conjecture. According to this analysis, there are only two types of T heads across languages: languages with overt subject movement have a T head as in (55), and languages without overt subject movement have FF(SU) adjoined to (55). As far as the structure of the VP shell headed by the light verb is concerned, it also has two forms: in languages with overt object movement, OB is in the Spec of VB; in languages without overt object shift, FF(OB) is adjoined to (the trace of) VB.

This approach has the advantage of keeping movement as "short" as
possible. Consider languages with no verb movement and no object shift, such as English. In Chomsky's (1995:361) analysis, the LF structure of the T head in these languages is as in (56), with FF(VB) instead of VB. Notice that the alleged impossibility of FF(OB) adjoining to the trace of VB does not necessarily arise in English; the checking of the Case-feature of FF(OB) could proceed cyclically, with FF(OB) adjoining to VB before VB moved to T. The poverty-of-interpretation conjecture is thus the only motivation for the longer and noncyclic movement of FF(OB) to T to be preferred over the shorter and cyclic movement to VB. In the analysis pursued here, the shorter and cyclic movement is the only option available, because the long movement violates the Minimal Link Condition.26

III.4.2. Linearization

With Kayne's (1994) influential hypothesis that linear order is determined by hierarchical structure, an interesting question arises: are traces computed for the purposes of linearization? The question is even more intriguing if one assumes that a trace is an exact copy of the moved element. Below I point out some conceptual problems for the claim that

---

26 Depending on how pied-piping of formal features works in the covert component, the approach pursued here is also consistent with the formal features of the object adjoining to the verb before the verb raises, and then being pied-piped when the formal features of the verb raise to T. Perhaps this is the way to implement Hornstein's (1995:chap. 8) non-QR analysis of quantifier scope interaction under a Move-F approach.
traces should be disregarded as far as linearization is concerned (see Kayne (1994) and Chomsky (1994, 1995: chap. 4)). I delay a full discussion of the issue until section III.6, where I show that the deletion of chain links discussed in section III.4.3 is triggered by linearization considerations.

III.4.2.1. Kayne (1994)

Kayne (1994: 10) briefly mentions the question of linearization of traces under the copy theory of movement when discussing the well-formedness of the phrase-marker in (58) below with respect to the LCA, as opposed to the one in (59). Recall that according to him, (58) is a well formed phrase-marker because V asymmetrically c-commands N and, therefore, see precedes John; in (59), on the other hand, V and NP c-command each other and, therefore, no linear order between see and John can be established (see section II.17).

(58)  
```
    K
   /\  
  J  VP
   /\   
  V NP
     /\  
    see N
       /\  
      John
```
(59)  

Assuming the distinction between phrase-markers such as (58) and (59), Kayne considers two possibilities as to how structures containing traces comply with the LCA:

The question arises of what happens if *John (or any phrase) is moved. If the result of movement is that DP (or NP) dominates just a trace, then *see and that trace will not be ordered at all with respect to one another, since the internal structure of DP (or NP), which ensured antisymmetry, will have been lost. This might conceivably be a tolerable consequence, since traces are in any event not visible. It is notable, though, that this question does not arise if movement transformations leave a copy, rather than a trace. (Kayne 1994:chap. 2, fn. 3)

Kayne's suggestion that the "invisibility" of traces at PF exempts them from being subject to the LCA is indeed at odds with his conceptual argument that the LCA applies to all syntactic representations, including LF (see Kayne 1994:sec. 5.2). Given the relevance of X'-Theory for, say, S-Structure and LF, and his claim that the LCA is the source of the major
properties of standard X'-Theory, Kayne (p. 49) concludes that the LCA underlies the entire set of syntactic representations. This entails, however, that S-Structure and LF would not allow a structure containing a head and the trace of its complement, for instance, for the same reason a structure such as (59) should not be permitted: it violates the LCA.

On the other hand, as Kayne observes, there seems to be no problem under the copy theory of movement in linearizing a trace in complement position with respect to the head it is a sister of, provided that the trace has internal structure. In other words, traces do not differ from unmoved elements with respect to the well-formedness conditions on phrase-markers imposed by the LCA.

Although this approach solves the problem of the linearization between a head and the trace of a complement with internal structure, it faces some problems regarding the linearization of the links of a chain with respect to one another. In section III.6 below, the issue of how nontrivial chains comply with the LCA is discussed in detail. For the purposes of this section, it suffices to note that if the links of a chain are in a sense the same element (they form a discontinuous object), any intervening material between two links of a given chain would asymmetrically c-command and be asymmetrically c-commanded by the same element. This state of affairs, however, is not compatible with the LCA and no linear order can obtain.
III.4.2.2. Chomsky (1994)

Chomsky (1994:28) suggests that "there is no reason for the LCA to order an element that will disappear at PF, for example, a trace." Although this hypothesis makes the plausible assumption that the (still to be explained) deletion of traces takes place after Linearize, it tacitly assumes that the output of Linearize may be a nonlinearized (perhaps mixed) object. However, if the phonological rules which apply after Linearize operate with a sequence of $X^0$ elements rather than a phrase-structure, as seems plausible, no PF object would be formed if Linearize failed to yield a sequence of $X^0$ items.

Furthermore, the logic of Chomsky's suggestion confers too much power to global computations. The linearization of a given element at a given derivational step is taken to be contingent on different operations targeting that element later on in the derivation. For instance, if this logic were to extend to morphological operations, which are taken to precede Linearize (see Chomsky (1994:28, 31)), we could ask why Morphology should not ignore traces, allowing contraction to proceed over them, given that they

---

27 In Chomsky (1994, 1995:chp. 4), the term LCA is used to refer both to the Linear Correspondence Axiom and to the mapping operation that conforms to this axiom, as becomes clear when it is suggested that the LCA may delete traces (see Chomsky (1994:28, 31)). I will avoid this ambiguity and use the term Linearize for the operation that maps a phrase-structure into a linear order of $X^0$ elements in accordance with the LCA.

28 Recall that a derivation is canceled if the pair $(\pi, \lambda)$ is not formed (see Chomsky (1995:225-226) and section II.4).
will be deleted later on in the derivation (after Linearize). Clearly, this would yield wrong results with respect to wanna-contraction over wh-traces (cf. (71b)).

Another suggestion by Chomsky (1994:28) is that "the LCA may (but need not) delete traces". This is intended to avoid the problem of linearizing the trace of a complement without internal structure with respect to the head it is a sister of (see section II.17). Notice, however, that there would be a redundancy under this approach: both Linearize and the phonological rules that apply to the output of Linearize can delete traces. In addition, although a motivation for the deletion of traces is provided when it is executed by Linearize, the motivation for the deletion after Linearize is still missing. The Minimalist assumption that the language faculty is an nonredundant system, however, would lead us to expect a unique motivation for deletion of traces and a unique operation to execute it.

III.4.3. Phonetic Realization

The most obvious difference between heads of chains and traces is their phonetic realization: whereas heads of chains are always phonetically realized if they have phonological features, traces are never pronounced.²⁹,³⁰

²⁹ Of course, the head of a chain is not phonetically realized if it does not have phonological features, as in the case of a chain headed by PRO, for example.

³⁰ See section III.7 for some apparent instances of phonetically realized traces.
Thus, any analysis which adopts some version of the copy theory of movement must explain why the derivation sketched in (60) yields the PF output associated with the sentence in (61a), but not with the one associated with (61b) or (61c):

(60)  

a. \[ \text{TP} \quad T \left[ \text{VP} \left[ \text{VP} \text{was} \right] \left[ \text{VP} \text{kissed John} \right] \right] \]  

b. COPY: \[ \text{TP} \quad T \left[ \text{VP was} \right] \left[ \text{VP kissed John} \right] \]  

c. MERGE: \[ \text{TP John} \left[ T \right. \left. \left[ \text{VP was} \right] \left[ \text{VP kissed John} \right] \right] \]  

(61)  

a. John was kissed.  


c. *John was kissed John.  

At first sight, the unacceptability of (61b) has an obvious explanation: the strong feature of T has not been checked, causing the derivation to crash at PF, because the phonological component is unable to deal with unchecked strong features (see section II.6.2). Although this is a straightforward account of one potential derivation of (61b), it does not extend to the derivation outlined in (60). In (60), a copy of John merges with the structure in (60a), becoming the specifier of T and thus being able to check the strong feature of T. If the copy checks this strong feature, the unacceptability of (61b) cannot be reduced to a problem of strong feature checking (see section III.4.3.2 for further discussion).
In turn, the unacceptability of the sentence (61c) resulting from the derivation in (60) raises an additional puzzle. The derivation of (61a) and (61b) from (60c) presumably involves an operation eliminating one of the copies of John, whereas no such operation is invoked in (61c). Thus, were the derivations of (61a) and (61b), on the one hand, and the derivation of (61c), on the other, to be compared for economy purposes, the derivation of (61c) should outrank the other two because it involves fewer operations, thus being more economical. Since (61c) is unacceptable, its derivation from (60c) must either crash or be canceled, thereby being irrelevant for the computation of economy (see sections II.2 and II.15).

These are the challenges that any analysis assuming the copy theory of movement has to face, as far as phonetic realization of chain links is concerned. I will postpone the discussion of my own proposal until section III.6. In the next sections I examine two proposals put forward to deal with these issues.

**III.4.3.1. Ellipsis and Deletion of Traces**

Chomsky (1993:35) suggests that deletion of traces in the phonological component is an obligatory variant of a more general process that converts the structure in (62a), for instance, into the sentence in (62b) in the
phonological component, by deleting E. Thus, in order to evaluate Chomsky’s suggestion concerning deletion of traces, we first need to establish if (62a) is actually related to (62b) by a derivational process.

(62)  
a. John said that he was looking for a cat, and so did Bill 
[Ε say that he was looking for a cat ]  
b. John said that he was looking for a cat, and so did Bill.

The main argument for (62a) to be transformationally related to (62b) is that both constructions are subject to a "parallelism constraint", which presumably holds at LF. For instance, if he in the first conjunct of either of the sentences of (62) is taken to refer to Tom and a cat is understood nonspecifically, the same interpretation will obtain in the second conjunct.

This argument is not very compelling, however, if the inclusiveness condition on the mapping from N to λ is assumed (see section II.3). If all the features of a given LF structure are brought into the derivation by the lexical items that feed it, the argument for the derivational relation between (62a) and (62b) based on parallelism would require that both specificity and reference be lexical features. Although it is possible that specificity is a lexical feature, many things are less clear as far as reference is concerned. Given the

---

31 After all, specificity appears to be associated with morphological case in some languages (see Enç (1991), among others).
relevance of intentions and beliefs for the determination of reference, it seems more appropriate to take reference to be in the realm of the C-I System.

If so, the referential properties of (62a) and (62b) cannot be used as an argument for their derivational relation. The hypothesis that the referential properties (and perhaps all parallelism restrictions) are post-LF phenomena is supported by the fact that the parallelism constraint also obtains across sentences, as illustrated by the discourse sequence in (63).

(63) A: — John said that he was looking for a cat.
    B: — So did Bill.

Another problem with a derivational approach to the pair of sentences in (62) concerns the generality of the proposed deletion operation. As pointed out by Chomsky and Lasnik (1993:565), given the pair in (62), it is not clear why the structures in (64), for instance, should not be derived from the ones in (65), which are however ill-formed.

(64) a. John said that he was looking for a cat, and Bill did too.
    b. John likes poetry, but not Bill.
(65)  a.  *John said that he was looking for a cat, and Bill did say that he was looking for a cat too.

       b.  *John likes poetry, but not Bill likes poetry.

Finally, the derivational approach to (62) requires unprecedented economy computations. Noting that the bracketed constituent in (62b) has a distinctive low-flat intonation, Chomsky and Lasnik (1993:564) propose that "the deletion rule (...) could say simply that material with this intonational property may optionally delete". Within the Minimalist framework, optionality must be due to the same derivational cost being ascribed to different options. However, deletion and lack of deletion clearly do not have the same derivational cost. Hence, Chomsky and Lasnik's proposal amounts to saying that at a certain point in the derivation, deletion is as costly as low-flat intonation. These options do not form a natural class, however. In the absence of independent evidence, it does not seem plausible to take deletion and low-flat intonation to be equally costly, if comparable at all.

Even disregarding these problems for a derivational relation between the sentences of (62), it is likely that deletion of traces is unrelated to ellipsis. First, traces are obligatorily deleted, as opposed to other potential "ellipsis material", which may be optionally deleted; a low-flat intonation of a trace does not make the pronunciation of the trace acceptable, as seen in (66b) with low-flat intonation on the trace of the subject. Notice that (66a)
is a well-formed structure; it is the structure that yields the sentence in (66c) after the trace of the subject is deleted. Thus, without an independent explanation for the ill-formedness of (66b) with low-flat intonation on the trace, Chomsky’s (1993) proposal actually amounts to saying that deletion of traces and ellipsis are different phenomena.

(66)  
   a. [that John said he was looking for a cat] is believed [that John said he was looking for a cat] by everyone  
   b. *That John said he was looking for a cat is believed that John said he was looking for a cat by everyone.  
   c. That John said he was looking for a cat is believed by everyone.  

There is another reason not to take deletion of traces to be a subcase of ellipsis. Deletion of traces operates under strict identity, whereas this need not be the case with ellipsis. Consider the initial numeration underlying (62a), represented in (67), and the initial numeration of (66a), represented in (68) (functional categories without phonological features are not represented).

(67) \{John₁, said₁, that₂, he₂, was₂, looking₂, for₂, a₂, cat₂, and₁, so₁, did₁, Bill₁, say₁\}
Whereas the repeated lexical items of (66a) are obtained by the copying operation (see section II.10.4) and are nondistinct in the initial numeration, as shown in (68), the doubled lexical items of (62a) are distinct in the initial numeration, as shown in (67). Deletion of traces thus operates with elements that are not distinguished in the initial numeration, while the alleged deletion in the mapping from (62a) to (62b) operates with elements that are morphologically identical, but distinctively specified in the numeration (e.g., John or even lexical items which are not identical at all (e.g., say is allegedly deleted in (62a) based on its relation to said).32

Chomsky (1995:252-253) takes the opposite view on the relation between ellipsis and deletion of traces, suggesting that ellipsis is a subcase of deletion of traces:

At some point in the derivation [from (26a) to (26b); JMN], the bracketed element must be marked as "subject to parallelism interpretation." Assume that this takes place before Spell-Out [footnote omitted; JMN]. The marking

---

32 The unacceptability of (65b), for instance, takes as the derivational source for (61b), also shows that a derivational analysis of ellipsis cannot be based on strict identity (see Lasnik (1994) and Uriagereka (1994, forthcoming: chap. 4) for further discussion.)
could be the removal of the distinctions indicated by numeration, in which case the bracketed element is in a certain sense nondistinct from the phrase it "copies" (the latter still marked by the numeration). Such a configuration might be interpreted at PF as assigning a copy intonation to the bracketed expression, and at LF as imposing the parallelism interpretations (...). Suppose that numeration markings on the copy are changed to those of the first conjunct instead of being deleted. Then the antecedent and its copy are strictly identical and constitute a chain, if a chain is understood as (constructed from) a pair of terms ($\alpha_1, \alpha_2$) that are identical in constitution. It will follow, then, that the copy deletes, by whatever mechanism deletes traces in the phonological component. At LF, the two kinds of constructions will be very similar, though not quite identical. (Chomsky 1995:252-253)

This suggestion attempts a unification of ellipsis and deletion of traces (although deletion of traces is still unexplained), but it ends up stressing their differences. First, ellipsis "chains" do not satisfy Last Resort or the c-command condition, which do apply to regular chains (see sections II.10.3 and II.14.2). Second, even if we assume that the operation which changes numeration markings to form ellipsis "chains" can be independently motivated, it is not obvious how this operation works in instances of ellipsis "chains" formed in discourse, as illustrated in (63).

To summarize, the discussion above shows that: (i) a derivational approach to ellipsis constructions such as (62b) from a structure such as (62a)
is problematic on several accounts; and (ii) even if these problems are overcome, it does not seem possible to subsume deletion of traces under ellipsis, nor does it seem possible to subsume ellipsis under deletion of traces.

III.4.3.2. Copying Phonological Features to the Head of the Chain

O'Neil and Groat (1995) propose an alternative model to the one laid out in Chomsky (1993), as far as the notions of Spell-Out and movement are concerned:

In our model, the derivation proceeds to create a single phrase-marker Kf, which is both Spelled-Out to the PF interface, and interpreted by the LF interpretive component. Thus, after Spell-Out, no further syntactic operations are applied to the phrase-marker. This single phrase-marker must therefore meet all the requirements of both the phonological and interpretive components, or it is an illegitimate representation, and the derivation crashes. The difference between "overt" and "covert" movement is not expressed in the timing of movement with respect to Spell-Out, but rather with respect to where a category is pronounced in its chain. Thus, in the simple case of a category forming a two-membered chain in the syntax, what is traditionally called "overt" movement amounts to pronouncing the head of the chain, while "covert" movement entails pronouncing the tail of
the chain. We postulate that chain formation, invariably a reflex of feature checking, copies phonological material to the head of the chain only when the features to be checked are strong. (O’Neil and Groat 1995:44)

An immediate question that this model raises concerns instances where a chain is involved in checking more than one strong feature, as illustrated by the chain headed by who in (69), for instance, which checks the strong feature of the embedded T and the strong feature of the matrix C. If the phonological realization of a chain link were inherently tied to its checking a strong feature, we should in principle expect (69) to be realized as (70a) and not as (70b).

(69) who\textsubscript{i} did you say [TP t\textsubscript{i} [VP t\textsubscript{i} saw Mary ] ]

(70) a. *Who did you say who saw Mary?  
b. Who did you say saw Mary?

A derivational approach to the movement of phonological features could perhaps account for the unacceptability of (70a). Thus, when who checks the strong feature of the embedded T in (70), its phonological features are placed in Spec of TP; in the next movement operation, the phonological

\[ \text{33 Thanks to Juan Uriagereka (p.c.) for bringing this potential solution to my attention.} \]
features of *who are then placed in the matrix Spec of CP, deriving the sentence in (70b). If that is so, however, the well-known wanna-contraction facts, illustrated by the sentences in (71), cannot be captured.

(71) a. Who do you want to/wanna dance with?
    b. Who do you want to/*wanna dance with Mary?

Consider the derivation of (71b), represented in (72), under the standard assumption that contraction takes place in the phonological component. The phonological features of *who are first placed in Spec of the embedded clause, where T has its strong D-feature checked; later on, the phonological features of *who move to the matrix Spec of CP, where the interrogative complementizer has its strong wh-feature checked. The sequence want to in (72) is then allowed to contract, because no phonological features intervene between these lexical items, yielding an incorrect result.

(72) who_i did you want [TP t_i to [VP t_i dance with Mary ] ]

Notice that we cannot ascribe the impossibility of contraction in (72) to the intervening formal or semantic features of the intermediate trace of who. Were that the case, the formal or semantic features of PRO in the structure in (73) should also block contraction, which is not the case, as seen
Consider now how Procrastinate is reinterpreted in this system. According to O'Neil and Groat (1995:45), "the stipulation, that a post-Spell-Out operation is cheaper than the equivalent pre-Spell-Out operation is replaced in our model by the stipulation that it is cheaper not to move phonetic features for any given movement operation; therefore they will not move unless forced to move by considerations of Greed".

At first sight, this derivation of the effects of Procrastinate seems more principled than Chomsky's (1993) version, because it simply compares the application of the operation that moves phonological features with the lack of application of this operation. However, if movement of formal and semantic features must leave copies in their original site, as seems plausible, O'Neil and Groat's account of Procrastinate effects is obtained at the cost of distinguishing two types of movement operations: (i) movement operations which leave copies behind, as is the presumably the case of operations involving movement of formal and semantic features; and (ii) movement operations which do not leave copies, as is the case of movement of

---

34 I am assuming here that PRO in (73) raises to the Spec of the embedded TP to check the EPP, as well as its "null" Case (see Chomsky and Lasnik (1993:561) and Martin (1992)).
phonological features. Movement of the subject from its VP internal position to the Spec of TP in English, for instance, appears to require two movement operations, one of each type. At this point, the advantage of this analysis over Chomsky's (1993) becomes rather dubious.

I will not pursue this issue any further. The discussion above suffices to show that the relation between strong features and phonetic realization of chain links does not seem to be well grounded either conceptually or empirically. I will return to this issue in section III.6, where I offer an alternative analysis of why a chain cannot have more than one link overtly realized.

III.4.4. Summary

The discussion in sections III.4.1 led us to conclude that traces do not differ from heads of chains as far as the computational system is concerned. Traces do enter into checking relations, thereby being able to move and to induce Minimal Link Condition violations. I proposed that the need to delete unchecked [-interpretable] features of traces is motivated by some plausible assumptions concerning the uniformity of chain links at LF with respect to feature composition (the Feature Uniformity Condition). Traces are not special in this regard, though; heads of chains may also have

---

35 See section II.10.2 for the derivation of the effects of Procrastinate under the Copy + Merge theory of movement.
[-interpretable] features deleted by Chain Uniformization (see discussion of (42)). Finally, it was suggested that the standard assumption that the head of the chain is the link which moves in successive cyclic movement may follow from economy considerations concerning the number of applications of deletion for purposes of the Feature Uniformity Condition (see sections III.6.2.5 and IV.2.3 for further discussion).

Conceptual considerations in section III.4.2 also led to the conclusion that traces should pattern with heads of chains in terms of linearization. We are then left with the unavoidable difference between traces and heads of chains regarding phonetic realization, which was discussed in section III.4.3.

In section III.6 below, I will propose that deletion of traces is determined by the conjunction of the hypothesis that traces are subject to the LCA and economy considerations concerning the number of applications of deletion. Before I present the specifics of the proposal, let us first reexamine the notion of distinctiveness relevant for chain formation, which is the subject of the next section.

III.5. Nondistinctiveness and Chain Formation

Consider the structure in (74), which is built from a numeration which has two instances of John.\textsuperscript{36} The first instance (John) merges with kissed

\textsuperscript{36} The indices in (74) are used only for expository purposes.
and then moves to the embedded subject position (John2); the second instance (John3) is inserted in the Spec of the matrix light verb and then moves to the matrix subject position (John4).

(74) \[ [TP \text{John}_4 \text{T} [\text{vp \text{John}_3} \text{ said that } [TP \text{John}_2 \text{ was } [\text{vp kissed \text{John}_1}]]]] ]

Although the four copies of John in (74) are in some sense identical, only two chains can be formed: CH$_1$ = (John$_4$, John$_3$) and CH$_2$ = (John$_2$, John$_1$). Any other pairing of copies of John violates the Minimal Link Condition and/or Last Resort. Forming a chain with the pair (John$_4$, John$_1$), for instance, violates the Minimal Link Condition because the other two copies intervene, whereas forming a chain with the pair (John$_3$, John$_2$) violates Last Resort because John$_3$ enters into no checking relation. Hence, regardless of whether or not a pair is formed by distinct instances of John, the Minimal Link Condition and Last Resort only allow the chains CH$_1$ and CH$_2$ to be formed.

However, it is not the case that the Minimal Link Condition and Last Resort render the notion of distinctiveness redundant in every situation. Take the embedded clause of (74), repeated below in (75), without taking its derivational history into consideration. If the pair (John, John) of (75) were to form a chain, it would satisfy both the Minimal Link Condition and Last
Resort; actual chain formation in this case will thus depend on whether or not the two instances of John are to be taken as distinct.

(75) \[ [TP \text{ John was } [\nu P \text{ kissed John }]] \]

As discussed in section II.10.4, Chomsky (1994:7, 1995:227) proposes that two identical lexical items should be marked as distinct for the computational system if they enter into the derivation through different applications of the operation Select. I argued in section II.10.4 that rather than encoding distinctiveness of lexical items through the operation Select, the system should encode nondistinctiveness of terms through the Copy operation. I will represent the relation that the Copy operation establishes between two terms by using superscripts.\(^{37}\) Thus, if the two instances of John in (75) were coindexed by the Copy operation, they could form a chain; otherwise, the system automatically interprets them as distinct.

In the first situation, the chain \(CH = (\text{John}_{i}, \text{John}_{i})\) does not satisfy the Feature Uniformity Condition, because the lower link has an unchecked Case-feature, as opposed to the upper link. Chain Uniformization then deletes this unchecked Case-feature, making it possible for CH to satisfy the Feature Uniformity Condition and for Full Interpretation to be met at LF. If

\(^{37}\) See section II.10.4 for the meaning of coindexation used here and for a suggestion as to how it is impossible to determine whether or not a term is a copy of another term through the history of the derivation.
the two occurrences of *John* in (75) cannot form a chain because they are distinct, the derivation will crash at LF because the unchecked Case-feature of the lower instance of *John* will induce a Full Interpretation violation. Recall that deletion of [-interpretable] features is a suboperation of Chain Uniformization, which in turn is linked to the Feature Uniformity Condition, not to Full Interpretation (see section III.4.1.2).

The discussion above focused on the relevance of nondistinctiveness for chain formation, convergence at LF, and interpretation at the C-I interface. Clearly, the notion of nondistinctiveness of terms plays no role at the PF level and at the A-P interface. Questions arise, however, about whether this notion is relevant for computations in the phonological component. If it can be shown that the phonological component operates with chains, we will have evidence that this component is sensitive to the nondistinctiveness of terms. With this in mind, let us reconsider the structure in (74), repeated in (76) with the indexation provided by the Copy operation.

(76) \[ \text{[TP John\textsuperscript{i} T [\text{VP John\textsuperscript{i} [\text{v said [CP that [TP John\textsuperscript{j} was [VP kissed John\textsuperscript{j}]]]]]]]}]} \]

Let us assume for the sake of discussion that PF should in some way reflect the number of occurrences of each lexical item specified in the initial
numeration of a given derivation. If so, the PF output of (76) should contain only two instances of John and the other two should be deleted. Within the framework assumed here, which makes no distinction between deletion and erasure, this deletion process must take place in the phonological component and not in the overt syntax; otherwise the elements deleted would receive no interpretation whatsoever at the C-I interface. If this mapping in the phonological component were sensitive only to the identity of the relevant sets of features, (76) could in principle surface as any of the sequences in (77), contrary to fact.

(77)  
a. John said that John was kissed.
   b. *John John said that was kissed.
   c. *John said that was kissed John.
   d. *Said that John was kissed John.

Notice that the unacceptability of the sentences in (77b)-(77d) cannot be attributed to an EPP violation because some clauses do not have subjects. Recall that within the Minimalist framework, the EPP is reinterpreted in terms of the strong feature of T. In (76), the strong feature of each T head was checked before Spell-Out; thus, the unacceptable sentences of (77) should pattern with the acceptable sentence in (78a), represented in (78b), where the embedded clause does not have its subject phonetically realized and the
strong feature of the embedded T was also checked before Spell-Out:

(78) a. Who did you say called yesterday?
   b. [ who^i did you say [ who^i T called yesterday ]

In addition, the unacceptability of (77b) and (77c) cannot be ascribed to whatever derives that-trace effects. If the complementizer of these sentences is deleted or is not part of the initial numeration, the resulting sentences are still ill-formed:

(79) a. *John John said was kissed.
   c. *John said was kissed John.

Finally, one cannot rely on some deviant interpretation assigned by the C-I interface to account for the unacceptability of the sentences in (77b)-(77d): the deletion operations which yield the sentences in (77) take place in the phonological component, independently of whatever interpretation C-I assigns to (76) (after further computations in the covert component).

The only possible PF output for the structure in (76) is the one in which the trace of each chain is deleted (cf. (77a)), showing that the phonological component applies the deletion operation (for reasons yet to be determined) to members of a chain and not simply to terms with identical
sets of features. This in turn suggests that the phonological component takes nondistinctiveness of terms into account when performing deletion.

Furthermore, under the plausible assumption that Form Chain is an operation of the mapping from N to λ, the phonological component is not able to identify the relevant pairs of instances of John in (57) as chains if the two possible chains have not been formed in the overt syntax. The derivation should then crash at PF or be canceled. I will return to the exact nature of this ill-formed result in section III.6.2.2 below.

**III.6. Deletion of Traces in the Phonological Component**

As pointed out in section III.4.3, any version of the copy theory of movement has to address the following questions: (i) why is deletion of chain links necessary to begin with?; and (ii) why does deletion target traces and not heads of chains? Applied to the derivation in (80), for instance, questions (i) and (ii) amount to asking why the structure in (80c) cannot have the PF output of (81a), and why the structure in (80c) surfaces as (81c) and not as (81b).

---

38 For a proposal that the grammar may be "fooled" by identical copies in terms of chain formation at LF, see Uriagereka (1994).

39 Evidence that this is the case regardless of chain formation will be provided in sections III.6.2.5 and in chapter IV.
Recall that it must be the case that the derivation of (81a) does not converge (see section III.4.3); otherwise, it would incorrectly rule out the derivation of (81c), which employs an extra application of deletion targeting the lower link of the chain CH = (Johni, Johni) in (80c). As for the status of the derivation of (81b), it is not possible to predict in advance whether it is canceled, crashes at PF, or converges but is more costly than the derivation of (81c). I address questions (i) and (ii) in sections III.6.1 and III.6.2. below.

### III.6.1 Nondistinctiveness and the Linearization of Nontrivial Chains

In section II.10.4, I proposed that the computational system encodes nondistinctiveness of terms through the Copy operation. More specifically, I suggested that when Copy targets a given term T, it assigns an unused index \( i \) to T if T has no index, and creates a copy of the indexed term \( T^i \); if T has
already been indexed as \( T^i \) by a previous application of the Copy operation, Copy targets \( T^i \) and simply creates another copy of it. This indexation device allows LF and the C-I interface to make the necessary distinction between terms with identical sets of features and nondistinct terms (see fn. 37).

As discussed in section III.5, deletion of traces indicates that the phonological component also differentiates terms with identical sets of features from nondistinct terms. In other words, the notion of nondistinctiveness is relevant for both the mapping from \( N \) to \( \lambda \), and the mapping from \( N \) to \( \pi \). Keeping this in mind, let us consider the asymmetric c-command relations in (82), in order to determine how this structure should be linearized according to the LCA.

(82)

\[
\begin{array}{c}
TP \\
\downarrow \\
John^i \\
\downarrow \\
w&as \\
\downarrow \\
kissed \\
\downarrow \\
John^i \\
\end{array}
\]

Consider the relation between the two copies of \textit{John} and the copula \textit{was} in (82), for instance.\(^{40}\) Since the upper copy of \textit{John} asymmetrically c-

\(^{40}\) The question of how two heads in a mutual c-command relation can be linearized is orthogonal to the issue of linearization of structures containing chains. See section II.17 for discussion.
commands was, we should obtain the order \(<\text{John}', \text{was}\>\), according to the LCA; likewise, since the copula asymmetrically c-commands the lower copy of John, the order \(<\text{was}, \text{John}'\>\) should be derived. Combining these two results, we should obtain the partial sequence \(\sigma = <\text{John}', \text{was}, \text{John}'\>\). Were the two instances of John distinct, \(\sigma\) would be a well-formed linear order, with the copula following an occurrence of John and preceding a different occurrence of John. However, since the two instances of John in (82) are nondistinct, was should precede and be preceded by the same element, John. \(\sigma\) is therefore not a linear order because it lacks asymmetry (if \(\alpha\) precedes \(\beta\), then it must be the case that \(\beta\) does not precede \(\alpha\)), which is a defining property of a liner order (see chapter II:fn. 94).

The structure in (82) also violates the irreflexivity condition on linear order (if \(\alpha\) precedes \(\beta\), then it must be the case that \(\alpha \neq \beta\); see chapter II:fn. 94 and Kayne (1994:134, fn. 8)). Since the upper copy of John asymmetrically c-commands the lower one, the former should precede the latter in accordance with the LCA. Given that the two copies of John in (82) are nondistinct, that would amount to saying that John should precede itself.

Failure to yield a linear order thus provides a straightforward account for the fact that the structure in (82) cannot surface as the sentence in (81a). To put it more generally, if the links of a chain count as nondistinct for linearization purposes in virtue of their being associated by the Copy
operation, we have an explanation for why a chain cannot surface at PF with more than one link overtly realized: the syntactic object containing such a chain cannot be linearized.

Under the assumption that the rules of the phonological component that apply after Linearize operate on a sequence of $X^0$ elements, the attempted derivation of (81a) from (82) is canceled, because Linearize yields no output and, therefore, no PF object is formed (see fn. 28). This is a welcome result, since the derivation of (81a) should be prevented from being compared with the derivation of (81c) for purposes of economy (see section III.4.3). If the derivation of (81a) is canceled because it cannot be linearized, no question of convergence or economy can be raised.41

I propose that deletion of chain links is thus required for a structure containing nontrivial chains to be linearized in accordance with the LCA. This proposal has the welcome conceptual advantage that it takes both heads of chains and traces to be subject to linearization, and that the lack of phonetic realization is not stipulated as an inherent property of traces. What is required at this point is an independent motivation for why deletion targets traces and not heads of chains. This is the topic of section III.6.2.3 below. Before beginning this discussion, let us first examine the more general case regarding the optimality of deletion for purposes of linearization.

41 Recall that the alternative derivation of the sentence (81a) in which the two instances of John are distinct crashes at LF, because the Case-feature of the lower instance induces a violation of Full Interpretation (see section III.5).
III.6.2. Optimality of Deletion

III.6.2.1. Full vs. Scattered Deletion

Consider the simplified structure in (83) below, in which the embedded object DP raises to the matrix subject position, leaving two copies behind. As discussed in section III.6.1, a structure such as (83) cannot be linearized as is. The highest instance of the tall man asymmetrically \( c \)-commands the verb appears, for instance, which in turn asymmetrically \( c \)-commands the other two instances of the tall man. Given that these three copies are nondistinct, no linear order between the tall man and appears can be established in accordance with the LCA. Thus, if Linearize applies to (83), the derivation will be canceled, because Linearize will yield no output for further computations in the phonological component and no PF object will be formed.

\[
(83) \quad [_{TP} \ [_{DP} \ \text{the tall man} ] \ . \ \text{appears} \ . \ [_{TP} \ [_{DP} \ \text{the tall man} ] \ . \ \text{to have been} \ . \ \text{kissed} \ . \ [_{DP} \ \text{the tall man} ] \ . ] ]
\]

I have proposed in section III.6.1 that deletion may allow a structure containing nontrivial chains to be linearized by eliminating "repeated" material which induces lack of asymmetry and irreflexivity in the intended
linear order. Nothing that has been said so far, however, prevents deletion from applying within the different links of a chain, in what may be called "scattered deletion". Assuming that the DP *the tall man* has a structure along the lines of (84), deletion could target different terms in each of the links of the DP-chain in (85), which is the structure in (83) with the DP *the tall man* replaced by its bare X'-Theory representation, yielding a structure such as (86), for instance.

\[(84) \quad L = \{\text{the}, \{\text{the}, \{\text{man}, \{\text{tall, man}\}\}\}\}\]

\[\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \Quad
the LCA, yielding the sentence in (87).  

(87) *The appears tall to have been kissed man.

I propose that although the derivation of (87) from the structure in (85) converges at PF, it is not the most economical derivation starting from (85). To put it more generally, scattered deletion does not yield an optimal derivation. Take (86), for example, under the assumption that deletion for purposes of linearization only targets terms. Given the structure of the tall man in (84), the derivation of (86) from (85) requires that the deletion operation apply (at least) five times, targeting the following terms: the term $K = \{ \text{man}, \{\text{tall, man}\} \}$ of the chain link in the matrix subject position, the terms the and man of the link in the intermediate subject position, and the terms the and tall of the link in the object position. Three other derivations starting from (85) which employ "full deletion" of chain links are more economical. If deletion targets the term $L = \{ \text{the}, \{ \text{the, \{man, \{tall, man\}\}} \}$ of two of the links of the DP-chain in (85), the structures in (88) will be derived.

---

42 Technically, the mapping from (83) to (84) is made possible by the recursive step of the definition of the LCA, repeated below (see Uriagereka (forthcoming:chap. 3)).

(i) **Linear Correspondence Axiom (LCA):**
A category $\alpha$ precedes a category $\beta$ iff:
(a) $\alpha$ asymmetrically c-command $\beta$; or
(b) $\gamma$ precedes $\beta$, and $\gamma$ dominates $\alpha$. 

Each structure of (88) can be linearized in accordance with the LCA, yielding the sentences in (89a)-(89c), respectively. Given that the derivation of any of the sentences in (89) employs only two applications of deletion, it blocks the derivation of (87), which requires (at least) five applications of this operation.

(89) a. The tall man appears to have been kissed.

b. *Appears the tall man to have been kissed.

c. *Appears to have been kissed the tall man.

Under the assumption that deletion targets one term per application, economy considerations concerning the number of applications of the deletion operation block scattered deletion within chains in favor of full deletion of chain links. I refer to the operation of the phonological component which converts (85), for instance, into structures such as (86) or
(88) as *Chain Reduction*:

(90)  *Chain Reduction*:

Delete the minimal number of terms of a nontrivial chain CH which suffices for CH to be mapped into a linear order in accordance with the LCA.

Notice that if Chain Reduction had deleted each of the three links of the DP-chain in (85), forming the object in (91a), the problem of lack of asymmetry and irreflexivity would be circumvented and (91a) could be linearized in accordance with the LCA, eventually yielding the sentence in (91b). The derivation of (91a) from (85), where the deleted material is nonrecoverable, is not optimal, however. Chain Reduction in this derivation employs (at least) three applications of deletion, when only two applications would suffice for the DP-chain to be mapped into a linear order, as shown in (88). Therefore, recoverability of deletion of terms in the analysis explored here follows straightforwardly from economy considerations, and nothing additional need be stated in the theory to ensure that it obtains (see section III.6.4 for further discussion).43

---

43 As was the case with Chain Uniformization, it is not necessary to specify that Chain Reduction deletes the smaller number of terms of a nontrivial chain CH (each application of deletion targeting a single term). Once it is postulated that Chain Reduction allows CH to be mapped into a linear order by having some of its terms deleted, economy considerations concerning the number of applications of the suboperation of deletion may independently determine the number of terms to be
(91)  a.  [TP appears [TP to have been kissed ] ]

b.  *Appears to have been kissed.

III.6.2.2. A Closer Look at Chain Reduction

As stated in (90), Chain Reduction of a nontrivial chain CH deletes some terms of CH so that the surviving terms of this nontrivial chain can be mapped into a linear order in accordance with the LCA. Notice that Chain Reduction proceeds without taking into consideration how the structure containing CH can be linearized. That Chain Reduction may apply in this "local" fashion is made possible by the interaction of the conditions constraining the operation Form Chain, which are repeated below:

---
deleted. For purposes of exposition, I will however use the formulation of Chain Reduction given in (90).---
(92)  *Form Chain:*

The syntactic objects $\alpha$ and $\beta$ can form the nontrivial chain $CH = (\alpha, \beta)$ only if:

(i)  $\alpha$ c-commands $\beta$;

(ii) $\alpha$ is nondistinct from $\beta$;

(iii) (at least) one given feature $F$ of $\alpha$ enters into a checking relation with a sublabel of $K$, where $K$ is the head (of the projection) with which $\alpha$ merges; and

(iv) there is no syntactic object $\gamma$ such that $\gamma$ has a feature $F'$ which is of the same type as the feature $F$ of $\alpha$, and $\gamma$ is closer to $\alpha$ than $\beta$ is.

Given that two nondistinct terms in a mutual c-command relation cannot enter into a checking relation, the Last Resort condition on Form Chain (see (92iii)) entails that if the pair $(\alpha, \beta)$ forms a chain, $(\alpha, \beta)$ must be in an asymmetric c-command relation. However, as discussed in section III.6.1, two nondistinct terms (see (92i)) in an asymmetric c-command relation cannot be part of a linear order which is formed in accordance with the LCA. Consider the terms $T_1$ and $T_2$, for instance, such that they are links of a nontrivial chain and $T_1$ asymmetrically c-commands $T_2$. According to
the LCA, \( T_1 \) should precede \( T_2 \) given that \( T_1 \) and \( T_2 \) are nondistinct (see (92i)), however, this would amount to saying that \( T_1 \) should precede itself, in violation of the irreflexivity condition on linear order (see chapter II:fn. 94 and section III.6.1).

Thus, when a structure \( \Sigma \) containing a nontrivial chain \( \text{CH} \) is spelled out, the phonological component has the information that the links of \( \text{CH} \) will make it impossible for a linear order to obtain, even without considering the linearization of \( \Sigma \) as a whole. Chain Reduction then deletes terms of \( \text{CH} \) in such a way that the irreflexivity condition on linear order is satisfied; that is, either the surviving terms of each chain link are distinct or only one chain link survives (scattered and full deletion, respectively; see section III.6.2.1). By satisfying the irreflexivity condition, the output of Chain Reduction in principle makes it possible for a linear order to be derived in compliance with the LCA. Whether or not a linear order will actually obtain will then depend on the structural properties of \( \Sigma \) as a whole.

Below I show that when compared to a global version of Chain Reduction such as the one given in (93), the "local" version of Chain Reduction given in (90) is to be preferred not only for conceptual reasons, given that it reduces computational complexity, but for empirical reasons as well.
(93)  

Chain Reduction:

Delete the minimal number of terms of a nontrivial chain CH in order for the largest structure containing CH to yield a linear order in accordance with the LCA.

Suppose, for instance, that after assembling the DP in (94a), the computational system starts assembling another structure by making a copy of the object of questioned and merging it with the verb arrested, which was pulled out from the numeration, as illustrated in (94b).44 After further computations, we obtain the two objects in (94c), which then merge, yielding the structure in (94d).

(94)  a.  [DP the fact that the police questioned John ]

b.  [DP the fact that the police questioned Johni ]

[VP arrested Johni]

c.  [DP the fact that the police questioned Johni ]

[TP T [VP indicates that [TP Johni was arrested Johni ] ] ]

d.  [TP [DP the fact that the police questioned Johni ] T [VP indicates that [TP Johni was arrested Johni ] ] ]

Recall that under the Copy + Merge theory of movement, which I am

---

44 This is an instance of what I have referred to as sideward movement (see section II.10.3 and chapter IV).
arguing for here, the sequence of operations Copy and Merge is not necessarily associated with chain formation (see chapter IV for detailed discussion). Nontrivial chains result from applications of the operation Form Chain (see (92)) in the course of the mapping from N to $\lambda$. Suppose then that no chain involving the nondistinct instances of John in (94d) has been formed prior to Spell-Out. Given that the phonological component recognizes nondistinct terms (see section III.5), the nondistinct instances of John in (94) will then yield a violation of the asymmetry condition on linear order, making it impossible for (94d) to be linearized in accordance with the LCA (see section III.6.1).

The verb was, for instance, asymmetrically c-commands the copy of John in the object position of arrested and is asymmetrically c-commanded by the copy John in its subject position. The LCA should therefore impose the contradictory requirement that was should precede and be preceded by the same element, namely, John. Once the structure cannot be linearized, no PF object is formed and the derivation is canceled (see fn. 28). Crucially, Chain Reduction cannot randomly delete some instances of John in (94d), allowing the structure to satisfy the asymmetry condition on linear order, because it can only apply to nontrivial chains (see (90) and (93)), and no nontrivial chain has been formed before Spell-Out in this scenario.

45 Whenever nondistinct copies induce a violation of the asymmetry condition on linear order, the irreflexivity condition is violated as well. For the sake of brevity, I will focus on only one of them when the discussion of both violations is unnecessary.
Consider now the case in which chains may have been formed before (94d) is spelled out. For the purposes of this discussion, let us focus on the c-command condition on Form Chain (see (92i)). In (94d), the only chain that can be formed with the instances of John is the one between the subject and the object of was arrested. The occurrence of John in the object of questioned neither c-commands nor is c-commanded by either of the other two copies. This derivation can now help us choose between the two versions of Chain Reduction in (90) and (93).

As discussed above, a chain such as CH = (John, John) formed between the subject and the object of was arrested in (94d) would prevent a linear order from being derived in accordance with the LCA, regardless of the structure which contains it; the fact that the nondistinct copies of John are in an asymmetric c-command relation does not allow the irreflexivity condition on linear order to be met. According to the "local" version of Chain Reduction in (90), CH should be reduced in such a way that its terms could be mapped into a linear order. Thus, Chain Reduction deletes either of the links of CH, overcoming the irreflexivity problem and in principle allowing the surviving link of CH to be mapped into a linear order.

Reduction of CH as described above does not allow the whole structure containing CH in (94d) to be linearized, however. The link that survives reduction and the copy in the object position of questioned induce a violation of asymmetry, canceling the derivation. To see this, take the verb
indicates, for instance. On the one hand, it is required to precede John in virtue of asymmetrically c-commanding the copy that survives reduction inside the TP "John was arrested John"; on the other hand, it is also required (by the recursive step of the definition of the LCA; see fn. 42) to be preceded by John, since the TP "the fact that the police questioned John", which dominates John, asymmetrically c-commands indicates. Therefore, according to the version of Chain Reduction in (90), there is no grammatical output for the derivation sketched in (94), regardless of whether or not a chain has been formed prior to Spell-Out.

The global version of Chain Reduction in (93), on the other hand, would delete the smallest number of links of the chain CH formed between the subject and the object of was arrested which would allow the whole structure in (94d) to be linearized. As we have just seen, if Chain Reduction deletes one of the links of CH, the derivation will be canceled; the link that survives and the instance of John in the object of questioned would induce an asymmetry violation preventing the structure from being linearized. If Chain Reduction deletes both links of CH, as represented in (95a), however, no violation of asymmetry will arise and the whole structure can be linearized. The sentence corresponding to (95a), given in (95b), is nevertheless unacceptable.
(95)  

a. [TP [DP the fact that the police questioned John] VP indicates that [TP was arrested] ] ]

b. *The fact that the police questioned John indicates that was arrested.

The restricted version of Chain Reduction in (90) is therefore to be preferred over the global version in (93), given that the former correctly rules out the unacceptable sentence in (95b) in all circumstances, whereas the latter wrongly rules (95b) in, if the subject and the object of was arrested in (94d) form a chain before Spell-Out. In chapter IV, we will examine derivations similar to the one in (94), which provide additional evidence for the restricted version of Chain Reduction in (90).45

III.6.2.3. Deletion of Traces vs. Deletion of Heads of Chains

We are still left with a problem from section III.6.2.1, if the derivations that convert the structure in (83) into the PF output associated with the sentences in (89), repeated below in (96) and (97), are equally economical. We need to explain why the only derivation which yields an acceptable sentence

45 As will become clear in chapter IV, the problem with (93) is not that it refers to the largest structure containing a nontrivial chain rather than the smallest one. In instances where the matrix Spec of CP is occupied by a chain link (in wh-movement structures, for instance), the smallest and the largest structure actually coincide, namely, it is the matrix CP. The point remains that Chain Reduction operates blindly, without considering the structure containing the nontrivial chain to be reduced.
(see (97a)) is the one in which the traces — but not the head of the DP-chain — are deleted (see (96b) and (97b)). I propose that checking theory provides the basis for an account of the paradigm in (97), as shown below.

\[(96) \quad [TP [DP \text{ the tall man }]^i] \text{ appears } [TP [DP \text{ the tall man }]^i] \text{ to have been kissed } [DP \text{ the tall man }]^i] \]

\[(97) \quad \begin{align*}
\text{a.} & \quad \text{The tall man appears to have been kissed.} \\
\text{b.} & \quad *\text{Appears the tall man to have been kissed.} \\
\text{c.} & \quad *\text{Appears to have been kissed the tall man.}
\end{align*} \]

So far, we have restricted the discussion of the relation between interpretability of formal features and checking theory to the mapping from N to O. Thus, checking operations render [-interpretable] features invisible at LF, eventually allowing the derivation to meet Full Interpretation and converge at this level; [+interpretable] features, on the other hand, are taken to be unaffected by checking operations, thereby being available for the relevant interpretation they receive at the C-I interface.

Let us now consider the role of formal features in the mapping from N to \(\pi\). It is very plausible that formal features are relevant for computations in the morphological subcomponent of the phonological component (see
chapter II:fn. 9). For instance, Morphology must ensure that phonological and Case-features are correctly paired; the phonological features of the pronoun *he*, for example, must be associated with nominative, and not with accusative Case. However, it is clear that formal features are not PF objects. Therefore, somewhere in the mapping from Spell-Out to \( \pi \), formal features must be eliminated; otherwise, Full Interpretation will not be met at PF.

Leaving the discussion of [+interpretable] features for the next section, let us focus on how the computational system eliminates [-interpretable] features in the course of the mapping from Spell-Out to \( \pi \). We have already assumed that a checking operation makes a [-interpretable] feature invisible at LF. A natural extension of this assumption is to take checking operations to render [-interpretable] features invisible at PF as well, given that no formal feature is interpreted at the A-P interface. Notice that this extension does not prevent checked [-interpretable] features from being active in the phonological component; they are invisible at both interface levels, but are accessible to the computational system. Recall that although unable to enter into further checking relations, checked [-interpretable] features are able to block movement operations, for instance (see section III.4.1.5).

This extension of checking theory is however insufficient to ensure

---

47 Hence the version of Spell Out adopted here, which ships formal features to both the covert and the phonological components (see Chomsky (1994:8) and the discussion in section II.5).
that legitimate PF objects are formed, because not all formal features are eliminated in the mapping from N to π by checking operations. The structure in (98), for instance, is spelled out with the Case-feature of a problem, for instance, being unchecked and therefore still visible at PF.

(98) \[ \text{[ there is [ a problem ]-CASE here] } \]

Given that a problem is a trivial chain, one cannot resort to the deletion suboperation of Chain Uniformization or Chain Reduction to eliminate this unchecked Case-feature. The phonological component must therefore resort to an additional operation obliterating unchecked formal features, in order for the derivation of (98) to satisfy Full Interpretation at PF. Furthermore, it is reasonable to assume that such an operation should follow Chain Reduction; otherwise it would be redundant in eliminating formal features of terms which would be deleted by Chain Reduction.

Consider the output of Linearize. It must involve a sequence of pairs \( \sigma = \langle (F, P)_1, (F, P)_2, ..., (F, P)_n \rangle \) such that \( F \) is a set of formal features and \( P \) is a set of phonological features. Let us then assume that the phonological component applies a rule such as (99) below to the output of Linearize, where deletion targets a single feature per application.\(^{48} \)

\(^{48}\) Similarly to the case of Chain Uniformization and Chain Reduction, the specification of the number of features to be deleted in the description of FF-Elimination in (99) may be taken to follow from economy considerations concerning the
Formal Feature Elimination (FF-Elimination):

Given the sequence of pairs \( \sigma = \langle (F, P)_1, (F, P)_2, ..., (F, P)_n \rangle \) such that \( \sigma \) is the output of Linearize, \( F \) is a set of formal features and \( P \) is a set of phonological features, delete the minimal number of features of each set of formal features in order for \( \sigma \) to satisfy Full Interpretation at PF.

Applied to the \( \sigma \) sequence of (98), FF-Elimination deletes the unchecked Case-feature of a problem, allowing the derivation to eventually satisfy Full Interpretation and converge at PF.

Let us now reconsider the derivation of (96), repeated below in (100) with the relevant Case-features represented. After being assembled and merged with the verb kissed, the DP the tall man raises to the Spec of each T head in order to check their strong D-features; in addition, the Case-feature of the topmost copy of the tall man enters into a checking relation with the Case-feature of the matrix T. Since Case is a [-interpretable] feature, this checking relation renders the Case-feature of the highest copy of the tall man invisible at LF and, according to the extension of the checking theory proposed above, invisible at PF as well. The lower copies of the tall man, on the other hand, have their Case-features unaffected by the Case-checking number of applications of the suboperation of deletion. As before, I will keep the description as in (99) for expository purposes.
relation involving the highest copy (see section III.4.1.2). I will represent this state of affairs by using subscripts to annotate checked features, as exemplified in (100).

\[(100) \ [\text{TP} \ [\text{DP} \ \text{the tall man}]^{\text{-CASE}} \ \text{appears} \ [\text{TP} \ [\text{DP} \ \text{the tall man}]^{\text{-CASE}} \ \text{to have been kissed} \ [\text{DP} \ \text{the tall man}]^{\text{-CASE}}] ]\]

This strongly derivational approach to the copy theory of movement thus requires that the notion of "sameness" relevant for Form Chain be defined based on nondistinctiveness (see section II.10.4), rather than identity (see (92ii)). Were the relevant notion of "sameness" defined in terms of identity, the topmost copy of the tall man in (100) could not form a chain with either of the other copies, because the Case checking relation it participated in would render it nonidentical with respect to the other copies.

As it stands, the DP-chain in (100) would violate the Feature Uniformity Condition, because the Case-feature of the topmost copy is invisible at LF, whereas the Case-features of the other copies are not. If not eliminated, the Case-features of the lower links of the DP-chain will induce a Full Interpretation violation at LF and at PF. These unchecked features can in principle be deleted by Chain Uniformization, by Chain Reduction or by FF-Elimination. Let us consider each of these alternatives in turn.

In order for the chain in (100) to satisfy the Feature Uniformity
Condition, Chain Uniformization must delete the Case-features of the lower links, yielding the structure in (101). As a by-product of Chain Uniformization, the DP-chain in (101) complies with Full Interpretation at LF.

\[(101) \quad [TP [DP \text{ the tall man }]^i_{\text{CASE}} \text{ appears } [TP [DP \text{ the tall man }]^i \text{ to have been kissed }]_{DP \text{ the tall man}^i}] \]

I have been tacitly assuming that Chain Uniformization applies in the covert component. However, given that the uniformity condition on the mapping from $N$ to $\lambda$ makes the same set of operations available in overt syntax and in the covert component (see section II.5), one wonders whether Chain Uniformization could apply to the chain of (100) overtly. If that were possible, it would enable the DP-chain to satisfy Full Interpretation at both LF and PF without any other operation eliminating the unchecked Case-features.

Such a derivation is blocked by economy considerations, however. Consider the operations which may apply overtly, namely, Select, Copy, Merge, and Form Chain. If Select, Merge or Form Chain do not apply overtly, the derivation will be canceled because no PF object is formed (see sections II.4 and III.6.2.2); as for Copy, if it does not apply overtly, a strong feature will not be checked and the derivation will crash at PF (see section
II.6). In contrast, Chain Uniformization is triggered by the need to satisfy the Feature Uniformity Condition, which is a convergence condition holding of LF objects. Since chains are not PF objects, it is irrelevant for the phonological component whether or not the chains it receives from Spell-Out are uniform in terms of feature composition. Thus, at no derivational step in the overt syntax is the computational system forced to apply Chain Uniformization.

After all the strong features are checked, the computational system may either apply Chain Reduction to the chains formed overtly or apply Spell-Out. Since Spell-Out is required for a derivation to be generated, it is costless, therefore being more economical than Chain Uniformization, which is an operation related to convergence conditions (see section II.4). Although available throughout the mapping from N to $\lambda$, Chain Uniformization is thus prevented from applying overtly by economy considerations. Hence, the structure in (100) is the one actually sent to the phonological component, and the structure in (101) obtains only in the covert component.

If the DP-chain of (100) is uniformized in the covert component, the unchecked Case-features of the lower links should then be eliminated in the phonological component in order for the derivation to converge at PF. The question is whether such an elimination is carried out by Chain Reduction, by FF-Elimination, or by a combination of both operations. As mentioned above, it is plausible to assume that FF-Elimination applies after Chain
Reduction; in other words, FF-Elimination is invoked just in case some formal features were not eliminated by checking or Chain Reduction. Furthermore, economy considerations prevent FF-Elimination from deleting a checked feature, since it is already invisible at PF.

Let us then see how the DP-chain of (100) is to be reduced. As discussed in section III.6.2.1, the optimal reduction of this chain involves only two applications of deletion targeting any two of its links, as shown in (102):

\[(102) \quad \begin{align*}
&\text{a. } [\text{TP } [\text{DP the tall man }]^i\text{-CASE appears } [\text{TP to have been kissed }]] \\
&\text{b. } [\text{TP appears } [\text{TP } [\text{DP the tall man }]^i\text{-CASE to have been kissed }]] \\
&\text{c. } [\text{TP appears } [\text{TP to have been kissed } [\text{DP the tall man}]^i\text{-CASE }]]
\end{align*}\]

If the DP-chain of (100) is reduced as in (102a), no further application of FF-Elimination is required for Full Interpretation to be satisfied, because the Case-feature of the copy that survives is checked and is therefore invisible at PF; the PF output in (103a) is then derived after further applications of phonological rules. By contrast, if the DP-chain is reduced as in (102b) or (102c), the convergent PF outputs in (103b) and (103c) are obtained only if FF-Elimination deletes the unchecked Case-feature of the copy that survives.

---

49 It is worth observing that intrinsic ordering of operations is not something extraneous to the phonological component (see Chomsky (1994:28, fn. 34)). In order for wanna-contraction to be accounted for, for instance, it must be the case that Morphology precedes Chain Reduction/Linearize.
The derivation in which the head of the chain survives Chain Reduction, such as (102a), is therefore more economical than derivations in which other links survive Chain Reduction, as in (102b) and (102c), because it requires fewer (if any) additional applications of deletion by FF-Elimination. The pattern of acceptability in (103) hence results.

(103)  a. The tall man appears to have been kissed.

b. *Appears the tall man to have been kissed.

c. *Appears to have been kissed the tall man.

Recall that in Chomsky's (1995:chap. 4) system, it is stipulated that a trace is affected by the checking relation that the head of its chain enters into (see section III.4.1.2); this symmetric theory of checking relations involving chain links then has to be combined with the further stipulation that traces must be erased in the phonological component.

The analysis developed here, on the other hand, derives the difference between heads of chains and traces regarding phonetic realization by pursuing the null hypothesis under the copy theory of movement that traces should have no inherent properties that would distinguish them from heads of chains. By keeping checking configurations as simple as possible, only allowing the elements in the checking domain of a given head H to enter into a checking relation with H, we obtain an asymmetry between
heads of chains and traces. The asymmetry resulting from this strongly derivational version of the copy theory of movement in turn enables us to derive the fact that traces cannot be pronounced from economy considerations concerning the number of applications of deletion by FF-Elimination.

III.6.2.4. [+interpretable] Feature Checking and Deletion of Traces

Consider the structures of the sentences in (104), provided in (105). In (105a) John moves from the Spec of the embedded light verb to check the strong D-feature of the infinitival T head, whereas in (105b) what moves from the object position to the Spec of the null interrogative complementizer to check its strong wh-feature (see sections II.14.3.3. and III.4.1.4).

(104)  a. I expected John to call me.
       b. What did you buy?

(105)  a. \[TP I expected [TP Johni to [\nu P Johni [\nu call me ] ] ] ]
       b. \[CP whati did+Q [TP you buy whati ] ]

In both chains of (105), it is the categorial feature (a [+interpretable]
feature) of the moved element which enters into a checking relation with a sublabel of the target. If [+interpretable] features remain unaffected by checking operations, as proposed by Chomsky (1995:chap. 4), the chain links of (105a) and (105b) will be identical with respect to the only checking relation that takes place overtly, as represented in (106) below. Thus, the structures in (106) provide no way to account for why Chain Reduction deletes the lower copies.

(106)  

\begin{align*}
\text{a.} & \quad [\text{TP I expected [TP John}^1\text{-D to [VP John}^1\text{-D [v call me ] ] ] }] \\
\text{b.} & \quad [\text{CP what}^1\text{-WH did+Q [TP you buy what}^1\text{-WH ] ]}
\end{align*}

I have proposed in section III.6.2.3 that if a [-interpretable] feature is checked, it is made invisible at both interface levels. Given that no formal feature is assigned an interpretation by the A-P system, this view of checking relations can be generalized by assuming that a checking operation renders a given feature F invisible at the level at which it would induce a Full Interpretation violation. In other words, a checking operation may render F invisible at PF, regardless of the interpretability of F at the C-I interface. A checked [-interpretable] feature will then be invisible at both PF and LF, whereas a checked [+interpretable] feature will be invisible at PF, but visible
at LF (see section III.6.2.5 below for further discussion).

Under this revised extension of checking theory, the appropriate representation of the sentence in (104) in the phonological component is (107) rather than (106), where the subscript convention is now generalized to mean 'invisible at the relevant interface':

\[(107) \quad \text{a. } [\text{TP I expected } [\text{TP John}^{-D} \rightarrow [\nuP \text{John}^{-D} \rightarrow [\nu \text{call me }]]]]
\]

\[\text{b. } [\text{CP what}^{-\text{WH}} \text{did+Q } [\text{TP you buy what}^{-\text{WH}}]]
\]

When applied to the chains of (107), Chain Reduction can in principle delete either link of the chains \(\text{CH}_1 = (\text{John}^{-D}, \text{John}^{-D})\) and \(\text{CH}_2 = (\text{what}^{-\text{WH}}, \text{what}^{-\text{WH}})\). If it deletes the upper link of either chain, FF-Elimination will be required to delete the unchecked D-feature and wh-feature, respectively; if Chain Reduction deletes the lower links, no application of FF-Elimination is required because the D-feature and the wh-feature of the upper links are checked and therefore invisible at PF. Again, the optimal derivation is the one in which Chain Reduction deletes every chain link except the head of the chain.

\[\text{To a certain extent, this is a revival of Chomsky’s (1993:33) notion of Greed, which requires that a given element move just in case it will satisfy some of its own morphological requirements as a result of the movement operation. If an analysis along the lines of the one I am proposing in this section is on the right track, the head of linked chains formed overtly will always have at least one feature rendered invisible at PF by a checking operation.}\]
III.6.2.5. Successive Cyclic Movement and Deletion of Traces

An apparent problem for the extension of checking theory proposed in section III.6.2.4 is presented by instances of successive raising such (108), roughly represented in (109).\(^{51}\)

(108) I believe John to be likely to be kissed.

(109) \([TP I believe [TP John\textsuperscript{i} to be likely [TP John\textsuperscript{i} to \[VP be kissed John\textsuperscript{i} ] ] ] ]\]

Let us examine the derivation of (109) stepwise. A copy of the most embedded object John is made and merges with the most embedded TP; its D-feature then enters into a checking relation with the T head. Suppose that, as proposed above, this checking relation renders the categorial feature of John invisible at PF, yielding (110a). After the structure in (110b) is assembled, two potential options are available for the higher T head to have

---

\(^{51}\) Instances of successive cyclic movement of wh-phrases such as the one illustrated in (i) also pose the same type of problem and presumably are subject to the same analysis I develop in this section. I will focus the discussion on A-movement, because successive cyclic wh-movement presents the additional problem of how to characterize the checking relation of a declarative complementizer and the trace of a wh-phrase in its Spec (see fn. 7).

(i) \([CP [ which picture of himself ]d John say [CP [ which picture of himself ] that Mary like [ which picture of himself ] ] ]\]
its strong D-feature checked: either the instance of John in the Spec of the embedded TP or the one in the embedded object position can in principle be copied and merge with the embedding TP.

(110)  

a. \([TP \text{John}^{i-D} \text{to } \left[\varphi' \text{be kissed John}^{i-D}\right]]\)

b. \([TP \text{to be likely } TP \text{John}^{i-D} \text{to } \left[\varphi' \text{be kissed John}^{i-D}\right]]\)

Let us consider the possibility in which the instance of John in the embedded subject position is the one copied, as represented in (111) before any checking operation proceeds. If the higher T head in (111) could be checked by a copy of the embedded subject despite the fact that the categorial feature of that instance of John has already been checked with respect to PF, the eventual linked chain \(CH = (\text{John}^{i-D}, \text{John}^{i-D}, \text{John}^{i-D})\) will remain as in (111) after the higher T has its strong feature checked. After the structure in (112) is finally assembled and spelled out, CH should in principle be reduced as (113a), (113b), or (113c). (113a) is more costly than the other two possibilities because it still requires that the D-feature of the copy that survives be deleted by FF-Elimination; hence the unacceptability of (114a). In turn, since (113b) are (113c) are equally costly, the analysis developed here predicts that the sentences derived by these possibilities are both acceptable. As shown in (114b) and (114c), this prediction is false.
Recall that the gist of Chomsky's (1995) proposal concerning the relation between interpretability and accessibility to the computational system is that if a [-interpretable] feature \( F \) is made invisible at LF by a checking operation, \( F \) cannot enter into further checking relations (see section II.14). Under the generalized version of checking theory proposed in sections III.6.2.3 and III.6.2.4, Chomsky's proposal can be naturally extended to checking operations which render a given feature invisible at PF as well. Put simply, a feature which has been made invisible at either interface level by a checking operation cannot participate in any further checking relation.
If this is so, there is no grammatical output for the mapping from (111) to (112), where the topmost instance of John is a copy of the intermediate one. Since the topmost copy has its D-feature rendered invisible at PF by a previous checking operation, it cannot check the strong D-feature of the T head of the intermediate clause. Given that the topmost copy of John enters into no checking relation with the T head, Last Resort prevents it from forming a chain with either of the lower copies (see (92iii)). The highest instance of John and one of the other copies will then induce a violation of asymmetry when Linearize applies to (112), causing the derivation to be canceled (see section III.6.2.2).

Let us now consider the possibility in which rather than copying the instance of John in the embedded subject position of (110b) to check the strong feature of the upper T head, the computational system copies the instance of John in object position, which does not have its D-feature checked. If that copy is the one which merges with the structure in (110b), we will have the structure in (115a) after merger, and the structure in (115b), after a checking relation is established between the strong feature of T and the D-feature of the upper copy of John. As seen above, the structure in (115b), which is identical to (111), makes the wrong prediction that both sentences in (114b) and (114c) should be acceptable, given that their derivation would have the same cost.
(115) a. \[ TP \text{John}^i-D \text{ to be likely } [TP \text{John}^i-D \text{ to } [VP \text{ be kissed John}^i-D ]] ] \\

b. \[ TP \text{John}^i-D \text{ to be likely } [TP \text{John}^i-D \text{ to } [VP \text{ be kissed John}^i-D ]] ] \\

The highest copy of John in (115b) cannot form a chain with the lowest copy, because the embedded subject intervenes, giving rise to a Minimal Link Condition violation. Recall that although the D-feature of the embedded subject has been rendered invisible at PF, it is still present in the structure and is therefore able to induce a Minimal Link Condition violation (see section III.4.1.5). The highest and the lowest copies could still be part of the linked chain \( CH = (\text{John}^i-D, \text{John}^i-D, \text{John}^i-D) \), if the two upper copies had formed a chain (see Chomsky and Lasnik (1993:563) and section III.4.1.1). This however would wrongly allow the linked chain to be reduced as (113b), yielding the sentence in (114b). Let us consider how this possibility can be ruled out.

We have already seen that if the highest subject is a copy of the embedded subject in (111), the derivation violates Last Resort, because no feature of the highest copy of John participates in a checking relation. In (115a), the highest copy of John can enter into a checking relation with the T head because its D-feature has not been checked, but the intermediate copy of John cannot. Suppose that when evaluating whether a pair \( (\alpha, \beta) \) forms a chain, Last Resort not only checks whether a feature of \( \alpha \) enters into a
checking relation with a sublabel of the category it merges with, but also whether \( \beta \) could actually enter into the same checking relation in the position of \( \alpha \), as in the reformulation of Last Resort and Form Chain given in (116) and (117).

(116) \textit{Last Resort:}

The chain CH = \((\alpha, \beta)\) can be formed if:

(i) (at least) one given feature F of \( \alpha \) enters into a checking relation with a sublabel of K, where K is the head (of the projection) with which \( \alpha \) merges; and

(ii) the corresponding feature F of \( \beta \) could enter into a checking relation with a sublabel of K
Form Chain:

The syntactic objects \( \alpha \) and \( \beta \) can form the nontrivial chain \( CH = (\alpha, \beta) \) only if:

(i) \( \alpha \) c-commands \( \beta \);

(ii) \( \alpha \) is nondistinct from \( \beta \);

(iii) (at least) one given feature \( F \) of \( \alpha \) enters into a checking relation with a sublabel of \( K \), where \( K \) is the head (of the projection) with which \( \alpha \) merges, and the corresponding feature \( F \) of \( \beta \) could enter into a checking relation with a sublabel of \( K \); and

(iv) there is no syntactic object \( \gamma \) such that \( \gamma \) has a feature \( F' \) which is of the same type as the feature \( F \) of \( \alpha \), and \( \gamma \) is closer to \( \alpha \) than \( \beta \) is.

If we assume the revised versions of Last Resort and Form Chain in (116) and (117), the highest copy of \textit{John} in (115b) cannot form a chain with the copy of \textit{John} in the embedded subject position, because the D-feature of the latter could not enter into a checking relation with the intermediate T head. Since the pair (\textit{John}^{i-D}, \textit{John}^{i-D}) in (115b) does not form a chain, no linked chain \( CH = (\textit{John}^{i-D}, \textit{John}^{i-D}, \textit{John}^{i-D}) \) can be formed either. The
topmost copy of *John* and either of the other copies will then induce a violation of the asymmetry condition on linear order, canceling the derivation, because no PF object is formed.

To summarize, Chomsky (1995:280) has proposed that a [+interpretable] feature remains unaffected by a checking operation, thus being able to participate in multiple checking relations. In the system I am proposing here, this amounts to saying that when participating in an overt checking relation, a [+interpretable] feature can optionally be checked with respect to PF, becoming invisible at this level. If it is checked, it patterns with checked [-interpretable] features in not being able to enter into any further checking relation; if it remains unchecked with respect to PF, it is allowed to enter into another checking relation. Since formal features (regardless of their interpretability at the C-I interface) must be eliminated in the mapping from N to π in order for a derivation to converge at PF, economy considerations dictate that two elements in an overt checking relation should have the greatest number of features checked with respect to PF (up to convergence). In other words, checking with respect to PF allows the number of applications of FF-Elimination targeting unchecked features to be minimized.

Thus, movement of the object to the Spec of TP to check the strong D-feature of T in (118) below also allows the moved element to have its D-feature checked with respect to PF, as illustrated in (119). Whether or not
this extra checking operation will yield a convergent derivation depends on
the kind of structure in which (119) is embedded. If (119) is embedded in a
structure which requires no further copies of John, as in (120), checking the
phonological part of the D-feature in the checking domain of the infinitival
T is not only allowed, but is actually forced by economy considerations. As
discussed in section III.6.2.2, the optimal PF realization of (120) will then be
the one resulting from the deletion of the trace of the DP-chain by Chain
Reduction, as shown in (121).

(118) \[ \text{TP John-D to } [\nu_P \text{ be kissed John-D }] \]

(119) \[ \text{TP John-D to } [\nu_P \text{ be kissed John-D }] \]

(120) \[ \text{TP I expected } \text{TP John-D to } [\nu_P \text{ be kissed John-D }] \]

(121) I expected John to be kissed.

By contrast, if (119) is embedded in a structure which requires that
further copies of John be created such as (110b), repeated below in (122a), no
convergent derivation is possible. As discussed above with respect to (111)
and (114), regardless of which instance of John is the source of the copy in
the upper subject position in (122b), the topmost copy of John cannot form a
chain with either of the lower copies, which then induces a violation of the asymmetry condition on linear order and cancels the derivation.

\[(122)\]  
\[a. \ [TP \text{ to be likely } [TP \text{ John}^{i-D} \text{ to } [\nu^P \text{ be kissed John}^{i-D} ] ] ] \]
\[b. \ [TP \text{ John}^{i-D} \text{ to be likely } [TP \text{ John}^{i-D} \text{ to } [\nu^P \text{ be kisses John}^{i-D} ] ] ] \]

Therefore, in a convergent derivation of successive cyclic movement, the [+interpretable] feature which enters into multiple checking relations must not be checked with respect to PF in intermediate positions. Rather than (122b), the computational system must have therefore assembled the structure in (123) below. Again, whether or not the D-feature of the highest copy of John in (123) can be checked depends on the type of structure (123) merges with. As before, if (123) is embedded in a structure such as the one in (124), the checking of this feature is forced by economy considerations.

\[(123)\]  
\[TP \text{ John}^{i-D} \text{ to be likely } [TP \text{ John}^{i-D} \text{ to } [\nu^P \text{ be kissed John}^{i-D} ] ] ] \]

\[(124)\]  
\[TP \text{ I expected } [TP \text{ John}^{i-D} \text{ to be likely } [TP \text{ John}^{i-D} \text{ to } [\nu^P \text{ be kissed John}^{i-D} ] ] ] ] \]

Regardless of which one of the lower instances of John is the source for the highest one in (124), the topmost copy of John can form the chain CH}
= (\text{John}^{i-D}, \text{John}^{i-D}) \) with the most embedded subject, which in turn can form the chain \( \text{CH}_2 = (\text{John}^{i-D}, \text{John}^{i-D}) \) with the embedded object. Given that the tail of \( \text{CH}_1 \) is the head of \( \text{CH}_2 \), the linked chain \( \text{CH}_3 = (\text{John}^{i-D}, \text{John}^{i-D}, \text{John}^{i-D}) \) can also be formed. If \( \text{CH}_3 \) is formed before Spell-Out, economy considerations concerning the application of FF-Elimination in the phonological component select the derivation in which Chain Reduction deletes the traces of \( \text{CH} \) (see section III.6.2.3), yielding the sentence in (125).

(125) I expect John to be likely to be kissed.

If either \( \text{CH}_1 \) or \( \text{CH}_2 \) is not formed overtly, the derivation will be canceled because the nondistinct copies of \text{John} will induce a violation of the asymmetry condition on linear order, making it impossible for a PF object to be formed. If both \( \text{CH}_1 \) and \( \text{CH}_2 \) — but not \( \text{CH}_3 \) — are formed overtly, the derivation will either be canceled or will yield the sentence in (125). If, for instance, Chain Reduction applies to \( \text{CH}_2 \) and deletes its head, yielding the structure in (126), the two remaining copies of \text{John}, which are not part of a chain, induce a violation of the asymmetry condition, preventing the structure from being linearized and canceling the derivation.

(126) \([\text{TP} \text{I expected} \ [\text{TP} \text{John}^{i-D} \text{to be likely} \ [\text{TP} \text{to} \ [\text{VP} \text{be kissed} \text{John}^{i-D} \ ] ] ] \) ]
If Chain Reduction deletes the tail of CH₂, yielding the structure in (127a), economy considerations concerning the number of applications of FF-Elimination will then force Chain Reduction to delete the tail of CH₁, yielding the structure in (127b), which will then surface with the PF output corresponding to (125). These considerations appear to make it possible to eliminate the notion of linked chains. I will leave this issue for future research.

(127)  

a. \[TP I expected [TP John^i-D to be likely [TP John^i-D to [vP be kissed ] ] ] ]

b. \[TP I expected [TP John^i-D to be likely [TP to [vP be kissed ] ] ] ]

III.6.3. On Why Strong Features Are Associated with Targets of Movement

Chomsky (1995:232) starts his discussion of strong features under the assumption that "[i]f F is strong, then F is a feature of a non-substantive category". The attempt to eliminate Agr-projections (see section II.7.3) in structures such as (128) below, however, leads Chomsky (1995:353) to assume that adjectives can also have a strong feature. The reason for this is that in languages with overt adjectival agreement, the structure illustrated in (129a) does not allow the adjective to check its [-interpretable] φ-features against its subject, given that checking relations and θ-relations are in complementary
distribution (see section II.7.5). Thus, adjectives in predicative constructions such as (129a) should have a strong feature licensing an extra Spec position as illustrated in (129b), where the subject of the adjective could move to and check the strong feature and the \(\phi\)-features of the adjective.\(^{52}\)

(128)  \[
\text{John} \_{i} \ {\text{is} \ [_{\text{AgrP}} \text{t} \_{i} \text{Agr} \ [_{\text{AP}} \text{t} \_{i} \text{intelligent}]}}
\]

(129)  a.  \[
\text{John} \_{i} \ {\text{is} \ [_{\text{AP}} \text{t} \_{i} \text{intelligent}]}
\]

b.  \[
\text{John} \_{i} \ {\text{is} \ [_{\text{AP}} \text{t} \_{i} [_{\text{A'}} \text{t} \_{i} \text{intelligent}]}}
\]

The generalization that seems to emerge from this discussion is that strong features are associated with targets of movement and not with moved elements. This generalization follows straightforwardly from the analysis developed here. Consider the abstract structure of the derivation in

\(^{52}\) Within the framework that I have developed here, where traces are as accessible to the computational system as heads of chains (see section III.4.1), one is not necessarily forced to conclude that in constructions such as (129a), the adjective has a strong feature. As Chomsky (1995:chap. 4, fn. 133) observes, it is not clear whether adjectives have an extra shell corresponding to the external vP of transitive constructions. Let us suppose that this is not the case with adjectives which take only one argument and that \textit{John} is the complement of \textit{intelligent} in (129a). Under the assumption that checking relations and \(\theta\)-relations are complementary (see Chomsky (1995:312) and section II.7.5), \textit{John} cannot check the \(\phi\)-features of the adjective in (129a), because they enter into a \(\theta\)-relation. At some point in the overt syntax, \textit{John} moves to check the strong feature of the T head. In the covert component, the formal features of the trace of \textit{John} then raise and adjoin to \textit{intelligent}, allowing the \(\phi\)-features of the latter to be checked. By taking traces to be accessible to the computational system and therefore able to move (see section III.4.1.4), no strong feature associated with adjectives in predicative constructions need be postulated in order to account for how the [-interpretable] \(\phi\)-features of the adjective are eliminated in the mapping from N to \(\lambda\).
which a term T which has a strong feature moves to the checking domain of the head H to check this strong feature:

(130)

\[ \begin{array}{c}
\text{HP} \\
\text{T'}, \text{H'} \\
\text{H}, \text{P} \\
\text{...T...}
\end{array} \]

In (130), H checks the strong feature of the upper copy of T, but not the strong feature of the lower copy because the latter is not in its checking domain (see section II.7.1). After (130) is spelled out, the derivation will crash at PF because the strong feature of the lower T has not been checked and strong features are not eliminable in the phonological component (see section II.6.2).

Therefore, the system in which only the chain link in the checking domain of a head H can be checked by a sublabel of H, such as the one I am arguing for in this dissertation, does not need to stipulate that strong features can only be associated with targets of movement.\textsuperscript{53} A system in

\textsuperscript{53} Of course, if a head H has a strong feature, it is allowed to move after its strong feature is checked, since the copy of H after the checking operation will also have a checked strong feature.

If both trivial and nontrivial head chains establish checking domains, as proposed by Chomsky (1993:10) (see section II.2.7.1), the system I am assuming here would also allow a head H with an unchecked strong feature to move, provided that a given element in the checking domain of the resulting nontrivial head chain CH = (Hᵢ, tᵢ)
which a checking relation between a chain link and a sublabel of a given head H affects both the link that is in the checking domain of H and its trace, as proposed by Chomsky (1995:chap. 4, fn. 12), must then stipulate that only targets of movement can bear strong features. Conceptual simplicity again leads us towards an asymmetric checking relation (see sections III.4.1.2 and III.6.2.3).

III.6.4. On Recoverability of Deletion

A system which allows information to be lost in the course of a derivation through deletion operations must ensure that such loss is severely constrained, as anecdotally shown in Fiengo and Lasnik (1972). In this sense, deletion is very different from Copy, for instance. "Unnecessary" applications of Copy cause the derivation to crash or be canceled, and therefore are detectable by the interpretive systems; "unnecessary" applications of deletion, on the other hand, may "fool" the system and filter out every feature that prevents convergence.

The system I am proposing here constrains applications of deletion by taking it to be part of the inner workings of the operations Chain

---

would be able to check the strong feature of CH. An example of this situation if found in Jonas and Bobaljik's (1993) analysis of English, according to which T raises overtly to Agrs independently of verb movement, and the subject then moves to Spec of Agrs, checking the relevant features of Agrs and the T-chain (including its strong D-feature). It is not clear, however, whether instances such as this one arise in a more restrictive system without Agr-projections.
Uniformization, Chain Reduction and FF-Elimination, as shown in (131)-(133) below.\(^{54}\)

(131) **Chain Uniformization:**
Delete the minimal number of features of a nontrivial chain CH in order to allow its links to satisfy the Feature Uniformity Condition.

(132) **Chain Reduction:**
Delete the minimal number of terms of a nontrivial chain CH which suffices for CH to be mapped into a linear order in accordance with the LCA.

(133) **Formal Feature Elimination (FF-Elimination):**
Given the sequence of pairs \(\sigma = <(F, P)_1, (F, P)_2, \ldots, (F, P)_n>\) such that \(\sigma\) is the output of Linearize, \(F\) is a set of formal features and \(P\) is a set of phonological features, delete the minimal number of features of each set of formal features in order for \(s\) to satisfy Full Interpretation at PF.

\(^{54}\) As mentioned in fn. 16, 43, and 48, the definitions in (131)-(133) may be simplified in that the number of elements to be deleted may be determined by economy considerations concerning the number of applications of the suboperation of deletion.
The operations in (131)-(133) have two points in common: (i) deletion is applied as little as possible (see fn. 60); and (ii) deletion always has a very specific goal to meet, be it satisfying the Feature Uniformity Condition or Full Interpretation at PF, or allowing a nontrivial chain to be mapped into a linear order. Being a suboperation constrained by economy considerations on the number of its applications, deletion is sufficiently constrained so that it is always recoverable.55

III.7. Apparent Counterexamples56

III.7.1. Wh-Copying

Some languages appear to allow traces of wh-phrases to be phonetically realized. The sentences in (134)-(138) exemplify such phenomenon:57

---

55 One could object that there are too many operations which take deletion as part of its inner workings. Although I in principle sympathize with the objection, it is worth pointing out that operations comparable to the one in (131)-(133) appear to be unavoidable within the framework discussed in chapter II. In Chomsky’s (1995:chap. 4) system, for instance, we find (i) the stipulation that the formal features of an intermediate trace are erased when a linked A-chain is formed (see section III.4.3.1); and (iii) the tacit assumption that the formal features with are shipped to the phonological component must be eliminated by PF, otherwise the derivation does not converge at this level (see section II.14).

56 Special thanks to Valentina Bianchi, Mirta Groppi, Eva Schlachter, and Raffaella Zanuttini, for discussion of relevant data.

57 I would like do thank Ger de Haan, Richard Kayne, Ana Perez-Leroux, and Craig Thiersch, for bringing some of the references on this subject to my attention. I would also like to thank Rozz Thornton for having clarified some points about the children's data discussed below.
Afrikaans (from du Plessis (1977)):

(134) *Met wie* het jy nou weer gesê *met wie* het Sarie met who did you now again said with who did Sarie gedog *met wie* gaan Jan trou? thought with who go Jan marry 'Whom did you say (again) that Sarie thought Jan is going to marry?'

Frisian (from Hiemstra (1986)):

(135) *Wa* tinke *jo* wa’*t* ik sjoen haw who think you who-that I seen have 'Who do you think that I have seen?'

German (from McDaniel (1986)):

(136) *Mit wen* glaubst du *mit wen* Hans spricht? with whom think you with whom Hans talks 'With whom do you think Hans is talking?'

Romani (from McDaniel (1986))

(137) *Kas* misline *kas* o Demiri dikhlâ? Whom you-think whom Demir saw 'Who do you think Demir saw?'
English child grammar (from Thornton (1990))

(138) Who do you think really who's in the can? (Tiffany 4;9)

If the terms with the same set of phonological features of the sentences above were related by a Copy operation, thus being nondistinct, they should induce violations of the irreflexivity and asymmetry conditions on linear order, canceling the derivation. If so, the acceptability of the sentences in (134)-(138) would be unaccounted for. The data above can however be compatible with the proposal about linearization of chains developed in this chapter, if the "repeated" elements in (134)-(138) are already present in the initial numeration, counting as distinct for purposes of linearization. Although a careful analysis of these constructions goes beyond the scope of this dissertation, two pieces of evidence suggest that the identical terms in (134)-(138) have not been related by the Copy operation.

The first piece of evidence concerns the generality of this pseudo-copying phenomenon. If the languages above allowed instances of phonetically realized traces, we should expect the traces of complex wh-phrases to be phonetically realized as well. As illustrated in (139) and (140) below, this is not the case, however. This pseudo-copying in general involves a single wh-word or (in some languages) a preposition plus a wh-word. This appears to suggest that the "repeated" terms in (133)-(138) are present in the initial numeration and encode some sort of wh-agreement,
rather than being introduced into the derivation by the Copy operation and forming a chain with nondistinct elements.

German (from McDaniel (1986)):

(139) *Wessen Buch glaubst du wessen Buch Hans liest?
whose book think you whose book Hans reads

'Whose book do you think Hans is reading?'

Romani (from McDaniel (1986)):

(140) *Save chave mislinea save chave o Demiri dikhlâ?
which boy you-think which boy Demir saw

'Which boy do you think Demir saw?'

The second piece of evidence comes from the fact that in some cases the apparent phonetically realized traces may actually have a different phonetic form. In German, for instance, the apparent wh-trace can be either the interrogative pronoun, as in (136), or the relative pronoun, as shown in (141) (see McDaniel (1986:125)).
German (from McDaniel (1986)):

(141) Mit *wem* glaubst du *mit den* Hans spricht?
with whom think you with whom Hans talks

'With whom do you think Hans is talking?'

A comparable state of affairs is found in Afrikaans. As shown in (142) below, a wh-word plus a pied-piped preposition can be replaced by a compound (see du Plessis (1977:724)). What is interesting for our purposes is that in a sentence such as (143), the lowest Spec of CP is occupied by either the compound form *waar* or the basic form *wat*, whereas the two upper Specs of CP are filled with the full compound form *waaroor*.

Afrikaans (from du Plessis (1977)):

(142) a. *Vir wat* werk ons nou eintlik?
for what work we now actually

'For what do we actually work?'

b. *Waarvoor* werk ons now eintlik?
wherefore work we now actually

'For what do we actually work?'
Afrikaans (from du Plessis (1977)):

(143) *Waaroor dink jy waaroor dink die bure*

whereabout think you whereabout think the neighbors

*wat/warr stry ons die meeste oor?*

what argue we the most about

'What do you think the neighbors think we are arguing about the most?'

The data in (141) and (143) show that at least for certain cases, this pseudo-copying phenomenon must be analyzed as involving different lexical items in the numeration. Given the idiosyncratic nature of this phenomenon, as discussed above, this conclusion can be generalized and we may take the "repeated" lexical elements in the data above to be all distinct in the initial numeration. If only one of these apparent copies is the "real" wh-phrase which has been moved, the other ones might be encoding some sort of wh-agreement. The fact that most of the times the wh-phrases are identical can be a reflex of the fact that agreeing forms generally share some common morphology.\(^{58}\)

---

\(^{58}\) Similar reasoning extends to instances of auxiliary doubling in English child grammar, as illustrated in (i) (from Guasti, Thornton, and Wexler (1995)). Rather than treating the lower auxiliary as a phonetically realized trace of the upper auxiliary, we may analyze the two instances as present in the initial numeration. Notice that in (i), it is even plausible that children are analyzing *didn't* as a single lexical item. At any rate, children also produce instances where the copies are distinct, as shown in (ii) (from Guasti, Thornton, and Wexler (1995)).
III.7.2. Clitic Reduplication

Some languages allow instances of clitic reduplication, which appear to suggest that the trace of the moved clitic may be overtly realized. The sentence in (144) below, which is acceptable in some dialects of Chilean Spanish (see Silva-Corvalán (1989)), exemplifies this phenomenon.

Chilean Spanish (from Silva-Corvalán (1989))

(144) Yo lo iba a matarlo.

I it was-going to kill-it

'I was going to kill it.'

If the second instance of *lo* in (144) were a copy of the upper instance, i.e. there is only one instance of *lo* in the initial numeration of the derivation of (144), the two nondistinct instances of the clitic should in principle yield violations of the irreflexivity and asymmetry conditions on linear order. There is reason to believe that the two instances of the clitic in (144) do not arise through copying, however. Were the lower instance of the clitic in (144) a copy of the upper, the two nondistinct instances of the clitic should in principle yield violations of the irreflexivity and asymmetry conditions on linear order. There is reason to believe that the two instances of the clitic in (144) do not arise through copying, however. Were the lower instance of the clitic in (144) a copy of the upper, the two nondistinct instances of the clitic should in principle yield violations of the irreflexivity and asymmetry conditions on linear order. There is reason to believe that the two instances of the clitic in (144) do not arise through copying, however. Were the lower instance of the clitic in (144) a copy of the upper, the two nondistinct instances of the clitic should in principle yield violations of the irreflexivity and asymmetry conditions on linear order. There is reason to believe that the two instances of the clitic in (144) do not arise through copying, however.

(i) a. What did he didn't wanna bring to school? (Darrell 4;1)
b. What kind of bread do you don't like? (Rosy 3;10)
c. Why could Snoopy couldn't fit in the boat? (Kathy 4;0)

(ii) a. What didn't Miss Piggy don't like to do? (Matt 4;3)
b. What do you can't eat?

59 I would like to thank Richard Kayne for bringing this type of construction to my attention.
clitic a phonetically realized trace, we should expect clitic reduplication to be parallel to regular instances of clitic movement in being subject to the same restrictions on clitic movement. As pointed out to me by Héctor Campos (p.c.), this is not the case. Contexts where clitic climbing is not possible may allow clitic reduplication, as shown in (145).

Chilean Spanish:

(145)  


it I-hate do-INF

'I hate to do it.'

b. Lo odio hacerlo.

it I-hate do-INF-it

'I hate to do it.'

It is reasonable to assume that it is irrelevant for the conditions governing applications of Form Chain before Spell-Out whether or not the nondistinct copies which are to form a chain will be deleted later on in the phonological component. If so, there should be no difference between the chain formed between the clitic and its trace in (145a) and between the apparent copies in (145b). The contrast in (145) thus suggests that (145b) does not involve an instance of clitic climbing with the trace being spelled out; rather it is probably the case that the upper instance of the clitic in (144) and
(145b) is an agreement morpheme of the upper verb.

Here I will not attempt to provide an analysis of the licensing conditions on this type of agreement or the interesting dialectal variation concerning clitic reduplication found in South American Spanish.⁶⁰ For our purposes, suffices it to note that once clitic reduplication does not go hand in hand with clear instances of clitic movement, multiple instances of identical clitics do not seem amenable to being analyzed as phonetically realized traces.

III.8. Conclusion

In this chapter I have argued in favor of the null hypothesis concerning the copy theory of movement in the Minimalist framework, namely, that traces and heads of chains do not have distinct intrinsic properties. In particular, I have shown that (i) traces are as accessible to the computational system as heads of chains (see section III.4.1); and (ii) traces do not differ from heads of chains regarding linearization (see sections III.4.2 and III.6.1). Differences between traces and heads of chains were shown to arise in the course of the derivation due to the fact that checking operations work very locally; a given head H can only enter into a checking relation with a chain link which is in its checking domain (see sections III.4.1 and

⁶⁰ I would like to thank Mirta Groppi, Héctor Campos, Marcela Depiante, and Beatriz Galdieri for helpful discussion about clitic reduplication in their dialects.
The reason why traces are not phonetically realized, as opposed to heads of chains, for instance, follows from the interaction of the fact that traces are subject to the LCA with economy considerations concerning the number of applications of deletion to eliminate unchecked formal features in the phonological component (the FF-Elimination operation; see section III.6.2.3). A syntactic object containing a nontrivial chain CH cannot be linearized in accordance with the LCA; since the links of CH are nondistinct (see sections II.10.4 and III.5), they induce a violation of the asymmetry condition on linear order, canceling the derivation, because no PF object is formed. In order to prevent this state of affairs, the phonological component can resort to the operation Chain Reduction, which deletes all but one link of a nontrivial chain.

Given that a given head only checks the relevant features of the chain link that is in its checking domain, the head of a chain CH will always have fewer unchecked formal features (if any) to be deleted by FF-Elimination than the lower links of CH. Thus, a derivation in which Chain Reduction deletes all the links except the head of the chain is always more economical than a derivation in which Chain Reduction deletes all of the links of the chain except one trace. Since FF-Elimination must delete the unchecked formal features (if any) of the link that survives in order for Full Interpretation to be satisfied at PF, the derivation in which the head of the
chain is the link that survives Chain Reduction requires fewer additional applications (if any) of deletion by FF-Elimination than the derivation in which a trace survives.

This analysis therefore allows us to dispense with the notion of a trace as a theoretical primitive. A trace is a descriptive notion which applies to a term with the following properties: (i) it is introduced into the derivation by the Copy operation; (ii) it is a member of a chain; and (iii) its deletion for linearization purposes is a more economical derivational step than the deletion of another link of its chain.

Under this analysis, the computations of the phonological component after the operations of the morphological subcomponent proceed as follows. If the phonological component has received from Spell-Out a syntactic object $\Sigma$ containing discontinuous objects (nontrivial chains), Chain Reduction must convert every nontrivial chain into a trivial chain; otherwise, these trivial chains cannot be mapped into a linear order. If possible, Linearize converts the structure resulting from applications of Chain Reduction into a linear order of pairs of sets phonological features and sets of formal features; otherwise, the derivation is canceled. If linearization is successful, FF-Elimination deletes unchecked formal features and the phonological rules can finally apply to the linear order of sets of phonological features, yielding

---

61 To be precise, it is targeted by the Copy operation at a given derivational step (see section II.10.4).
\( \pi \), which can be legitimate or not.

Appendix: Null Operator Constructions

In an attempt to provide an account for reconstruction effects found in null operator constructions, Munn (1994) proposes that null operators and their traces are copies of the element that licenses them. Under this view, the structure of the sentence in (146), for instance, which involves \textit{tough}-movement, should be as in (147b), rather than (147a).

(146) John is tough to please.

(147) a. \([\text{TP John} i \text{ is tough } [\text{CP O} i [\text{TP PRO to please } t i ] ]]\)

b. \([\text{TP John} i \text{ is tough } [\text{CP John} i [\text{TP PRO to please John} i ] ]]\)

As discussed in section III.6.2.2, Chain Reduction applies to nontrivial chains without taking into consideration the actual structure which contains those chains (see section III.6.2.2). Thus, if Munn's proposal were to be recast in terms of the analysis of trace deletion developed here, it would not be sufficient to simply say that the intermediate copy of \textit{John} in (147b) forms a chain with the lowest copy. When Chain Reduction applies to this chain, it will delete only one of the copies. The copy that survives Chain Reduction
and the copy in the matrix subject position will then induce violations of the irreflexivity and asymmetry conditions on linear order, preventing a PF object from being formed and canceling the derivation. Thus, if the null operator of (147a) is to be taken as a copy of the matrix subject, it must be the case that the two lower copies form a linked chain with the highest one. If this happens, the optimal reduction of the linked chain presumably involves the deletion of the two lower copies, yielding the sentence in (146).

In order for such an approach to succeed, there are several details to be spelled out. For instance, it must be shown that the kind of strong feature which the intermediate copy of John in (147b) checks cannot be checked by PRO. Otherwise, the Minimal Link Condition would prevent the two lower copies of John from forming a chain, which would then cause the derivation to be canceled later on; the lowest copy of John and either of the upper copies would induce violations of the irreflexivity and asymmetry conditions on linear order, making it impossible for the relevant structure to be linearized.

If it can be shown that PRO does not induce a Minimal Link Condition violation in (147b), the lack of tough-movement constructions in languages such as Greek and Bulgarian, for instance, can then be related to the fact that the infinitival constructions of these languages are inflected.62 If we have pro rather than PRO in the structures corresponding to (147b), pro would

---

62 I would like to thank Spyridoula Varlokosta for an illuminating discussion about tough-movement in Greek.
presumably have the feature which could enter into a checking relation with the embedded complementizer, thus preventing the two lower copies from forming a chain and causing the derivation to be canceled.

I will not pursue such an approach to tough-motion constructions and other null operator constructions in this dissertation (see Nunes (1995b) for discussion). I would just like to point out that if something along the lines suggested above is correct, it will enable us to account for why null operators are not phonetically realized: its deletion for purposes of linearization is more economical than another link of its chain, namely, the element that licenses it.
CHAPTER IV

SIDEWARD MOVEMENT AND

LINEARIZATION OF CHAINS AT PF

IV.1. Introduction¹

Let us recap the main features of the Copy + Merge theory of movement developed in chapters II and III. I have proposed that rather than being a primitive operation of the computational system, Move should be taken as a description of the complex interaction among the operations Copy, Merge, Form Chain, and Chain Reduction (see section II.10). In this system, nontrivial chains result from the application of the operation Form Chain in the course of the mapping from N to 0. Two nondistinct terms related by the operation Copy can form a chain if they satisfy the c-command condition on chain links, Last Resort and the Minimal Link Condition (see sections II.10.3, II.10.6 and III.6.2.5), as stated in (1):

¹ Part of the material discussed in this chapter has been presented at NELS 26 (MIT/Harvard University, October, 1995), at the International Conference on Interfaces in Linguistics (Universidade do Porto, November, 1995), and at the University of Southern California (Fall, 1995). I am thankful to these audiences.
(1) **Form Chain:**

The syntactic objects $\alpha$ and $\beta$ can form the nontrivial chain $\text{CH} = (\alpha, \beta)$ only if:

(i) $\alpha$ c-commands $\beta$;

(ii) $\alpha$ is nondistinct from $\beta$;

(iii) (at least) one given feature $F$ of $\alpha$ enters into a checking relation with a sublabel of $K$, where $K$ is the head (of the projection) with which $\alpha$ merges, and the corresponding feature $F$ of $\beta$ could enter into a checking relation with a sublabel of $K$; and

(iv) there is no syntactic object $\gamma$ such that $\gamma$ has a feature $F'$ which is of the same type as the feature $F$ of $\alpha$, and $\gamma$ is closer to $\alpha$ than $\beta$ is.

Pursuing the null hypothesis under the copy theory of movement, I have shown that traces and heads of chains do not differ in terms of their accessibility to the computational system or the conditions they are subject to (see section III.4). Differences between heads of chains and traces arise in the course of the derivation due to the fact that each link of a chain may participate in a different checking relation (see section III.6.2.3). Given a head
H and a chain CH, only with the link of CH which is in the checking domain of H may enter into a checking relation with H. Thus, the links of the chain CH = (he\textsuperscript{i}, he\textsuperscript{i}) in (2) below differ in terms of the features they have visible at the interface; the Case-feature of the upper copy has been rendered invisible by virtue of having entered into a checking relation with the Tense head \textit{will}, as represented by the subscript, whereas the lower copy of he remains unaffected by this checking relation.

(2) \[ [\text{TP he}^\text{-CASE} [\text{T will [VP be elected he}^\text{-CASE }]]] \]

The checking theory developed in chapters II and III is a reformulation of the one proposed in Chomsky (1995:chap. 4). The first difference between the two systems is that in the proposal developed here, a checking operation does not require further operations of erasure; a checking operation simply renders a given [-interpretable] feature invisible at the interface (see sections II.14 and III.6.2.4). The second difference is that checking theory is extended in the system proposed here in that overt checking relations may render features invisible at both interface levels, rather than being restricted to LF, as in Chomsky’s (1995:chap. 4) system (see sections III.6.2.4 and III.6.2.5). The basic idea behind this extension is that once formal features must be eliminated in the course of the mapping from N to \( \pi \) in order for the derivation to converge at PF, the system piggybacks on the existing checking
operation which makes features invisible at LF to render formal features invisible at PF as well. In this way, the system minimizes the number of applications of FF-Elimination, which is the operation of the phonological component which eliminates formal features in order for Full Interpretation to be ensured at PF, as stated in (3) (see chapter III: fn. 48).

(3) **Formal Feature Elimination (FF-Elimination):**

Given the sequence of pairs \( \sigma = <(F, P)_1, (F, P)_2, \ldots, (F, P)_n> \) such that \( \sigma \) is the output of Linearize, \( F \) is a set of formal features and \( P \) is a set of phonological features, delete the minimal number of features of each set of formal features in order for \( \sigma \) to satisfy Full Interpretation at PF.

Under this extension of the checking theory proposed in Chomsky (1995: chap. 4), overt checking operations render [-interpretable] features invisible at both LF and PF, and optionally render [+interpretable] features invisible at PF. Convergence conditions on chain formation and economy considerations regarding the number of applications of FF-Elimination determine when an overt checking relation renders a [+interpretable] feature invisible at PF. Assuming that a checked feature cannot enter into further checking relations regardless of the interface level at which it is
rendered invisible (see section III.6.2.5), a given [+interpretable] feature F of a
term T will be rendered invisible at PF by an overt checking operation only
if F's being checked will not prevent T from forming a chain with terms
nondistinct from T. Hence, the categorial feature of the topmost copy of he
in (4) below can be checked for PF purposes, but not the categorial feature of
the intermediate copy. Checking the D-feature of the former does not
prevent the linked chain (in the sense of Chomsky and Lasnik (1993:563))
CH₁ = (he₁-D-CASE, he₁-D-CASE, he₁-D-CASE) from being formed, as
represented in (4a). By contrast, if the D-feature of the intermediate copy is
checked, as represented in (4b), Last Resort blocks the formation of the chain
CH₂ = (he₁-D-CASE, he₁-D-CASE) between the two higher links (see (1iii)), and
the Minimal Link Condition prevents the chain CH₁ = (he₁-D-CASE, he₁-D-
CASE) from being formed between the highest and the lowest copy of he (see
(1iv)); consequently, no linked chain CH₃ = (he₁-D-CASE, he₁-D-CASE, he₁-D-
CASE) can be formed either.

(4)  a. [TP he₁-D-CASE [TP he₁-D-CASE to [VP be elected he₁-D-CASE ] ]]

     b. [TP he₁-D-CASE [TP he₁-D-CASE to [VP be elected he₁-D-CASE ] ]]

Assuming that traces and heads of chains are subject to the same
conditions and operations, there arises a problem with the linearization of the structure in (4a). Since the links of the linked chain \( \text{CH}_1 = (\text{he}^1\text{-D-CASE}, \text{he}^i\text{-D-CASE}, \text{he}^i\text{-D-CASE}) \) are nondistinct from one another, they induce violations of the irreflexivity and asymmetry conditions on linear order, making it impossible for (4a) to be linearized in accordance with Kayne's (1994) LCA. Since the highest copy of the pronoun \textit{he} asymmetrically \( c \)-commands the other two nondistinct copies, \textit{he} should precede itself; in turn, the adjective \textit{likely}, for instance, is required to precede \textit{he} since it asymmetrically \( c \)-commands the two lower copies of \textit{he}, but it is also required to be preceded by \textit{he}, given that it is \( c \)-commanded by the topmost nondistinct copy of \textit{he}. Once the structure in (4a) yields no linear order, no PF object is formed and the derivation is canceled (see Chomsky (1995:225) and sections II.5 and III.6.1); hence, the unacceptability of the sentence in (5), where all the links of the linked \( \text{CH}_1 \) in (4a) are phonetically realized.

\[
(5) \quad ^*\text{He is likely he to be elected he.}
\]

I have proposed that deletion of chain links in the phonological component is forced upon a nontrivial chain \( \text{CH} \) in order to allow the syntactic object containing \( \text{CH} \) to be linearized. I have referred to the operation which allows structures involving nontrivial chains to be mapped into a linear order as Chain Reduction, as formulated in (6) (see...
sections III.6.2.1 and III.6.2.2).

(6) Chain Reduction:

Delete the minimal number of terms of a nontrivial chain CH which suffices for CH to be mapped into a linear order in accordance with the LCA.

Applied to the linked chain $CH_1 = (he^{1-D\text{-}\text{CASE}}, he^{1-D\text{-}\text{CASE}}, he^{1-D\text{-}\text{CASE}})$ of (4a), Chain Reduction deletes two of its links, yielding either of the structures in (7). The choice between the derivations of the structures in (7) is determined by economy considerations concerning the number of applications of FF-Elimination which are eventually required to apply to the link that survives Chain Reduction. If the output of Chain Reduction is the structure in (7a), no further application of FF-Elimination is required because the D- and Case-features of he were rendered invisible at PF by a checking operation; the structures in (7b) and (7c), on the other hand, each require two applications of FF-Elimination targeting the unchecked features of the surviving copy. The more economical derivation in (7a) therefore rules out the ones in (7b) and (7c); hence, the pattern of acceptability found in (8).
(7)  
   a. \([TP \text{ he}^{i}\text{-D\text{-CASE}} [T \text{ is likely } [TP \text{ to } [VP \text{ be elected } ] ]]\]
   b. \([TP [T \text{ is likely } [TP \text{ he}^{i}\text{-D\text{-CASE}} \text{ to } [VP \text{ be elected } ] ]]\]
   c. \([TP [T \text{ is likely } [TP \text{ to } [VP \text{ be elected he}^{i}\text{-D\text{-CASE }} ] ]]\]

(8)  
   a. He is likely to be elected.
   b. *Is likely he to be elected.
   c. *Is likely to be elected he.

As stated in (6), Chain Reduction applies only to nontrivial chains. Thus, the optimal reduction of the structure in (4b), where only the lower copies of he can form a chain, is illustrated in (9). Given that the two nondistinct copies of he in (9) do not form a nontrivial chain, Chain Reduction is inapplicable. The two instances of he in (9) then induce violations of the asymmetry and irreflexivity conditions on linear order, canceling the derivation because no linear order can be established and no PF object can be formed. The relevance of this point will become clear as the discussion proceeds.

(9) \([TP \text{ he}^{i}\text{-D\text{-CASE}} [T \text{ is likely } [TP \text{ he}^{i}\text{-D\text{-CASE}} \text{ to } [VP \text{ be elected } ] ]]\]

Under this analysis, the notion of trace is not a primitive in the theory of grammar. Rather, a trace is a descriptive notion applied to a chain link
whose deletion for linearization purposes is a more economical derivational step than the deletion of another link of its chain, in that the number of applications of FF-Elimination is minimized. This is a significant achievement from a conceptual point of view, given that the copy theory of movement can be maintained in its simplest form and the lack of phonetic realization of traces is not stipulated in the theory.

A crucial difference between Chomsky's (1994, 1995:chap. 4) Move-based theory of movement and the Copy + Merge alternative proposed in chapters II and III is that in the latter, chain formation is not automatically associated with the sequence of derivational steps involving Copy and Merge. Rather, nontrivial chains result from applications of Form Chain in the course of the mapping from N to λ. The Copy + Merge theory of movement thus permits instances of "sideward movement", as represented in (10), where the term T of the syntactic object K is copied and merges with a syntactic object L, unconnected to K. At the derivational step represented in (10b), the pair (T₁, T₂) cannot form a chain because neither copy c-commands the other (see (1i)).

\[
\begin{array}{c}
\text{(10) a.} \\
\begin{array}{c}
\text{K} \\
\text{L} \\
\text{...T...}
\end{array}
\end{array}
\]

---

2 This strongly derivational approach is therefore immune to Brody's (1995) criticism against the redundancy between movement and chain formation.
One potential case of sideward movement concerning apparent noncyclic adjunction of relative clauses was discussed in section II.10.3. In sentences such as (11), for instance, the relative clause plausibly does not adjoin to the wh-phrase before it merges with the verb *like*; otherwise a Principle C effect would arise (see section III.2). Under the Copy + Merge theory of movement, a cyclic derivation for (11) is readily available, as shown in (12).

(11) Which portrait that Picasso painted did he like?

(12) a.  $K = [Q+ \text{did he like} \ [ \text{which portrait} \ ]]$

        $L = [\text{that Picasso painted}]$

b.  $K = [Q+ \text{did he like} \ [ \text{which portrait}^i \ ]]$

        $L = [\text{that Picasso painted}]$

        $M = [\text{which portrait}^i]$

---

3 See Uriagereka (forthcoming:chap. 4), who discusses the possibility of a sideward movement analysis of overt head movement.
(13) \[ O' = \left[ \left[ \left[ \text{which portrait} \right] \right] \left[ \text{that Picasso painted} \right] \right] \left[ \text{did he like} \right] \left[ \left[ \text{which portrait} \right] \right] \]
After (13) is shipped to the phonological component, the optimal reduction of the CH = ([which portrait ]-WH, [which portrait ]-WH) is the one which requires no application of FF-Elimination for the wh-feature of the copy that survives Chain Reduction (see section III.6.2.3), namely, the one which converts (13) into (14). After further computations, (14) yields the PF output associated with the sentence in (11).

\[ \text{O''} = [ [ [ \text{which portrait } ]^{i-\text{WH}} [\text{that Picasso painted } ] ] \text{ Q+did he like } ] \]

In this chapter, I provide additional evidence for the Copy + Merge theory of movement by showing that parasitic gap constructions and constructions involving across-the-board (ATB) extraction (see Ross (1967) and Williams (1978), for instance) can receive a straightforward analysis in terms of sideward movement. Various researchers have proposed that these two constructions should be treated alike (see Pesetsky (1982), Haïk (1985), Williams (1990), and Munn (1993), among others). Although I follow these authors in attempting to provide a unified account of these constructions, I will not attempt to assimilate one construction to the other. I will instead submit both parasitic gap and ATB constructions to the same analysis of "standard" movement operations developed in chapters II and III, and reviewed above. To the extent that sideward movement is constrained
enough to allow parasitic gaps and ATB gaps to be analyzed as "regular" traces, the Copy + Merge theory of movement receives strong conceptual and empirical support.

The chapter is organized as follows. In section IV.2, I review how the Copy + Merge theory of movement, combined with the analysis of deletion of chain links for purposes of linearization, can provide a straightforward account of the properties standardly attributed to parasitic gap constructions. In section IV.3, this analysis is extended to ATB constructions. In section IV.4, I discuss the requirement in languages with a rich morphological case system that the relevant gaps in parasitic gap and ATB constructions match in terms of morphological case. Finally, some conclusions are presented in section IV.5.

IV.2. Parasitic Gaps

(1991), Postal (1993, 1994), and Manzini (1994, among others).\(^4\) (16) below summarizes some of the properties of parasitic gap constructions discovered in this literature.

\begin{enumerate}
    \item Which paper did you file without reading?
    \item Which paper did you file without reading PG\(_i\)?
\end{enumerate}

(16) \begin{enumerate}
    \item Parasitic gaps are licensed at S-Structure (see Chomsky (1982), for instance).
    \item Parasitic gaps cannot be c-commanded by the "real" gap (see Taraldsen (1981) and Engdahl (1983)).
    \item Parasitic gaps are subject to island effects (see Kayne (1983, 1984), and Chomsky (1986a), for instance).
    \item Parasitic gaps can only be NPs (see Chomsky (1986a) and Cinque (1990), for instance).
    \item Parasitic gaps cannot be licensed by nonreferential NPs (see Cinque (1990)).
\end{enumerate}

Several different analyses have been put forward to account for the

\footnote{4 For expository reasons, I will focus on examples of parasitic gaps licensed by wh-movement in questions. The analysis proposed for this type of parasitic gap extends \textit{mutatis mutandis} to instances licensed by wh-movement in relative clauses and tough-movement. Where a detailed structure is not necessary, I will represent the "real" gap by \(t\) and the parasitic gap by \(PG\).}
properties listed in (16). Parasitic gaps have been analyzed as (i) traces of across-the-board extraction (see Williams (1990), for instance); (ii) traces of wh-phrases which are not the result of movement (see Frampton (1990), for instance); (iii) null resumptive pronouns (see Cinque (1990), for instance); and (iv) traces resulting from movement of null operators (see Chomsky (1986a), for instance). The proposal in (iv) in turn has different implementations: for Lasnik and Stowell (1991), the trace of the kind of null operator that appears in parasitic gap constructions is a null epithet; and for Browning (1987) the null operator itself is pro, whereas for Weinberg (1988) it is PRO.

Some of the proposals mentioned above are partially or totally incompatible with the Copy + Merge theory of movement proposed here, according to which movement is not an operation of the computational system, but rather the description of the interaction of the operations Copy, Merge, Form Chain, and Chain Reduction (see sections II.10 and III.6). If this approach to movement is on the right track, it rules out analyses which crucially depend on the notion of movement as a theoretical primitive, such as the analysis of parasitic gaps in terms of across-the-board extraction, which allow multiple identical constituents to move from multiple positions "at once" (see Williams (1978, 1990), for instance). I will show below that parasitic gap and ATB extraction constructions can be analyzed in terms of the interaction of operations which are responsible for standard instances of
movement, without enriching the set of operations of the computational system.

The Copy + Merge theory of movement is also incompatible with analyses which take a parasitic gap to be a special kind of trace which does not result from movement (see Frampton (1990), for instance). Under the analysis developed in Chapter III, traces are not theoretical primitives; rather, trace is a descriptive notion which applies to a chain link whose deletion for linearization purposes is a more economical derivational step than the deletion of another link of its chain, in that it minimizes the number of applications of FF-Elimination. Within the framework developed here, to say that a parasitic gap is a trace which does not arise through movement presumably amounts to saying that it is not introduced into the derivation by Copy. However, if a new term is introduced into the derivation by Select or Merge, it will be interpreted as distinct from every other term in the derivation, thus being unable to form a chain with any other term (see sections II.10.4 and III.5). In the analysis of parasitic gap constructions to be developed in the following sections, parasitic gaps will be treated as regular traces.

Also incompatible with the system advocated here or with any version of the copy theory of movement, are analyses in which the links of a chain may be different kinds of syntactic objects, such as the analysis proposed by Lasnik and Stowell (1991). They claim that the trace of a non-quantificational
null operator, such as the one found in parasitic gap constructions, is a null epithet, a new type of empty category. This analysis attempts to capture the fact that parasitic gaps appear to pattern with epithets in inducing an amelioration of weak cross over effects, as illustrated in (17) (Lasnik and Stowell's (20b) and (66b))

(17) a. [ who, did you stay with ti [ before [ his, wife ] had spoken to PGi ] ]
   b. [ which assailant, did Mary escape from ti [ before [ his, partner ] joined up with [ the bastard ]i ] ]

Hornstein (1995:chap. 6) however shows that the amelioration effect with respect to weak cross over found in (17a) is not restricted to constructions involving nonquantificational null operators and may arise even in constructions involving true quantificational phrases, as illustrated in (18) (Hornstein's (20a) and (20b)).

(18) a. *[ [ his, mother ] gave [ his, picture ] to [ every student ]i ]
   b. [ [ his, mother ] gave [ every student ]i [ his, picture ] ]

Hornstein (1995:chap. 6) provides a uniform account of data such as (17) and (18) by analyzing the weak cross over restriction in terms of Linking
Theory (see Higginbotham (1983, 1985), as stated in (19):\(^5\)

(19)  \textit{Weak Cross Over Principle:}

A pronoun cannot be linked to a variable on its right.

According to (19), the reason why (17a), (17b), and (18b) allow a bound reading of the relevant pronoun without inducing a Weak Cross Over effect is that the pronoun need not be linked to a variable on its right in order to be interpreted as a bound pronoun. Thus, the pronoun of each sentence of (17) can be linked to the variable on its left, and the pronoun inside the subject in (18b) can be linked to the pronoun inside the object; in (18a), on the other hand, one of the pronouns must be linked to a variable on its right in order for the relevant interpretation to obtain, violating (19).

The point of this short digression is to show that data such as (17) do not necessarily require an expansion of the inventory of empty categories or additional assumptions regarding chain formation in order to allow the postulated chain between the null operator and parasitic gap. I will therefore keep to the simplest assumption that two terms can form a chain only if they are nondistinct from one another; a parasitic gap will thus be taken to be nondistinct from the terms it forms a chain with.

The analyses that are relevant for a direct comparison with the analysis of parasitic gaps in terms of sideward movement and linearization of chains which I propose below, are therefore the ones which take parasitic gaps to be (regular) traces of null operators, and the ones which treat parasitic gaps as null resumptive pronouns. Here, I do not intend to provide a comprehensive review of the literature associated with either approach. I will limit myself to highlighting the large conceptual and empirical differences between these two approaches and the sideward movement approach with respect to the properties listed in (16). For the sake of discussion, I will take Chomsky (1986a) and Cinque (1990) as standard representatives of the null operator and the null resumptive pronoun approaches, respectively.

The discussion is organized as follows: in section IV.2.1, I outline a general analysis of parasitic gaps in terms of sideward movement, which is made available by the Copy + Merge theory of movement, combined with the proposal about linearization of chains discussed in section III.6. Sections IV.2.2 to IV.2.4 then address each of the properties listed in (16) and compare the analysis of parasitic gaps in terms of sideward movement with the analyses in terms of null operator movement or null resumptive pronouns.
IV.2.1. General Approach

Let us now see how the parasitic gap construction given in (15) and repeated in (20) below, can be derived in the system developed in chapters II and III.

(20) \[ [ \text{which paper}_i \text{ did you file}_t \text{ without reading PG}_i ] \]

Suppose that the computational system starts with the initial numeration N in (21) below and operates until the derivational stage where N has been reduced to N' in (22a) and the two syntactic objects K and L in (22b) have been assembled through several applications of Select and Merge.

(21) \[ N = \{\text{which}_{1}, \text{paper}_{1}, Q_1, \text{you}_{1}, \text{did}_{1}, v_2, \text{file}_{1}, \text{without}_{1}, \text{PRO}_{1}, T_1, \text{reading}_{1}\} \]

(22) a. \[ N' = \{\text{which}_{0}, \text{paper}_{0}, Q_1, \text{you}_{0}, \text{did}_{1}, v_1, \text{file}_0, \text{without}_{1}, \text{PRO}_{1}, T_1, \text{reading}_0\} \]
Presumably, possible continuations of (22) involving merger of L with K or any object formed from the lexical items still available in $N'$ do not converge. However, the derivational step sketched in (22) may lead to a convergent result if the computational system makes a copy of *which paper* and merges it with *reading* (an instance of sideward movement), as illustrated in (23).

The derivation proceeds until the numeration $N'$ in (22a) is reduced to
N'' in (24a) and the objects K in (23) and O in (24b) are assembled.\textsuperscript{6,7}

\begin{equation}
(24) \quad \begin{aligned}
a. \quad N'' &= \{\text{which}_0, \text{paper}_0, \text{Q}_1, \text{you}_0, \text{did}_1, \text{v}_0, \text{file}_0, \text{without}_0, \text{PRO}_0, \\
&\quad \quad \quad \text{T}_0, \text{reading}_0\} \\

b. \quad O &= \text{PP} \\
&\quad \quad \quad \text{without} \quad \text{TP} \\
&\quad \quad \quad \text{PRO}' \quad \text{T}' \\
&\quad \quad \quad \text{T} \quad \text{vP} \\
&\quad \quad \quad \text{PRO}' \quad \text{v}' \\
&\quad \quad \quad \text{v} \quad \text{VP} \\
&\quad \quad \quad \text{L} = \text{reading} \quad [\text{which paper}]_i
\end{aligned}
\end{equation}

The next derivational step involves adjunction of O to K, as shown in (25).\textsuperscript{8}

\textsuperscript{6} Nothing would essentially change IF the computational system had first assembled the PP in (24b) and then copied the wh-phrase, merging it with file and forming K in (22b).

\textsuperscript{7} Recall that PRO moves to Spec of TP to check the strong D-feature of T. The checking configuration between the upper copy of PRO in (24b) and the nonfinite T also allows them to check their null Case-features (see Chomsky and Lasnik (1993:561) and Martin (1992)).

\textsuperscript{8} For expository purposes, I will assume here that O in (25) adjoins to the projection of the light verb shell of the matrix clause. In section IV.2.3 below, I discuss
Notice that in (25), no chain formation between the nondistinct copies of *which paper* can take place. The pair ([ which paper ]', [ which paper ]) satisfies neither the Last Resort condition (none of the copies of *which paper* enters into a checking relation) nor the c-command condition on Form Chain, which is repeated below in (26).

---

the issue of the height of the adjunction site and the issue of whether the PP should be taken as a complement (see Larson (1988)) in order for the linear order between the matrix vP and the PP to be accounted for, in accordance with Kayne's (1994) LCA.
(26) *Form Chain:*

The syntactic objects $\alpha$ and $\beta$ can form the nontrivial chain $CH = (\alpha, \beta)$ only if:

(i) $\alpha$ c-commands $\beta$;

(ii) $\alpha$ is nondistinct from $\beta$;

(iii) (at least) one given feature $F$ of $\alpha$ enters into a checking relation with a sublabel of $K$, where $K$ is the head (of the projection) with which $\alpha$ merges, and the corresponding feature $F$ of $\beta$ could enter into a checking relation with a sublabel of $K$; and

(iv) there is no syntactic object $\gamma$ such that $\gamma$ has a feature $F'$ which is of the same type as the feature $F$ of $\alpha$, and $\gamma$ is closer to $\alpha$ than $\beta$ is.

After the remaining lexical items of $N''$ in (24a) are pulled out and merge with the structure in (25), the structure in (27a) is derived (copies irrelevant for the discussion were omitted). Under the assumption that the complementizer $Q$ in (27a) has a strong wh-feature, another copy of *which*
paper is made and merges with the structure in (27a), forming (27b).⁹

(27)  a. [\text{CP} \text{Q+did} [\text{TP} \text{you} [\nuP [\nuP \text{file} [\text{which paper }]^i] [\text{PP} \text{without PRO reading} [\text{which paper }]^i]]]]

b. [\text{CP} [\text{which paper }]^i \text{Q+did} [\text{TP} \text{you} [\nuP [\nuP \text{file} [\text{which paper }]^i]] [\text{PP} \text{without PRO reading} [\text{which paper }]^i]]]]

Given that no further checking relation involves the wh-feature of which paper, convergence conditions do not prevent the copy in the Spec of CP from having its wh-feature checked for PF purposes when it enters into a checking relation with the strong feature of Q. Economy considerations minimizing the number of applications of FF-Elimination then ensure that this feature is so checked (see sections III.6.2.5 and IV.1). Thus, a more appropriate representation of (27b) is as in (28).

(28)  [\text{CP} [\text{which paper }]^\text{-WH} \text{Q+did} [\text{TP} \text{you} [\nuP [\nuP \text{file} [\text{which paper }]^\text{-WH}] [\text{PP} \text{without PRO reading} [\text{which paper }]^\text{-WH}] ]]]

⁹ Since no features of either instance of which paper in (27a) were checked, it does not matter which of the instances is copied to form (27b). See section IV.2.3 below, where the choice of the term to be copied has consequences in terms of convergence and economy.
In (28), the copy of *which paper* in the Spec of CP can in principle form a chain with either of the lower copies (see Frampton (1990), for instance). In each case, the two instances of *which paper* are nondistinct (see (26ii)); one copy c-commands the other (see (26i); the copy in the Spec of CP enters into a checking relation (see (26iii)); and there is no term containing a wh-feature which is closer to the Spec of CP than either of the lower copies is (see (26iv)). Crucially, closeness is defined in terms of c-command, as shown in (29); thus, given that neither of the lower copies c-commands the other, one is not closer to the copy in the Spec of CP than the other is. Finally, if the 0-Criterion is to be interpreted within the Minimalist framework as a condition on 0-role assignment, as I argued for in section II.16, the wh-phrase in the Spec of CP in (28) is allowed to be part of both chains, regardless of the 0-role of each chain.

\[(29) \text{ Closeness:} \]

Where \( \gamma \) c-commands \( \alpha \) and \( \alpha \) c-commands \( \beta \), \( \alpha \) is closer to \( \gamma \) than \( \beta \) is iff:

(i) \( \alpha \) c-commands \( \beta \); and

(ii) \( \alpha \) is not in the same minimal domain as \( \gamma \) or \( \beta \).

Recall that Form Chain is an operation which applies in the course of
the mapping from N to \( \lambda \); although the phonological component is able to differentiate nondistinct terms from terms with identical sets of features, it presumably does not form chains (see section III.5). Let us then consider the situation where the copy of *which paper* in the Spec of CP in (28) has not formed a chain with either of the lower copies prior to Spell-Out. In this case, the phonological component will treat each of the instances of *which paper* as a trivial chain. Since these instances are nondistinct, they induce violations of the irreflexivity and asymmetry conditions on linear order (see sections III.6.1 and IV.1), preventing (28) from being linearized in accordance with the LCA. Once no linear order is established, no PF object is formed and the derivation is canceled.

Notice that if two of the wh-phrases of (28) were deleted in the phonological component, the derivation could converge at PF even if no wh-chain had been formed before Spell-Out. However, the system proposed in chapter III does not take deletion of terms as a general condition on convergence; rather, deletion of terms is taken to be part of the inner workings of Chain Reduction, repeated below in (30), which in turn operates with nontrivial chains, not with mere sequences of nondistinct terms (see sections III.5, III.6.2.2, and III.6.5). Thus, if no wh-chain has been formed overtly in (28), each wh-phrase will be treated by the phonological component as a trivial chain and Chain Reduction is inapplicable. The advantages of restricting deletion of terms in this fashion will become
evident as we proceed.

(30)  *Chain Reduction:*

Delete the minimal number of terms of a nontrivial chain CH which suffices for CH to be mapped into a linear order in accordance with the LCA.

Let us now consider the possibility under which only the chain between the wh-phrase in the Spec of CP in (28) and the copy in the object of *file* has been formed before Spell-Out. After (28) is shipped to the phonological component, the optimal reduction of this chain involves the deletion of the wh-phrase in the object position, since the upper copy requires no application of FF-elimination targeting its wh-feature (see sections III.6.2.3 and IV.1). However, the structure resulting from applying Chain Reduction to this chain cannot be linearized, as shown in (31) below. Since the copy in Spec of CP and the copy within the PP did not form a chain before Spell-Out, Chain Reduction is inapplicable; these two nondistinct copies then induce violations of the irreflexivity and asymmetry conditions on linear order, preventing the structure in (31) from being linearized and canceling the derivation.
A similar state of affairs arises if only the chain between the copy in the Spec of CP and the object of reading in (28) is formed before Spell-Out. After this chain is reduced in the phonological component as in (32), the two surviving nondistinct copies induce violations of the irreflexivity and asymmetry conditions on linear order, and the derivation is canceled.

\[
(31) \quad \left[ \text{CP [ which paper ]} \right]^{i-WH} \text{Q+did } [\text{TP you } [\text{VP file } [\text{PP without PRO reading [ which paper ]}^{i-WH} ] ] ]}
\]

Notice that if Chain Reduction were allowed to employ global computations, taking into consideration the linearization of the structure containing the nontrivial chain to be reduced, as stated in (33) below, for instance (see discussion in section III.6.2.2), it would make wrong empirical predictions for the last two scenarios discussed above.

\[
(32) \quad \left[ \text{CP [ which paper ]} \right]^{i-WH} \text{Q+did } [\text{TP you } [\text{VP file [ which paper ]}^{i-WH} ] ] [\text{PP without PRO reading } ] ] ]
\]

---

\[10\] See section IV.2.4 for some discussion about chain formation across adjunct islands.
Chain Reduction:

Delete the minimal number of terms of a nontrivial chain CH in order for the structure containing CH to yield a linear order in accordance with the LCA.

If only a wh-chain had been formed in (28) before Spell-Out, the "global" version of Chain Reduction in (34) would take into consideration the linearization of the structure of (28) as a whole and would delete both links of the relevant chain, allowing (28) to comply with the irreflexivity and asymmetry conditions on linear order. Thus, the global version of Chain Reduction in (33) would convert the structure in (28) into either (34a) or (35a) depending on which wh-chain had been formed. However, both derivations yield unacceptable sentences, as shown in (34b) and (35b).

(34)  a. \([CP \ Q+did \ [TP \ you \ [VP \ [TP \ file] \ [PP \ without \ PRO \ reading] \ [which \ paper]i-WH] \]]\)

b. *Did you file without reading which paper?

(35)  a. \([CP \ Q+did \ [TP \ you \ [VP \ [TP \ file \ [which \ paper]i-WH] \ [PP \ without \ PRO \ reading] \]]\]

b. *Did you file which paper without reading?
The unacceptability of the sentences in (34b) and (35b) provide additional evidence for the "local" version of Chain Reduction given in (30), which deletes terms of a nontrivial chain in such a way that the surviving terms could in principle be mapped into a linear order in accordance with the LCA, without considering whether an actual linearization obtains. Recall that since Form Chain requires that the links of a nontrivial chain be nondistinct terms in an asymmetric c-command relation (see discussion in section III.6.2.2), they cannot be mapped into any linear order which conforms with the LCA, because they violate the irreflexivity condition on linear order. Thus, the information that a nontrivial chain cannot be mapped into a linear order in accordance with the LCA is available to the computational system, regardless of the structure which actually contains the nontrivial chain. Chain Reduction then simply deletes terms of CH in such a way that the irreflexivity condition on linear order is satisfied; that is, either the surviving terms of each chain link are distinct or only one chain link survives (scattered and full deletion, respectively; see section III.6.2.1).

The system adopted here, which incorporates the "local" version of Chain Reduction in (30), therefore requires that in order for a continuation of (28), repeated below in (36), to yield a PF object, the wh- phrase in Spec of CP forms a distinct chain with each of the two lower copies before Spell-Out. If this happens, Chain Reduction in the phonological component will apply to each chain individually. As seen above, if applied to the chain formed by
the highest copy and the one inside the adjunct, for instance, the structure in (37) is derived.

(36) \[ CP [ which paper ]^l\text{-WH} Q+\text{did} [ TP you [ vP [ vP file [ which paper ]^l\text{-WH} ] [ PP without PRO reading [ which paper ]^l\text{-WH} ] ] ] ]

(37) \[ CP [ which paper ]^l\text{-WH} Q+\text{did} [ TP you [ vP [ vP file [ which paper ]^l\text{-WH} ] [ PP without PRO reading [ which paper ]^l\text{-WH} ] ] ] ]

The optimal application of Chain Reduction to the other wh-chain in (37) then yields the structure in (38) below. (38) can be linearized in accordance with the LCA without giving rise to violations of the irreflexivity or asymmetry conditions on linear order. After further computations in the phonological component, (38) will have the PF output associated with the sentence in (39).\(^\text{11}\)

(38) \[ CP [ which paper ]^l\text{-WH} Q+\text{did} [ TP you [ vP [ vP file [ PP without PRO reading ] ] ] ]

\(^{11}\) For expository purposes, I will assume in this chapter that vP precedes its adjunct in structures such as (38). See fn. 19 below for further discussion.
(39) Which paper did you file without reading?

The Copy + Merge theory of movement, which allows constrained instances of sideward movement, combined with the analysis of deletion of chain links in terms of linearization, thus provides a straightforward analysis of parasitic gaps. Under the approach outlined above, no construction-specific mechanism was employed and the inventory of operations and categories was kept constant. A chain formed before Spell-Out which has a "parasitic gap" as one of its links must undergo Chain Reduction in the phonological component in order for it to be linearized like any other chain; hence, a parasitic gap is not phonetically realized for the same reason traces are not phonetically realized: its deletion is more economical than the deletion of another link of its chain. If the "parasitic gap" copy does not form a chain with another term before Spell-Out, the structure cannot be linearized in compliance with the irreflexivity and asymmetry conditions on linear order and the derivation will be canceled.

The two instances of the wh-phrase in an acceptable sentence such as (40) below therefore must have been formed by distinct applications of Select and Merge and not by the Copy operation. The fact that the two wh-phrases are interpreted as distinct shows that this is actually the case.

(40) Which paper did you file after reading which paper?
The generalization that emerges from this picture is that given the nontrivial chains CH₁ and CH₂, a licit parasitic gap construction can arise only if (i) the links of CH₁ and the links of CH₂ are nondistinct; (ii) CH₁ and CH₂ have been formed before Spell-Out; and (iii) CH₁ and CH₂ have one link in common.

Compare these results with the way competing analyses derive the lack of phonetic realization of the "parasitic chain". Since analyses based on null operator movement take a parasitic gap to be a trace of a null operator, they appear to conform with the analysis I proposed above in reducing the lack of phonetic realization of the parasitic gap to the lack of phonetic realization of traces in general. However, there is a crucial difference between these two approaches. In order to completely derive the lack of phonetic realization of the parasitic gaps, the null operator approach must first explain why the null operator of parasitic gap constructions does not alternate with an overt operator, as illustrated in (41). Notice that null operators do alternate with overt operators in some constructions, as illustrated by the relative clauses in (42) (modulo the restriction on doubly filled Comps).

(41)  
   a. which paper did you file [ O₁ after reading t₁ ]
   b. *which paper did you file [ which₁ after reading t₁ ]
Thus, although the null operator approach may have an answer for why the parasitic gap is not phonetically realized, the reason why the head of the "parasitic chain" in sentences such as (41b) cannot be pronounced still remains to be derived.\footnote{Munn's (1994) analysis of parasitic gaps, in which the null operator is a copy of the element that licenses it, may be immune to this criticism, if the "null operator" copy forms a chain with the element that licenses it. See the appendix of chapter III and fn. 26 below.} Under the analysis of parasitic gaps in terms of sideward movement and linearization of chains, on the other hand, the contrast in (41) presents no problem. The "parasitic gap" of (41a) forms a chain directly with which paper, without the mediation of any other element; hence, the reason for the unacceptability of (41b) is the same for the unacceptability of the sentences in (43), where no parasitic gap is involved.

(43)  

\begin{enumerate}
\item \textbf{a.} *Which article did you read which after buying that book?  
\item \textbf{b.} *I read this article which after buying that book.  
\end{enumerate}

Analyses which treat parasitic gaps as null resumptive pronouns are subject to a similar criticism. If overt and null resumptive pronouns differ only in their set of phonological features, as seems reasonable, a parasitic gap should then freely alternate in being realized as an overt or a null
resumptive pronoun. From a Minimalist perspective, this point is even more relevant; after Spell-Out, overt and null resumptive pronouns should in principle be identical for any computation in the covert component or for interpretation at the C-I interface. As pointed out by Engdahl (1983), however, there are some environments which do not allow overt resumptive pronouns, but do allow parasitic gaps. Thus, the lack of alternation between an overt and a null resumptive pronoun in constructions such as (44), for instance, poses a potential problem for analyses of parasitic gaps in terms of resumptive pronouns.

(44)  
   a. Which man does everyone who meets ends up liking?  
   b. [ [ which man ], does [ [ everyone ] [ who meets PGt ] ends up liking ti ] ] 
   c. *Which man does everyone who meets him ends up liking?  
   d. [ [ which man ], does [ [ everyone ] [ who meets him ] ends up liking ti ] ] 

   I do not intend to claim with the above remarks that it is not possible to assign distinct features to overt and null resumptive pronouns and then obtain different interpretations at LF.13 All that I am saying is that first, it is

13 Cinque (1990:155), for instance, suggests that overt resumptive pronouns give rise to (stronger) weak cross over effects, as opposed to null resumptive pronouns (see also Chomsky (1982:44)).
not obvious how to do it in a principled fashion and in compliance with Minimalist guidelines; and second, that the analysis I proposed above is exempt from the potential problem of the distribution of overt or null resumptive pronouns, because it derives the lack of phonetic realization of parasitic gaps by treating them as regular traces.\textsuperscript{14}

Let us now examine how the analysis of parasitic gaps in terms of sideward movement and linearization of chains that I am proposing is able to account for the properties listed in (16).

**IV.2.2. S-Structure Licensing**

In pre-Minimalist analyses of parasitic gaps, the contrast between (45a), on the one hand, and (45b) and (45c), on the other, was taken to show that a parasitic gap is licensed at S-Structure (see Chomsky (1982), for instance). Given that the only syntactic levels of representation within the Minimalist framework are the interface levels LF and PF, the contrast in (45) presents a challenge for any analysis of parasitic gaps within the Minimalist Program.

\textsuperscript{14} Incidentally, if Hornstein's (1995) analysis of weak cross over (see section IV.1) is on the right track, it would exclude a null resumptive pronoun approach to parasitic gap constructions such as (44a). In order to be interpreted as a bound pronoun, the alleged resumptive pronoun of (44c) would have to be linked to a variable on its right, giving rise to a weak cross over violation (see (19)). The analysis proposed here, which takes the parasitic gap in (44a) to be a regular trace, is immune to this potential problem.
(45)  
   a. Which report did you file without reading?
   b. *John filed every report without reading.
   c. *Who filed which report without reading?

The analysis of parasitic gaps I am pursuing here has a simple account of the unacceptability of (45b) and (45c). Under the relevant reading, (45c), for example, is to be derived by sideward movement of which report from the object of filed to the object of reading (see fn. 6). That is, at a given derivational point where the syntactic objects K and L in (46) below have been formed, the computational system makes a copy of the instance of which report in the object of filed and merges it with reading, yielding the objects in (47). After further computations, the (simplified) structure in (48) is finally assembled.

(46)  
   a. K = \[v_p \text{ filed } [\text{ which report }]\]
   b. L = reading

(47)  
   a. K = \[v_p \text{ filed } [\text{ which report }]\]
   b. M = \[\text{ reading } [\text{ which report }]\]

(48)  
[\text{ who } [v_p \text{ filed } [\text{ which report }]] [PP \text{ without PRO reading } [\text{ which report }]]]
The pair ([ which report ]i, [ which report ]i) in (48) cannot form a chain, because it does not satisfy Last Resort or the c-command condition on Form Chain (see (26iii) and (26i)). The unacceptability of (45c) then follows: deletion of the second wh-phrase of (48) is unmotivated. Within the system I have proposed in chapter III, deletion of terms in the phonological component is a suboperation of Chain Reduction, which operates with nontrivial chains, not with mere sequences of nondistinct terms. Thus, after (48) is shipped to the phonological component, each instance of which report will be treated as a trivial chain and cannot be deleted by Chain Reduction.

As discussed in section IV.2.1, if either instance of the wh-phrase in (48) were deleted, the resulting structure could be linearized and eventually converge; however, deletion for purposes of linearization does not involve global computations taking into consideration the whole structure, but is restricted to nontrivial chains (see section III.6.2.2). Since neither wh-phrase in (48) can be deleted in the phonological component, the two nondistinct wh-phrases induce violations of the irreflexivity and asymmetry conditions on linear order, making it impossible for (48) to be linearized; hence, no PF object can be formed and the derivation is canceled. Therefore, the unacceptability of sentences such as (45a) and (45b) under the relevant derivation provides further evidence that the phonological component distinguishes chains from sequences of nondistinct terms (see section III.5).
While the S-Structure condition on parasitic gaps must be stipulated in the analyses of parasitic gaps in terms of resumptive pronouns or traces of null operators, it is derived in the system developed here from the linearization computations in the phonological component. To the extent that the analysis of parasitic gaps in terms of sideward movement and linearization of chains derives S-Structure effects without postulating a non-interface level, it gains solid conceptual support from a Minimalist perspective. It should be noted again that no new condition or operation was added to the theory. The same reasoning applied to linearization of chains in successive cyclic movement (see section III.6.2.5) accounts for the licensing of parasitic gaps.

IV.2.3. Anti-c-command Requirement

The licensing of parasitic gaps is also dependent on the structural position of the "real" gap (see Taraldsen (1981) and Engdahl (1983)). As shown in (49), a parasitic gap cannot be c-commanded by the licensing gap. (50) further shows that the unacceptability of (49) in English is not due to some Case incompatibility between the two gaps.\(^{15}\)

\(^{15}\) Recall that in section II.14.4, I argued that English nominal elements are underspecified in terms of the specific kind of case-feature they bear. In the following representations, I will therefore use the notation CASE rather than ACC. or NOM. For English NPs. I will return to this issue in section IV.4, where I contrast English with languages where Case is not underspecified.
(49)  
   a.  *Who called you before you met?  
   b.  *[ who, ti called you [ before you met PGi ] ]  

(50)  
   a.  Which papers did John say were unavailable before reading?  
   b.  [ [ which papers ], did John say ti were unavailable [ before reading PGi ] ]  

In this section, I show that the anti-c-command restriction on the distribution of parasitic gaps follows from the interaction between the conditions governing the application of Form Chain, repeated below in (51), and the linearization of structures containing nondistinct terms discussed in sections III.6.2.5 and IV.1.
(51) *Form Chain:*

The syntactic objects $\alpha$ and $\beta$ can form the nontrivial chain $\text{CH} = (\alpha, \beta)$ only if:

(i) $\alpha$ c-commands $\beta$;

(ii) $\alpha$ is nondistinct from $\beta$;

(iii) (at least) one given feature $F$ of $\alpha$ enters into a checking relation with a sublabel of $K$, where $K$ is the head (of the projection) with which $\alpha$ merges, and the corresponding feature $F$ of $\beta$ could enter into a checking relation with a sublabel of $K$; and

(iv) there is no syntactic object $\gamma$ such that $\gamma$ has a feature $F'$ which is of the same type as the feature $F$ of $\alpha$, and $\gamma$ is closer to $\alpha$ than $\beta$ is.

Let us start by examining the derivation of (49a) under the assumption that the temporal PP in (49a) is adjoined to the VP shell headed by the main verb (see below for other possibilities). If so, at a given derivational stage, the structure in (52) has been assembled.
In a convergent continuation of (52), the matrix light verb must assign its θ-role (see section II.16). If there is no element in the numeration which could merge with the matrix vP in (52) and allow the matrix verb to comply with the θ-Criterion, the computational system may copy the wh-phrase inside the adjunct PP and merge it with the matrix vP, as shown in (53).

Further computations yield the structure in (54) below, where another copy of who is made in order to check the strong feature of T, and the null
interrogative complementizer Q unselectively binds the wh-phrase in subject position (see Chomsky (1995:291) and section II.14.3.3).\textsuperscript{16} When the copy of who in the Spec of TP enters into a checking relation with T, their Case-features are rendered invisible at the interface, as represented by the subscript in (54).\textsuperscript{17}

\begin{itemize}
\item\textsuperscript{16} I will take the absence of T-to-C movement in questions involving a subject wh-phrase as evidence that the wh-phrase does not move to the Spec of CP and is unselectively bound by the interrogative complementizer. See Chomsky (1995:291), section II.14.3.3, and chapter II:fn. 78 for discussion.
\item\textsuperscript{17} In the continuation of (54), no further checking relation involving the categorial feature of who is necessary. Thus, assuming that the wh-feature of who is a type of D-feature (see Chomsky (1995:289)), economy considerations minimizing the number of applications of FF-Elimination require that the ([+interpretable]) wh-feature of the copy in the Spec of TP in (54) be checked for PF purposes in virtue of having entered into a checking relation with the strong feature of T (see sections III.6.2.5 and IV.1). For convenience, only Case-features are represented in (54).
\end{itemize}
It could be the case that the structure in (54) yields an unacceptable result because PP adjoins to a complement which is presumably a thematic position (see Chomsky (1986b, 1994, 1995:chap. 4)). Even if we put this possibility aside, a derivation involving a structure such as (54) cannot yield a licit object, because it cannot be linearized, as shown below.

Let us examine whether the copies of who in (54) can form chains. Since the copy in the Spec of vP enters into a thematic relation with the light verb, it cannot participate in a checking relation (see section II.7.5). Thus, although it can form a chain with the highest copy, Last Resort does not allow it to form a chain with the lowest copy (see (51iii)). Now consider the highest and the lowest copies. Intervening between these two copies are the Spec of the matrix vP, the Spec of the TP in the adjunct clause, and the Spec of vP inside the adjunct clause; all of these positions are occupied by elements that have D- and Case-features, i.e. the features which participate in checking relations in the Spec of the matrix TP (see fn. 17). Hence, the Minimal Link Condition (see (51iv)) prevents the lowest copy of who from forming a chain with the highest one. Therefore, only the two upper copies of who in (54) can form a chain. If this chain is formed before Spell-Out, it will be reduced in the phonological component as in (55), in accordance with economy considerations minimizing applications of FF-Elimination (see sections III.6.2.3 and IV.1).
(55) \[
[\text{CP Q [TP who}^{\text{CASE}} \text{ [VP [VP called you ] [PP before you met ]}]
\text{[who}^{\text{CASE}}
\text{]]]]]
\]

The two nondistinct copies of who in (55) induce violations of the irreflexivity and asymmetry conditions on linear order, preventing the structure from being linearized and canceling the derivation. The unacceptability of sentence in (49a), repeated below in (56), then follows from an unmotivated application of deletion to a trivial chain (see sections IV.2.1 and IV.2.2).

(56) *Who called you before you met?

The derivation is also canceled if the PP of (26) is adjoined to the VP shell headed by the light verb, as illustrated in (57), or to the matrix TP, as illustrated in (58).
(57) CP
Q TP
  who\textsuperscript{i-CASE} T'
    T vP
      vP
        who\textsuperscript{i-CASE} v' before you met who\textsuperscript{i-CASE}
          v VP
            called you

(58) CP
Q TP
  TP
    T' before you met who\textsuperscript{i-CASE}
      T vP
        who\textsuperscript{i-CASE} v' called you
          v VP
            called you
As in (55), the copy inside the adjunct in (57) cannot form a chain with the copy in the Spec of TP because of the Minimal Link Condition, or with the copy in the Spec of vP because of Last Resort.\(^{18}\) As for (58), the copy in the adjunct position cannot form a chain with either of the other copies because the c-command condition on Form Chain is not met (see (51i)). Therefore, after the structures in (57) and (58) are shipped to the phonological component, the copy of who inside the adjunct clause and one of the other copies which has survived Chain Reduction induce violations of the irreflexivity and asymmetry conditions on linear order, canceling the derivation. Again, the unacceptability of (56) is due to an unmotivated application of deletion to a trivial chain (see sections IV.2.1 and IV.2.2).

Thus, regardless of the position of the adjunct in constructions such as (56), there arises a situation where two different chains involving nondistinct copies of who do not have a link in common. The links of these chains which escape Chain Reduction then induce violations of the

\(^{18}\) In instances such as (i) below, where PP is adjoined to vP and the "parasitic gap" copy is in the Spec of the embedded TP, the Minimal Link Condition does not prevent the "parasitic gap" copy from forming a chain with the copy in the Spec of the matrix TP. However, since both copies in the Spec of TP have their Case-features checked, Last Resort prevents them from forming a chain, with eventually cancels the derivation because the structure cannot be linearized. Putting aside a possible ECP effect, the unacceptability of (ia) is due to the fact that Chain Reduction has deleted the two links of the wh-chain inside the adjunct clause, when a single application of deletion would suffice to allow this chain to comply with the irreflexivity condition on linear order (see sections III.6.2.2 and IV.2.1).

(i) a. *Who called before met you?
   b. \[CP Q [TP who^{i-CASE} [vP [who^{i-CASE} called ] [PP before [TP who^{i-CASE} [vP who^{i-CASE} met you ] ] ] ]]]]
irreflexivity and asymmetry conditions on linear order, canceling the
derivation because no PF object is formed. By contrast, the two possible
nontrivial chains involving copies of *which paper* in (50b), repeated below
in (59), have a link in common, which eventually allows (59) to be
linearized in accordance with the LCA, as shown below.

\[
\text{(59)} \quad \left[ \text{CP [ which papers ]}^i \text{ did+Q [TP John [VP say [ which papers ]}^i \text{ were [ which papers ]}^i \text{ unavailable ] [PP without reading [ which papers ]}^i \text{ ] ] ] ] ]}
\]

Suppose, for instance, that the computational system starts the
derivation of (59) by assembling the matrix *VP*, as shown in (60).

---

19 I will not attempt to choose between these three possibilities of adjunction (see Thompson (1994, 1995) for discussion concerning the effects of the height of temporal adjuncts for Binding Theory and antecedent contained deletion). For concreteness, I will assume from now on that adjuncts such as the PP in (56) are adjoined to the VP shell headed by the light verb.

Notice also that I am assuming for expository purposes that the VP of (57), for instance, precedes its adjunct. According to Kayne’s (1994) theory, however, the order VP-PP means that either the whole VP or each of its constituents asymmetrically c-commands PP. Suppose, for instance, that instead of an adjunction structure for the temporal PP of (57), PP is the complement of a head H and VP is the specifier of H, as represented in (i) below. If so, the copy of *who* inside PP cannot form a chain with the copy inside VP because of the c-command condition on Form Chain (see (51i)), or with the copy in the Spec of TP (see the discussion of (57) and (58) and fn. 18) because of the consequences of a structure such as (i) for reconstruction effects and linearization at LF.

\[
\text{(i) [CP Q [TP who^i T [VP who^i called you ] [H H [PP before you met who^i ] ] ] ] ]}
\]

20 I will disregard the possibility of there being an extra copy of *which papers* in the Spec of CP in (60), which is orthogonal to the present discussion (see chapter III:fn. 7).
In (60), the phrase *which papers* is copied and merges with $T'$, becoming the specifier of $T$. This checking configuration then allows the strong D-feature of $T$, as well as the Case-features of both $T$ and the wh-phrase in its Spec to be checked. In principle, this checking configuration could also allow the wh-phrase to have its categorial feature checked for PF purposes (see section III.6.2.4). However, if we follow Chomsky (1995:289) in taking a wh-feature to be a type of D-feature, convergence considerations prevent the copy in the Spec of TP from having its wh-feature checked; if this feature were checked, Last Resort (see (51iii)) would prevent the copy of *which papers* from forming a chain with the copy which later on merges
with the matrix $C'$ to check the strong wh-feature of $Q$.

Lack of chain formation between the nondistinct copies in the Spec of the embedded TP and in Spec of the matrix CP in (59) would then make it impossible for (59) to be linearized in the phonological component, canceling the derivation. Thus, in a convergent derivation, the copy in Spec of TP in (59) checks its Case-, but not its wh-feature, as represented in (60).

After forming the matrix $vP$ in (60), the computational system then starts building the PP by selecting the verb *reading* from the numeration. In order for the derivation to converge, another copy of *which papers* should be made and merged with *reading*. If the Copy operation targets the wh-phrase in the Spec of TP in (60), the derivation will not converge for the following reason: since the wh-phrase in the Spec of CP has already checked its Case-feature, it will not be able to check the Case-feature of the verb in the adjunct clause. Thus, the source of the wh-phrase which merges with *reading* is the copy inside AP, which has no features checked, as represented in (61).

\[
\text{(61) } \quad [\text{PP without reading [ which papers ]}-\text{WH-CASE}] \]

---

$^{21}$ Recall that under the reformulation proposed in section III.6.2.5, Last Resort not only checks whether the upper link of a chain enters into a checking relation with a sublabel of its host, but also whether the lower link could actually enter into the same checking relation as the upper link (see (51iii)). Thus, regardless of which wh-phrase is the source for the instance of *which papers* in the Spec of CP in (59), Last Resort will prevent the instance in the Spec of CP and the one in the Spec of TP from forming a chain, if the latter has its wh-feature checked.
The PP in (61) is then adjoined to the matrix vP, and the computational system proceeds to assemble the matrix CP, which is represented in (62).

Since the complementizer Q in (62) has a strong wh-feature, another copy of which papers is created and merges with CP, allowing Q to have its
strong feature checked. Since no further checking operations require overt copying of *which papers*, the copy in Spec of Q will then be able to check its wh-feature for PF purposes (see section III.6.2.5). The question then is which instance of *which papers* is the source of the copy in the Spec of Q. Notice that if Copy targets either the copy within AP or the copy within PP, FF-Elimination will have to delete the unchecked Case-feature of the newly created copy in the phonological component; if Copy targets the wh-phrase in Spec of TP, on the other hand, FF-Elimination will not be required to delete the Case-feature of the resulting copy, because its source has already checked its Case-feature. Economy considerations thus require that Copy target the wh-phrase in the Spec of TP. To put it generally, given a derivation with the nondistinct terms $\alpha$ and $\beta$ as potential targets for the Copy operation, Copy will target the term with more features checked, up to convergence.\(^{22}\) Thus, the representation of (59) displaying the relevant checked features is as in (63).

\(^{22}\) This is not testable in (59), because the phrase *which papers* can appear in both nominative and accusative positions. Potential test cases involve the wh-phrases *who* and *whom*, which show morphological case distinction, as shown in (i) below. However, the stylistic difference between these wh-phrases appears to obscure judgments. To the extent that one can abstract away from this extra factor, the contrast in (i) may be taken to show that, as in (63), the wh-phrase in the matrix Spec of CP has its Case-feature checked; hence the marginality of the default accusative form in (ib) (see section II.14.4). For a discussion of the licensing of parasitic gaps in languages with rich case morphology, see section IV.4.

(i)  a. Who did you say was nice without talking to?
    b. ??Whom did you say was nice without taking to?
The copy in the Spec of TP in (63) forms the chain $\text{CH}_1 = ([\text{which papers }]^l\text{-WH-CASE}, [\text{which papers }]^l\text{-WH-CASE})$ with the copy inside AP, and the chain $\text{CH}_2 = ([\text{which papers }]^l\text{-WH-CASE}, [\text{which papers }]^l\text{-WH-CASE})$ with the copy in the Spec of CP. Since the head of $\text{CH}_1$ is the tail of $\text{CH}_2$, the
linked chain CH₃ = ([ which papers ]⁻WH-CASE, [ which papers ]⁻WH-CASE, [ which papers ]⁻WH-CASE) is then formed (see Chomsky and Lasnik (1993:563) and section III.4.1). In addition, the copy in the Spec of CP forms the chain CH₄ = ([ which papers ]⁻WH-CASE, [ which papers ]⁻WH-CASE) with the copy within the adjunct PP.

Let us assume that all these chains have been formed before Spell- Out. Thus, the optimal reduction of the linked chain CH₃ in the phonological component involves the deletion of its two lower links, as shown in (64) (see section III.6.2.2).

(64) [CP [ which papers ]⁻WH-CASE did [TP John [νP [νP say were unavailable ] [PP without reading [ which papers ]⁻WH-CASE ] ]]]

Applying to CH₄ in (64), Chain Reduction then deletes the lower link of CH₄, yielding the structure in (65). After (65) is linearized and further computations take place in the phonological component, the PF output associated with (66) is obtained.

(65) [CP [ which papers ]⁻WH-CASE did [TP John [νP [νP say were unavailable ] [PP without reading ] ] ]]]
Which papers did John say were unavailable without reading?

To summarize, the analysis of parasitic gap constructions in terms of sideward movement and linearization of chains that I am proposing here accounts for the fact that a parasitic gap cannot be c-commanded by the "real" gap in the following way. Last Resort (see (51iii)) or the Minimal Link Condition (see (51iv)) may prevent a "parasitic gap" copy from forming a chain with a nondistinct copy which c-commands it; if that happens, at least two nondistinct terms will survive Chain Reduction in the phonological component and will induce violations of the irreflexivity and asymmetry conditions on linear order, canceling the derivation because no PF object can be formed. Again, the analysis of parasitic gaps in terms of sideward movement and linearization of chains derives the anti-c-command requirement on parasitic gaps without introducing any new principle or operation. The same factors which interact to determine when a "trace" can be deleted are employed to account for why a "parasitic gap" copy cannot be c-commanded by a "real gap" copy.

The competing alternatives, on the other hand, require extra assumptions to account for the anti-c-command requirement. Arguing for a null operator approach to parasitic gap constructions, Chomsky (1986a:65), for instance, suggests that the null operator must be 0-subjacent to the licensing gap; that is, there can be no barriers intervening between the
licensing gap and the null operator that binds the parasitic gap. In turn, in a
null resumptive pronoun approach such as Cinque's (1990), the
resumptive pronoun in the parasitic gap position must be A'-bound at S-
Structure in order to be licensed; thus, an intervening c-commanding gap in
an A-position presumably prevents the null resumptive pronoun from being
identified.

I will not discuss the adequacy of these proposals in detail. For the
current purposes, suffice it to note that they either require a considerable
enrichment of the theoretical apparatus and/or crucially rely on licensing
conditions at S-Structure, an option which is not available within the
Minimalist Program.

IV.2.3. Island Effects

The main motivation for the analyses of parasitic gaps in terms of null
operator movement is that parasitic gap constructions exhibit island effects
(see Kayne (1983, 1984), and Chomsky (1986a), among others). The pattern of
acceptability in the parasitic gap constructions in (67), for example, to a
certain extent replicates the pattern found in the corresponding instances of

\footnote{Chomsky (1986b:63) also raises the possibility that after the "licensing chain"
and the "parasitic chain" are connected by a process of chain composition, a subchain
of the composed chain will violate the Chain Condition (see section III.4.1.1), if the
parasitic gap is c-commanded by the real gap. On the undesirability of the Chain
Condition in the Minimalist framework, see discussion in section III.4.1.1.}
wh-movement given in (68) (adapted from Chomsky (1986a:55-56)):\(^{24}\)

\[
(67) \quad [ [ \text{which man} ] \text{, did John interview } t_i \text{ before} \\
\quad \text{a. } [ \text{expecting you to tell you to give the job to } PG_i ] ] \\
\quad \text{b. } [ \text{announcing the plan to speak to } PG_i ] ] \\
\quad \text{c. } [ \text{asking his boss whether they should hire } PG_i ] ] \\
\quad \text{d. } [ \text{asking which job to give to } PG_i ] ] \\
\quad \text{e. } [ \text{expecting you to leave without meeting } PG_i ] ]
\]

\[
(68) \quad [ [ \text{which man} ] \text{, did John} \\
\quad \text{a. } [ \text{expect us to tell you to give the job to } t_i ] ] \\
\quad \text{b. } [ \text{announce the plan to speak to } t_i ] ] \\
\quad \text{c. } [ \text{ask his boss whether they should hire } t_i ] ] \\
\quad \text{d. } [ \text{ask which job to give to } t_i ] ] \\
\quad \text{e. } [ \text{expect you to leave without meeting } t_i ] ]
\]

The similarities between (67) and (68) can be accounted for if a null operator moves from the position of the parasitic gap in (67) to an A'-position where it can be licensed by the other wh-chain, as represented in (69). Recall that in Chomsky's (1986a) system, for instance, the null operator

\(^{24}\text{In general, islands containing "real gaps", "parasitic gaps" or wh-phrases in situ yield marginal results, being rather sensitive to lexical choices, among other factors. The judgements to be discussed regarding island effects should thus be taken as relative to one another, as usual.}\)
must move to a position where it can be 0-subjacent to the licensing gap (see section IV.2.2).\textsuperscript{25,26}

\textsuperscript{25} Chomsky (1986a), citing Kearney (1983), takes the absence of reconstruction in the parasitic gap position in constructions such as (i) as evidence for the "parasitic" nature of the empty category in the object of \textit{read}. The reverse pattern found in parasitic gaps inside relative clauses adjoined to subjects, as illustrated in (ii) (from Munn (1994)), shows that reconstruction in either gap is in principle possible, regardless of its "parasitic" nature. See Nunes (1995a), where an analysis of (i) and (ii) in terms of linearization at LF is offered.

(i) a. Which books about himself did John file before Mary read?  
b. *Which books about herself did John file before Mary read?

(ii) a. *Which picture of herself did every boy who saw say Mary liked?  
b. Which picture of himself did every boy who saw say Mary liked?

\textsuperscript{26} As mentioned in the appendix of chapter III, Munn (1994) proposes that null operators are copies of the elements that license them. Under this view, a sentence such as (i) below, for instance, has the structure in (ii) before Spell-Out.

(i) Which paper did you file without reading?

(ii) \([CP [which paper]]^\text{did+Q} [TP you [\text{file} [which paper]]^\text{Q} [PP \text{without} [CP [which paper]]^\text{C} [TP \text{reading} [which paper]]^\text{Q}]]]])]]

As discussed in the appendix, in order for this proposal to be compatible with the analysis of trace deletion developed in section III.6, it must be the case that the "null operator" copy forms a chain with the element that licenses it; otherwise, the nondistinct copies that do not form a chain induce violations of the irreflexivity and asymmetry conditions on linear order, making it possible for a PF object to be formed and canceling the derivation. If a linked chain between the copy in the Spec f the matrix CP and the two copies inside the adjunct clause in (ii) is formed, however, there would be no explanation for why the sentence in (iii), represented in (iv) in order for a linear order to obtain would be the structurally analogous to the one required to be formed in (ii). In addition, the analysis of null operators as copies also fails to account for the acceptability of parasitic gaps inside relative clauses in subject position (see the contrast between (70) and (71), for example). On the reconstruction effects which motived Munn's analysis of null operators in parasitic gap constructions as copies of the element that licenses them, see Nunes (1995a).

(iii) ??Which book did you file this paper after reading?

(iv) \([CP [which book]^\text{did+Q} [TP you [\text{file this paper}][PP \text{without} [CP [which book]C [TP \text{reading} [which book]]]]]]])
(69)  [ [ which man ] , did John interview t, before
  a. [ OP, expecting us to tell you to give the job to t, ] ]
  b. ?[ OP, announcing the plan to speak to t, ] ]
  c. ??[ OP, asking his boss whether they should hire t, ] ]
  d. *[ OP, asking which job to give to t, ] ]
  e. *[ OP, expecting you to leave [ without meeting t, ] ] ]

The similarities between overt wh-movement and movement of a
null operator (whether or not it is taken to be a copy of the element that
licenses it (see fn. 26)) break down in constructions involving relative
clauses in subject position. Whereas a parasitic gap inside a relative clause
adjoined to a subject is acceptable, overt wh-movement from the same
position does not yield an acceptable result, as respectively shown in (70) and
(71).

(70)  a. Which man does everyone who meets end up liking?
       b. [ [ which man ] , does [ [ everyone ] [ who meets PG, ] ] end up
         liking t, ] ]

(71)  a. *Which man does everyone who meets feel happy?
       b. *[ [ which man ] , does [ [ everyone ] [ who meets t, ] ] feel
          happy ] ]
Chomsky (1986a) accounts for the acceptability of (70) by relying on the Vacuous Movement Hypothesis, according to which subject wh-phrases do not move to Spec of CP (see fn. 16). If this is the case in constructions such as (70b), Chomsky argues, the null operator moves to the Spec of CP at S-Structure, as represented in (72), and then is deleted at LF, vacating this position before movement of who. As evidence for this analysis, Chomsky presents constructions such as (73) below, which involve overt movement to the Spec of the CP of the relative clause. The unacceptability of (73a) is taken to show that the null operator in (73b) cannot move to the Spec of the embedded CP, which is filled by to whom, and thus fails to be licensed at S-Structure.

(72) \[ [ [ which man ]_i \text{ does [ [ everyone ] } [OP_i [ \text{ who meets } t_i ] ] \text{ end up liking } t_i ] \]

(73) a. *Which book did everyone to whom we gave like?

b. *[ [ which book ]_i \text{ did [ [ everyone ] [ [ to whom ]_j we gave } PG_j [ t_j ] ] like } t_i ]

However, as pointed out by Frampton (1990:fn. 21) citing Richard Kayne (see also Williams (1990)), the unacceptability of constructions such as (73a) may be independently due to their crossing dependencies (see Pesetsky
When crossing is avoided and vacuous movement is irrelevant, as in (74) and (75), a well-formed result obtains. Thus, the contrast between (70) and (71) still remains to be explained under the null operator analysis of parasitic gaps, regardless of whether or not the null operator is treated as a copy of the element that licenses it.

(74) a. Which paper did everyone we asked about praise?
   b. \[ [ \text{which paper} ]_i \text{ did } [ [ \text{everyone} ] [ \text{OP}_j \text{ we asked } t_j \text{ about } PG_i ] ] \text{ praise } t_i ]

(75) a. Which guy did every joke we told to delight?
   b. \[ [ \text{which guy} ]_i \text{ did } [ [ \text{every joke} ] [ \text{OP}_j \text{ we told } t_j \text{ to } PG_i ] ] \text{ delight } t_i ]

Recall that constructions such as (70) also pose problems for the analysis of parasitic gaps in terms of null resumptive pronouns (see section IV.2.1), because an overt resumptive pronoun in the position of the parasitic gap in (70) leads to unacceptability, as shown in (76).

(76) a. *Which man does everyone who meets him end up liking?
   b. *[ [ \text{which man} ]_i \text{ does } [ [ \text{everyone} ] [ \text{who meets him}_i ] ] \text{ end up liking } t_i ]
In order to account for island effects within his analysis of parasitic gaps in terms of resumptive pronouns, Cinque (1990:sec. 3.3.4) proposes that (i) the null resumptive pronoun moves at LF either by itself (see also Haïk (1985)) or within a larger phrase under pied-piping; and (ii) that this movement is subject to (a version of) the Connectedness Condition (see Kayne (1984)). For reasons internal to his system, however, movement of the null resumptive pronoun only, or movement of NP* by pied-piping in constructions such as (77) below, should produce an ill-formed result, contrary to the case. Cinque (1990:149) then proposes that neither movement takes place in this kind of parasitic gap; rather the resumptive pronoun is licensed through connectedness by the real gap on its right (but see fn. 14). Again, parasitic gaps within relative clauses adjoined to subjects receive a different treatment.

(77)  \[ \text{[ which man ]}_{i} \text{ does } \text{[NP* ]}_{i} \text{[ everyone ]}_{i} \text{[ who meets pro. ]}_{i} \text{ end up liking } t_{i} \] 

The brief discussion above shows that neither the null operator nor the null resumptive pronoun approaches to parasitic gap constructions have a uniform treatment of island effects found in parasitic gap constructions, which is also able to account for parasitic gaps inside relative clauses in subject position.
A full discussion of how island effects other than the ones induced by the Minimal Link Condition are to be captured within Minimalism goes beyond the scope of this dissertation. Even in the absence of a complete reanalysis of Subjacency and CED/ECP effects in the Minimalist framework, the core of my proposal can nonetheless be evaluated. I will continue to explore the idea that parasitic gaps form chains without the mediation of a "null operator" copy. I will furthermore suggest that the formation of a chain involving a parasitic gap is comparable to unselective binding of wh-phrases in situ by an interrogative complementizer, as far as economy considerations are concerned. Whatever turns out to be responsible within the Minimalist framework for the island effects found in constructions involving wh-phrases in situ should also extend to island effects in parasitic gap constructions.

In sections IV.2.4.1-IV.2.4.3 below, I examine the similarities between parasitic gaps and wh-phrases in situ with respect to island effects. The discussion is then summarized in section IV.2.4.4.

**IV.2.4.1. Parasitic Gaps, Wh-phrases in Situ, and Island Effects**

Before we consider the formation of chains in parasitic gap constructions, let us examine structures with similar structural properties. Suppose that the computational system has assembled the structure in (78)
and that the complementizer Q has a strong wh-feature. If there is no other lexical item in the numeration which can check the strong feature of Q, the computational system then must make a copy of one of the wh-phrases in (78) and place it in the checking domain of Q. If the wh-phrase in the matrix clause is copied, (79a) is derived; if the wh-phrase within the adjunct clause is copied, (79b) is the result.

(78) [ Q+did you [ [ go where ] [ without notifying who(m) ] ] ]

(79) a. Where did you go without notifying who(m)?
   b. *Who(m) did you go where without notifying?

The contrast in acceptability between (79a) and (79b) cannot be ascribed to the Minimal Link Condition. Since neither wh-phrase c-commands the other, neither is closer than the other to the matrix Spec of CP. The unacceptability of (79b) cannot be ascribed to the impossibility of extracting out of adjunct islands either. Although marginal (see fn. 24), the sentence in (80) below, with extraction from the adjunct island, is on a par with (79a), rather than with (79b).

(80) ?Who(m) did you go to Brazil without notifying?
This set of data can be accounted for, if island violations other than violation of the Minimal Link Condition somehow entail derivational cost. That is, given a choice, the computational system will prefer options which do not involve island violations. If so, the derivation of (79b), although convergent, is ruled out by the more economical derivation in (79a), which involves no island violations. The derivation of (80), on the other hand, has no competing alternative which converges; thus, it is shipped to the interface systems and its marginality should presumably be attributed to the interpretation that the C-I interface assigns to structures involving island violations.

Consider now the structure in (81), in which the wh-phrase which paper was copied from the object position of one of the verbs and merged with the other. Assuming that the numeration has been exhausted and that Q has a strong feature, the computational system makes an additional copy of which paper and merges it with the structure in (81), yielding (82).

(81) \[ Q+\text{did} \ [ \text{you} \ [ \ [ \text{file} \ [ \text{which paper} ] ] \ ] \ ] \ [ \text{without reading} \ [ \text{which paper} ] ] ] ]

(82) \[ \ [ \text{which paper} ] ] Q+\text{did} \ [ \text{you} \ [ \ [ \text{file} \ [ \text{which paper} ] ] ] \ ] \ [ \text{without reading} \ [ \text{which paper} ] ] ] ]
In principle, the instance of *which paper* in the Spec of CP can form a chain with either wh-phrase it c-commands. According to the discussion about the contrast between (79a) and (79b), the computational system chooses to form a chain with the object of *file*, because it is more economical given that it involves no island violations. However, if the computational system only forms a chain between the copy in Spec of the matrix CP and the copy in the matrix object position, no derivation will be generated: after (82) is shipped to the phonological component, the copy that escapes Chain Reduction and the copy inside the adjunct clause will induce violations of the irreflexivity and asymmetry conditions on linear order, canceling the derivation because no linear order can be established and no PF object is formed. The computational system then forms an additional chain between the copy in the matrix Spec of CP and the copy inside the adjunct clause, enabling the structure in (82) to be realized as (83) after both chains are reduced and further computations in the phonological component apply (see section IV.2.1).

(83)  Which paper did you file without reading?

What is relevant for our purposes is that at the derivational step after (82) is assembled, it is more economical for the computational system to form a chain with the "real" gap. Thus, the parasitic gap of (82) is parallel to
the wh-phrase in situ in (79a), in the sense that a chain involving either of them entails derivational cost and is not a preferred option. In the case of parasitic gap constructions, the disfavored option must also be exercised, though; if the "parasitic gap" copy does not form a chain with a c-commanding copy, no linear order can be established and no PF object is formed. In the case of wh-phrases in situ within islands, on the other hand, they are interpreted as a variable by being unselectively bound by an interrogative complementizer (see Baker (1970), Chomsky (1995:291), section II.14.3.3, and Nunes (1995a)).

It is thus plausible to conjecture that islands interfere in the acceptability of this process of unselective binding of a wh-phrase in situ in a similar way that they interfere with the acceptability of a parasitic chain. To a certain degree of approximation, this prediction holds true. Although island effects appear to be weaker in parasitic gap constructions than in constructions with wh-phrases in situ, the two constructions exhibit the same scale of acceptability with respect to type and number of islands. Thus, along with the paradigm in (67), repeated below in (84), we have the paradigm in (85).

---

27 Parasitic gaps and wh-phrases in situ also pattern alike with respect to connectedness effects (see Kayne (1984)).

28 The fact that island effects are stronger in constructions involving unselective binding of a wh-phrase in situ than between the links of a "parasitic chain" may be due to computational complexity. Whereas the "licensing chain" and the "parasitic chain" are associated with a single semantic variable, in multiple interrogatives the moved wh-phrase and the wh-phrase in situ are each associated with a different semantic variable.
(84)  [ [ which man ]_i did John interview t_i
   a.  [ expecting us to tell you to give the job to PG_i ] ]
   b.  ?[ announcing the plan to speak to PG_i ] ]
   c.  ??[ asking his boss whether they should hire PG_i ] ]
   d.  *[ asking which job to give to PG_i ] ]
   e.  *[ expecting you to leave without meeting PG_i ] ]

(85)  Who interviewed John
   a.  after telling us to give the job to which person?
   b.  ?after announcing the plan to speak to which person?
   c.  ??after asking his boss whether they should hire which person?
   d.  *after asking which job to give to which person?
      (wide scope reading)
   e.  *after allowing you to leave without meeting which person?

As noted above, I will not attempt to develop an analysis of Subjacency or CED/ECP effects within the Minimalist framework in this dissertation. My aim here is simply to show that once these effects also arise in unselective binding of wh-phrases in situ by an interrogative complementizer; it is not necessary that a null operator must be postulated in order for island effects to be accounted for, regardless of whether or not
the null operator is a copy of the element that licenses it (see fn. 26). In other words, whatever accounts for island effects in constructions with wh-phrases in situ should in principle apply to parasitic gap constructions as well. In the next sections, we find additional evidence which lends support to this approach.

IV.2.4.2. Categorial Restrictions

Cinque (1990) notes that Adriana Belletti’s observation that only NPs can be extracted out of islands, as illustrated in (86), extends to parasitic gap constructions, as well. Thus, adjectival, prepositional, and adverbial phrases, for instance, cannot license parasitic gaps, as shown in (87) (see Postal (1994) for relevant discussion).

(86)

a. [ [ which man], did you go home [ after meeting t₁ ] ]

b. *[ [ to which man ], did you travel [ after writing t₁ ] ]

c. *[ [ how sick ], did he go to the hospital [ after feeling t₁ ] ]

d. *[ [ how kind ], did he go to sleep [ after behaving t₁ with your friends ] ]
(87) a. [ [ which paper ] : did you file t : [ without reading PG ] ]

b. *[ [ to whom ] : did John give directions t : [ after talking PG ] ]

c. *[ [ how important ] : can one become t : [ without feeling PG ] ]

d. *[ [ how kind ] : did he behave t : with you [ without behaving PG with your friends ] ]

In order to account for the paradigm in (87), the null operator approach to parasitic gaps has to assume that a null operator can only be an NP (see Aoun and Clark (1985), for instance) or that a non-nominal null operator cannot move to the same position that a nominal null operator can move to, thus failing to be licensed by chain composition (see Chomsky (1986a)). These approaches, however, have nothing to say regarding the sentences in (86), where presumably there is no null operator. Cinque (1990), on the other hand, accounts for (86) and (87) by taking these constructions to both involve null resumptive pronouns, and proposing that only NPs have a null resumptive pronoun at their disposal.

It is interesting to note that the categorial restriction observed in extraction out of islands and in parasitic gap constructions partially extends to wh-phrases in situ. As widely discussed in the ECP literature, multiple interrogative constructions give rise to unacceptable results if a wh-phrase in situ inside an island is an adjectival or an adverbial phrase, as illustrated
in (88c) and (88d). On the other hand, no contrast appears to arise between nominal and prepositional wh-phrases in situ within islands, as shown in (88a) and (88b).

(88)  
   a. Who went home after meeting whom?  
   b. Who traveled to Brazil after writing to whom?  
   c. *Who went to the hospital after feeling how sick?  
   d. *Which kid went to sleep after behaving how kind with your friends?

The acceptability of prepositional wh-phrases in situ within islands may be due to independent factors. If wh-phrases in situ are unselectively bound by an interrogative complementizer in order for an operator-variable relation to be established, nothing appears to force the interrogative complementizer to bind a PP when it can bind the nominal wh-phrase which is the complement of the preposition instead. If so, the interrogative complementizer in (88b) unselectively binds whom, rather than to whom; hence the lack of contrast between (88a) and (88b). This possibility is not available in instances of extraction out of an island or parasitic gap constructions because the PP in Spec of CP must form a chain with its copy inside the adjunct island; hence, the unacceptability of (86b) and (87b).

Ideally, the categorial restrictions on extraction out of islands, parasitic
gaps, and wh-phrases in situ inside islands should receive a unique explanation. I will not attempt to pursue such an explanation here, since it depends on how Subjacency and CED/ECP effects are to be captured within Minimalism. For the purposes of the current discussion, it suffices to note that if wh-phrases in-situ are unselectively bound by an interrogative complementizer (see Baker (1970), Chomsky (1995:sec. 4.5), section II.14.3.3, and Nunes (1995a)), the correct description of the set of data in (86), (87), and (88) seems to be that no operator-variable relation can be established if the wh-phrase within an island is a non-nominal phrase (see Szabolcsi and Zwarts (1993), Reinhart (1994), and Hornstein (1995:chap. 6-7), for relevant discussion).

IV.2.4.3. "Nonreferential" NPs

Another set of data that Cinque (1990) presents in favor of a null resumptive pronoun analysis of parasitic gaps involves the distribution of "nonreferential" NPs, such as time and measure phrases. Although nonreferential NPs can undergo long distance wh-movement, as illustrated in (89), they cannot be extracted out of islands, as shown in (90), or license parasitic gaps, as shown in (91).
(89)  a.  [ [ how many kilos ], did Mary think John said he weighed t_i ]
    b.  [ [ how many weeks ], did Mary think John said he spent t_i in Brazil ]

(90)  a.  *[ [ how many kilos ], did John go off his diet [ before weighing t_i ] ]
    b.  *[ [ how many weeks ], did John change his life [ after spending t_i in Brazil ] ]

(91)  a.  *[ [ how many kilos ], does he weigh t_i [ without believing that he weighs PG_i ] ]
    b.  *[ [ how many weeks ], did he spend t_i in Berlin [ without wanting to spend PG_i in London ] ]

Cinque (1990:117) relates the unacceptability of constructions such as (90) and (91), which in his analysis involve a null resumptive pronoun, to the fact that nonreferential NPs cannot be pronominalized, as illustrated in (92). Thus, whatever rules out the overt pronouns in (92) should also rule out the alleged null resumptive pronouns in (90) and (91).  

---

29 As pointed out to me by Ellen Thompson (p.c.), the unacceptability of (92b) may be due to the interpretation restrictions imposed by before, rather than a problem with the pronoun itself. In (i) below, for instance, pronominalization is acceptable.

(i) John spent three weeks in Brasil, while Mary spent them in Italy.
(92) a. *John wants to weigh 100 kilos, because his favorite actor weighs them.
    
b. *John spent three weeks in Brazil, before Mary spent them there.

This restriction on parasitic gaps and extraction out of islands also extends to wh-phrases in situ within islands, as illustrated in (93) below. Again, if the ban on nonreferential NPs is stated in terms of an operator-variable relation across islands, the data in (90), (91), and (93) should fall under the same account.

(93) a. *Who went off his diet before weighing how many kilos?
    
b. *Who changed his life after spending how many weeks in Brazil?

**IV.2.4.4. Summary**

The discussion in sections IV.2.4.1-IV.2.4.3 shows that neither the null operator analysis of parasitic gaps nor the resumptive pronoun analysis can provide a general account of the categorial and referentiality restrictions on parasitic gaps, extraction out of islands, and wh-phrases in situ within islands. If these restrictions on parasitic gaps are to be attributed to some
intrinsic property of null operators, the same restrictions on extraction out of islands and wh-phrases in situ within islands must be accounted for independently. If the restrictions are ascribed to some property of null resumptive pronouns, constructions involving parasitic gaps and extraction out of islands may receive a uniform account, but no explanation is provided for why these restrictions also hold of in-situ wh-phrases within islands.

If the categorial and referentiality restrictions are general conditions on a operator-variable relation, as suggested above, all three constructions fall under the same account. However, it remains to be shown why this restriction should hold (see Szabolcsi and Zwarts (1993), Reinhart (1994), and Hornstein (1995:chap. 6-7), for some suggestions) and how the thorny issue of subjacency and CED/ECP effects is to be recast in Minimalist terms.

IV.3. Across-the-Board Extraction

Let us now consider constructions involving across-the-board (ATB) extraction (see Ross (1967) and Williams (1978), among others) such as the one illustrated in (94).

(94)  [ [ which paper ]i, didj John tk file ti and Mary tj read tij ]
The similarities between parasitic gap and ATB constructions have led researchers to attempt to assimilate one of the two to the other. For example, Ha•k (1985) and Williams (1990) have proposed that parasitic gap constructions are to be treated in terms of ATB extraction. Munn (1993) and Postal (1993) pointed out several problems for this approach. Among these problems are the fact that analyses of parasitic gaps as ATB constructions have to assume that every structure which allows a parasitic gap is optionally coordinative, and the fact that these analyses crucially rely on the ATB operation, which is construction-specific.

The opposite approach is taken by Munn (1993), who proposes that ATB constructions involve movement of a null operator, as in parasitic gap constructions. This approach has the conceptual advantage of eliminating the ATB formalism from the grammar. If the analysis of parasitic gaps developed in section IV.2 is on the right track, however, no null operator needs to be invoked in parasitic gap constructions. If we were to recast Munn's proposal in terms of the analysis proposed in section IV.2 and if ATB constructions should fall under the same analysis as parasitic gap constructions, ATB constructions should involve no null operator either.

Below I show that, like parasitic gap constructions, ATB constructions can be analyzed in terms of sideward movement and linearization of chains. In others words, I will attempt to demonstrate that despite their apparent differences, ATB movement and regular movement receive a uniform
treatment under the Copy + Merge theory of movement and the proposal about linearization of chains developed in chapters II and III.

I will follow Munn (1987) and assume that coordinating conjunctions head a phrase of their own, which Munn dubs Boolean Phrase, as illustrated in (95) below, where and has two specifiers.\(^{30}\)

\[(95) \quad [_{BP} \text{John}_{B'} [_{B'} \text{Mary}_{B'} \text{and} \text{Sue} ] ] \]

Under this hierarchical view of coordination, the structure in (94) is to be represented as in (96), where the Boolean head and takes a TP as its specifier and another TP as its complement.

\[(96) \quad [_{CP} [ \text{which paper}_{1} \text{ did}_{1} [_{BP} [_{TP} \text{John}_{t_{1}} \text{ file}_{t_{1}} ] [_{B'} \text{and} [_{TP} \text{Mary}_{t_{1}} \text{ read}_{t_{1}} ] ] ] ] ] \]

Let us consider the relevant details of the derivation of (96), whose initial numeration is as in (97).

\[(97) \quad \text{N} = \{ \text{which}_{1}, \text{ paper}_{1}, \text{ did}_{1}, \text{ Q}_{1}, \text{ John}_{1}, \text{ v}_{2}, \text{ file}_{1}, \text{ and}_{1}, \text{ Mary}_{1}, \text{ read}_{1} \}\]

\(^{30}\) See Munn (1993) and Kayne (1994), among others, for different proposals for the structure of BP.
At some given derivational step, the syntactic objects K and L in (98) have been formed. Since merger of file with either K or any other syntactic object formed from the lexical items still to be selected from the numeration does not yield a convergent derivation, the computational system makes a copy of the term which paper in K and merges it with file, as shown in (99).

(98)  
\[ K = [TP \text{ did } [\nu P \text{ Mary } \nu [VP \text{ read } [\text{ which paper }]i]]] \]

b. L = file

(99)  
\[ K = [TP \text{ did } [\nu P \text{ Mary } \nu [VP \text{ read } [\text{ which paper }]i]]] \]

b. M = [VP file [ which paper ]i]

Further applications of Select and Merge reduce the numeration to N' in (100) and form the object O in (101).

(100)  
\[ N' = \{\text{which}_0, \text{paper}_0, \text{did}_0, Q_i, \text{John}_0, \nu_0, \text{file}_0, \text{and}_1, \text{Mary}_0, \text{read}_0\} \]

(101)  
\[ K = [TP \text{ did } [\nu P \text{ Mary } \nu [VP \text{ read } [\text{ which paper }]i]]] \]

b. O = [VP John \nu [VP file [ which paper ]i]]
Given that there are no Tense heads available in N’, the computational system makes a copy of did and merges it with O, yielding P in (102).

(102)  
   a.  \[ K = [TP \text{ did} [\text{VP Mary v VP read [ which paper ]}]] \]  
   b.  \[ P = [TP \text{ did} [\text{VP John v VP file [ which paper ]}]] \]

The subjects of both K and P in (102) move to the Spec of TP to check the strong feature of each instance of did, and the Boolean head and merges with one of the resulting structures, forming the (simplified) objects R and S in (103). R and S then merge, yielding U in (104), which in turn merges with the interrogative complementizer Q, forming V and exhausting the numeration.

(103)  
   a.  \[ R = [BP \text{ and } [TP \text{ Mary did} \text{ read [ which paper ]}]] \]  
   b.  \[ S = [TP \text{ John did} \text{ file [ which paper ]}]] \]

(104)  
   a.  \[ U = [BP \text{ [TP John did} \text{ file [ which paper ]}][B' and [TP Mary did} \text{ read [ which paper ]}]]) \]  
   b.  \[ V = [CP Q [BP \text{ [TP John did} \text{ file [ which paper ]}][B and [TP Mary did} \text{ read [ which paper ]}]]) \]
Since the numeration has been exhausted and the complementizer Q in (104b) has strong features to be checked, the computational system makes a copy of \textit{did} and adjoins it to Q, and then makes another copy of \textit{which paper} and merges it with the resulting structure, yielding the object in (105).

\[(105) \quad X = [\text{CP} \left[ \text{which paper} \right] \text{i} \text{did} \text{j} + \text{Q} \left[ \text{BP} \left[ \text{TP} \text{John did} \text{j file} \right. \right. \left. \left[ \text{which paper} \right] \text{i} \right] \left[ \text{B'} \text{and} \left[ \text{TP} \text{Mary did} \text{j read} \left[ \text{which paper} \right] \text{i} \right] \right] \right] \]

Since no further checking relation involving the wh-feature of one of the copies of \textit{which paper} or the V-feature of one of the copies of \textit{did} is necessary, economy considerations require that the highest copies in (105) have their categorial features checked for PF purposes (see sections III.6.2.5 and IV.2.1), as represented in (106).

\[(106) \quad [\text{CP} \left[ \text{which paper} \right] \text{i-\text{WH}} \text{j-\text{V} } + \text{Q} \left[ \text{BP} \left[ \text{TP} \text{John did-V file} \right. \right. \left. \left[ \text{which paper} \right] \text{i-\text{WH}} \right] \left[ \text{B'} \text{and} \left[ \text{TP} \text{Mary did-V read} \left[ \text{which paper} \right] \text{i} \right] \right] \right] \]

If (106) is Spelled-Out before the wh- and T-chains are formed, the nondistinct copies of \textit{which paper} and \textit{did} will induce violations of the irreflexivity and asymmetry conditions on linear order, canceling the derivation. Thus, in order for a derivation to be generated, each copy in (106)
must be part of a chain. The pairs involving the copies inside BP, i.e. ([ which paper ]-WH, [ which paper ]-WH) and (did-V, did-V), cannot form chains because they do not satisfy the c-command condition. However, the highest copy of which paper and the highest copy of did can form a different chain with each of their lower copies (see section IV.2.1). If these four chains are formed before Spell-Out, a derivation can now be generated after Chain Reduction applies to each chain in the phonological component, as shown below.

Consider the chain formed between the wh-phrase in the Spec of CP and the one in the object position of read. If Chain Reduction deletes the upper link, FF-Elimination is required to delete the wh-feature of the lower copy. If Chain Reduction deletes the lower link, on the other hand, no application of FF-Elimination deleting the wh-feature of the upper link is necessary; since the wh-feature of the upper link has been checked, it has been rendered invisible at the PF level and does not induce a violation of Full Interpretation (see section III.6.2.5). Thus, the optimal reduction of this chain involves the deletion of the lower link, yielding the structure in (107a) below.

(107)  

\[
\text{[CP [ which paper ]-WH [ did-\text{V} ]+Q [BP TP John did-\text{V} file } \\
\text{[ which paper ]-WH ] [TP and [TP Mary did-\text{V} read ] ] ] ]}
\]
Application of Chain Reduction to the chain formed between the two remaining instances of *which paper* in (107) results in the deletion of the lower link, in compliance with economy requirements regarding the number of applications of FF-Elimination, as shown in (108).

(108) \[
\begin{align*}
\text{[CP [ which paper ]} & \text{\_WH [ did\_V ] }+Q \text{[BP [ TP John did\_V file ] [B and [TP Mary did\_V read ] ] ] ]}
\end{align*}
\]

Let us now consider the T-chains of (108). Take the chain formed between the highest copy of *did* and the copy in the second conjunct. The optimal reduction of this chain involves the deletion of lower copy, since the upper copy has fewer features to be deleted by FF-Elimination. Applying to this chain, Chain Reduction converts (108) into (109) below. By the same reasoning, the optimal application of Chain Reduction to the chain formed by the highest copy of *did* and the copy inside the first conjunct involves the deletion of the lower copy, as shown in (110).

(109) \[
\begin{align*}
\text{[CP [ which paper ]} & \text{\_WH [ did\_V ] }+Q \text{[BP [ TP John did\_V file [ which paper ]] ] }+Q \text{[BP [ TP Mary read [ which paper ]] ] ] ]}
\end{align*}
\]

(110) \[
\begin{align*}
\text{[CP [ which paper ]} & \text{\_WH [ did\_V ] }+Q \text{[BP [ TP John file [ which paper ]] ] }+Q \text{[BP [ TP Mary read [ which paper ]] ] ] ]}
\end{align*}
\]
Further computations in the phonological component finally yield the PF object associated with the sentence in (111).

(111) Which paper did John file and Mary read?

It should be clear that the analysis of ATB constructions in terms of sideward movement and linearization of chains I am proposing is not meant to account for all the properties of extraction out of conjuncts, such as the unacceptability resulting from extraction out of a single conjunct, for instance (for relevant discussion, see Williams (1978) and Munn (1993), among others). Nor am I claiming that parasitic gap and ATB constructions have the same properties (see Postal (1993) for an inventory of many interesting differences between parasitic gap and ATB constructions). All that I am proposing is that the core property of ATB constructions, namely that a given element appears to be moving from more than one site, is captured very naturally under a framework which takes Move to be the description of a complex interaction among the operations Copy, Merge, Form Chain, and Chain Reduction. To the extent that this analysis succeeds, it allows ATB extraction to be treated as standard overt movement, dispensing with the ATB formalism, which is a welcome result from a conceptual point of view.31

---

31 Apparent counterexamples to the elimination of the ATB formalism involve wanna-contraction constructions such as (i), which was pointed out to me by David
IV.4. Sideward Movement and Case Theory

In section II.14.4, I proposed that English nominal elements are underspecified in terms of the specific kind of Case-feature they bear and that the morphological subcomponent realizes pronouns with unchecked Case-features as accusative. This proposal enabled us to account for existential constructions such as (112) below. In the mapping from N to \( \lambda \), the pronoun of (112) has an underspecified Case-feature. Thus, at LF its formal features can adjoin to T and check the Case-feature of T and its own Case-feature; at PF the pronoun is realized as the accusative him by default, since its Case-feature has not been checked.

(112) (there's Mary, there's Sue and) there's him/*he

This proposal also allows us to account for sentences such as (113a), represented in (113b), under the sideward movement analysis of parasitic Lightfoot (p.c.). (i) appears to indicate that wanna-contraction is possible only if both instances of to cliticize in an across-the-board fashion. This conclusion is not forceful, though. The pattern in (i) may follow from the two possibilities of conjunction represented in (ii); only (iib) seems to permit contraction, apparently yielding an across-the-board effect.

(i) a. I want to dance and to sing.
   b. I wanna dance and sing.
   c. I wanna dance and (*to) sing.

(ii) a. [ want [BP [ to dance ] [B' and [ to sing ] ] ] ] ]
    b. [ want to [ BP [ dance ] [B' and [ sing ] ] ] ]

---

Lightfoot (p.c.). (i) appears to indicate that wanna-contraction is possible only if both instances of to cliticize in an across-the-board fashion. This conclusion is not forceful, though. The pattern in (i) may follow from the two possibilities of conjunction represented in (ii); only (iib) seems to permit contraction, apparently yielding an across-the-board effect.
gaps outlined in section IV.2.1 (see IV.2.3 and fn. 22). If Case-features in English nominals are underspecified in terms of their values, the copy of which papers in the subject of were unavailable in (113b) is able to check nominative Case, whereas the copy in the object of reading is able to check accusative Case.

(113) a. Which papers did John say were unavailable without reading?

b. $[CP \ [ \text{which papers} ]^{i} \ \text{did} \ [TP \ \text{John} \ [VP \ [VP \ \text{say} \ [ \text{which papers} ]^{i} \ \text{were unavailable} \ [PP \ \text{without reading} \ [ \text{which papers} ]^{i} \ ] ] ] ]$

Interesting questions arise with respect to languages with a rich morphological case system, where it is plausible to assume that nominal elements have their Case-features fully specified as they enter into the numeration. Consider the Russian parasitic gap constructions in (114) (from Franks 1993:525), for example.
(114) a. mal’čik, [ *kotoromu/*kotorogo ]i Maša davala boy who(DAT.)/(GEN.) Masha(NOM.) gave den’gi ti do togo, kak (ona) stala izbegat’ PG_i money until she started to-avoid 'the boy who Masha gave money to until she started to avoid him'

b. devuška, [ kotoroj ]i Ivan daval den’gi ti girl who(DAT.-GEN.) Ivan(NOM.) gave money do togo, kak (on) stal izbegat’ PG_i until he started to-avoid 'the girl who Ivan gave money to until he started to avoid her'

In (114), the verb davala 'gave' requires a dative NP as its indirect object, and the verb izbegat’ 'to avoid' requires that its direct object have genitive case morphology.32 In (114a), neither the masculine-singular-dative relative pronoun nor its genitive counterpart is able to license both the real gap and the parasitic gap. In (114b), on the other hand, the feminine singular

32 The issue of whether this requirement should be stated in terms of inherent Case assignment or structural Case checking is orthogonal to the purposes of the present discussion.
relative pronoun has a syncretic form for dative and genitive, which allows both gaps to be licensed.

A similar pattern can also be found in ATB constructions, as illustrated by the Polish constructions in (115) (from Dyla (1984:703)).\textsuperscript{33} The verbs lubi 'likes' and nienawidzi 'hates' in (115) require an accusative and a genitive object, respectively. Neither the accusative nor the genitive relative pronoun in (115a) is able to license the ATB extraction of the objects of lubi and nienawidzi. By contrast, the relative pronoun \textit{którego} in (115b), which is ambiguous between accusative and genitive, is able to license this ATB extraction.

\textsuperscript{33} For additional restrictions on these constructions, see Franks (1993).
a. dziewczyna, [ *która/*której ]i Janek lubi t
    girl, who(ACC.)/(GEN.) Janek(NOM.) likes

and Jerzy nienawidzi ti

'a girl who Janek likes and Jerzy hates'

b. chłopiec, którego Maria lubi ti
    boy who(ACC.-GEN.) Maria likes

and Ewa nienawidzi ti

' the boy who Maria likes and Ewa hates'

The analysis of parasitic gap and ATB constructions in terms of sideward movement and linearization of chains proposed in sections IV.2
and IV.3 has a straightforward account of the pattern of acceptability illustrated in (114) and (115). Let us assume that lexical items with different morphological cases are to be treated as distinct as they enter into the numeration in languages with a rich morphological case system.34 If so, at a given step of the derivation that yields (114a) and (115a), a distinct wh-

---

34 See Uriagereka (1994, forthcoming:chap. 4), who suggests that languages which encode number and gender distinctions may treat lexical items with identical features other than number or gender as distinct in numeration.
phrase is merged in each object position, as respectively sketched in (116a) and (116b), where glosses were used instead of the actual words.

(116) a. \[TP \text{Masha} \ [VP \ [VP \text{gave money who}_{\text{DAT.}}] \ [PP \text{until she started to-avoid who}_{\text{GEN.}}]]] \]

b. \[BP \[TP \text{Janek likes who}_{\text{ACC.}}] \ [B' \text{and [TP Jerzy hates who}_{\text{GEN.}}]]\]

Suppose, for instance, that the first relative pronoun of each structure in (116) moves to Spec of CP, as shown in (117). The only nontrivial wh-chain that is formed in either structure is the one between the two nondistinct copies; the wh-phrase inside the adjunct in (117a) or the second conjunct in (117b) cannot form a chain with either of the two other wh-phrases because it is distinct from them. After the structures in (117) are spelled out, Chain Reduction will delete the lower copy of \(\text{who}_{\text{DAT.}}\) in (117a) and the lower copy of \(\text{who}_{\text{ACC.}}\) in (117b). The relative pronoun \(\text{who}_{\text{GEN.}}\) in both (117a) and (117b) cannot be deleted by Chain reduction, because it is a trivial chain. The unacceptability of the constructions in (114a) and (115b) therefore is due to the unmotivated application of deletion to the wh-word inside the adjunct or the second conjunct. Similar results would obtain if the second relative pronoun of the structures in (116) were the ones to move to Spec of CP.
(117)  

(a) \[ \text{[NP [NP boy ] [CP who(DAT.)i [TP Masha [VP [VP gave money who(DAT.)i [PP until she started to-avoid who(GEN.) ] ]]}}] \]

(b) \[ \text{[NP [NP girl ] [CP who(ACC.)i [ [BP [TP Janek likes who(ACC.)i ] ] ] [B' and [TP Jerzy hates who(GEN.) ] ]] ]} \]

As for the acceptability of the sentences in (114b) and (115b), let us make the plausible assumption that syncretic forms in languages such as Russian or Polish are underspecified with respect to the specific morphological cases they are compatible with. Thus, multiple instances of a syncretic form in a derivation may arise through applications of Select, in which case the relevant forms will be treated as distinct, or through applications of Copy, in which case the relevant terms will be treated as nondistinct (see section II.10.4). Constructions such as (114b) and (115b) can be accounted for if the multiple instances of the syncretic form arise via the Copy operation. In other words, (114b) and (115b) must involve sideward movement of the relative pronoun, as illustrated in (118), followed by movement to Spec of CP, as shown in (119).

(118)  

(a) \[ \text{[TP Ivan [VP [VP gave money who(DAT.-GEN.)i-WH ] [PP until he started to-avoid who(DAT.-GEN.)i-WH ] ]]} \]

(b) \[ \text{[BP [TP Maria likes who(ACC.-GEN.)i-WH ] [B' and [TP Jerzy hates who(ACC.-GEN.)i-WH ] ]]} \]
In both (119a) and (119b), economy considerations lead the wh-phrase in the Spec of CP to check its wh-feature for PF purposes (see sections III.6.2.5 and IV.1). In both structures, the copy in the Spec of CP can form a different chain with each of the other copies. Assuming that these chains are formed before Spell-Out, Chain Reduction will apply to each chain individually, always keeping the copy in the Spec of CP, since it has its wh-feature checked (see section III.6.2.3). The output of these applications of Chain Reduction is respectively given in (120) and (121), which then yields the PF output associated with (114b) and (115b).

(120) \[[NP [NP girl] [CP who(\text{DAT-GEN}.)]_\text{-WH} [TP Ivan [VP [VP gave money who(\text{DAT-GEN}.)]_\text{-WH} [PP until he started to-avoid who(\text{DAT-GEN}.)]_\text{-WH} ]]]]\]

(121) \[[NP [NP boy] [CP who(\text{ACC-GEN}.)]_\text{-WH} [BP [TP Maria likes who(\text{ACC-GEN}.)]_\text{-WH} [b and [TP Jerzy hates who(\text{ACC-GEN}.)]_\text{-WH} ]]]]]
The requirement that parasitic gaps and real gaps, on the one hand, and the gaps in ATB constructions, on the other, should match in terms of their morphological case in languages with a rich morphological case system thus follows from the fact that these elements are nondistinct copies related by the Copy operation.

IV. 5. Conclusion

Parasitic gap and ATB constructions pose very interesting questions for theoretical linguistics. Given the scarcity of these constructions as primary linguistic data, it is unlikely that a child can learn their properties by being exposed to them. This is even more so, if children are degree-0 learners (see Lightfoot (1991)); since parasitic gap and ATB constructions typically involve multiclausal structures, their properties presumably cannot be learned by degree-0 learners. Under standard assumptions of the Principles and Parameters Theory, it is also unlikely that UG has principles with the specific task of governing the well-formedness of parasitic gap and ATB constructions. Their properties should rather follow from general principles which are independently motivated (see the discussion in Chomsky (1982: sec. 4), for example).

In this chapter, I have developed an analysis of parasitic gap and ATB constructions which attempts to meet these conceptual considerations in a
Minimalist fashion. I have shown that under the Copy + Merge theory of
movement developed in chapters II and III, the same principles and
operations which account for the properties of standard movement
operations also account for the movement properties of parasitic gap and
ATB constructions. Under the Copy + Merge theory of movement, Move is
not an operation of the computational system; rather, it is a description of
the interaction of the operations Copy, Merge, Form Chain, and Chain
Reduction. Once movement operations are so understood, parasitic gap and
ATB constructions are taken to involve instances of "sideward movement",
where the computational system makes a copy $T'$ of the term $T$ of a given
syntactic object $K$ and merges $T'$ with a syntactic object $L$, unconnected with
$K$.

Recall that the possibility of sideward movement in the grammar is
constrained by the conditions on Form Chain, repeated below in (122), and
by the way how nontrivial chains are linearized in the phonological
component (see section IV.2.1).
(122)  *Form Chain:*

The syntactic objects $\alpha$ and $\beta$ can form the nontrivial chain $CH = (\alpha, \beta)$ only if:

(i) $\alpha$ c-commands $\beta$;

(ii) $\alpha$ is nondistinct from $\beta$;

(iii) (at least) one given feature $F$ of $\alpha$ enters into a checking relation with a sublabel of $K$, where $K$ is the head (of the projection) with which $\alpha$ merges, and the corresponding feature $F$ of $\beta$ could enter into a checking relation with a sublabel of $K$; and

(iv) there is no syntactic object $\gamma$ such that $\gamma$ has a feature $F'$ which is of the same type as the feature $F$ of $\alpha$, and $\gamma$ is closer to $\alpha$ than $\beta$ is.

In the case sketched above, for instance, $T$ and $T'$ cannot form a chain, because neither copy c-commands the other. Thus, if after sideward movement, the chains involving the nondistinct terms $T$ and $T'$ do not have a link in common, $T$ and $T'$ will induce violations of the irreflexivity and asymmetry conditions on linear order, preventing the structure from being linearized and canceling the derivation. Crucially, deletion of terms in
the phonological component is a suboperation of Chain Reduction, which only applies to nontrivial chains; hence, sideward movement of *which paper* in (123a) below, for instance, cannot yield the parasitic gap construction in (123b). Thus, under the framework developed here, the effects of the S-Structure licensing condition on parasitic gaps are accounted for straightforwardly, without postulating non-interface levels (see section IV.2.2).

(123)  
a. [ who [ [ filed [ which report ]\(^i\) ] [ without reading [ which report ]\(^i\) ] ] ]  
b. who filed [ which report ]\(^i\) without reading PG\(^i\)

If the chains involving T and T\(^i\) happen to have one link in common, on the other hand, linearization of the relevant structure is possible after Chain Reduction applies to each chain individually. In the structure in (124a), for instance, which involves sideward movement of *which paper* from one object position to the other, the highest copy of *which paper* forms a different chain with each of the lower copies. After both chains undergo Chain Reduction in an optimal way (minimizing applications of FF-Elimination), the structure in (124b) is derived and surfaces as (124c) (see section IV.2.1).
(124)  

a. \[[_{\text{CP}} [\text{which paper }]_{\text{WH}} Q+\text{did} [_{TP} \text{you} [_{\text{VP}} \text{file} \\
\text{[which paper]}_{\text{WH}}] ] ]\] 

b. \[[_{\text{CP}} [\text{which paper }]_{\text{WH}} Q+\text{did} [_{TP} \text{you} [_{\text{VP}} \text{file} ]_{\text{PP}} \text{without PRO reading} ] ] ]\] 

c. Which paper did you file without reading?

Consider now the unacceptability of (125a), which is represented in (125b). (125a) is standardly taken to show that the licensing of parasitic gaps is subject to an additional requirement of anti-c-command. Under the Copy + Merge approach to movement operations, the impossibility of (125) follows from the fact that the structure in (125b) cannot be linearized. The copy of who inside the adjunct PP cannot form a chain with either of the other copies because of the c-command or the Minimal Link Condition (see section IV.2.3 for details). The copy inside the PP and either of the other copies of who will then induce violations of the irreflexivity and asymmetry conditions on linear order, canceling the derivation because no PF object is formed. Again, an apparently idiosyncratic condition on parasitic gap licensing is derived within the Copy + Merge theory of movement by the same principles and operations which account for standard instances of movement.
(125)  a. *Who, it called you before you met PG? 
  
  b. \([CP Q [TP who-CASE \{VP \{who-CASE called you \} [PP before you met who-CASEi \} \} \} \) \]

Finally, the requirement that the "real" gap and the parasitic gap or the gap in a second conjunct in ATB extraction must have the same morphological case in a language with rich morphology follows straightforwardly if parasitic gaps and the gap in the second conjunct are copies of the "real" gap.

Still to be accounted for (within the Minimalist Program) is the fact that parasitic gaps are subject to island effects and to categorial and "referential" restrictions. As discussed in section IV.4, however, wh-phrases in situ, which according to Chomsky (1995:291) are unselectively bound by an interrogative complementizer and do not involve covert adjunction to the complementizer, are also subject to these restrictions. Thus, it is plausible to assume that these are not conditions on "movement" (Form Chain), but on operator variable relations. If so, whatever accounts for the restrictions on the distribution of wh-phrases in situ should in principle be able to account for the distribution of parasitic gaps, as well.

To the extent that parasitic gap and ATB constructions are handled in the framework developed in chapters II and III without any construction specific operations or the introduction of principles which are not
independently needed, they lend strong conceptual and empirical support to the Copy + Merge theory of movement and the proposal about linearization of chains.
CHAPTER V

CONCLUSION

This dissertation has explored the simplest version of the copy theory of movement, according to which traces and heads of chains are not intrinsically distinct, but are rather subject to the same principles and accessible to the same operations of the computational system. In the approach pursued here, the differences between traces and heads of chains arise in the course of the derivation due to the fact that checking is a very local operation which affects only the chain link which is the checking domain of a given head. Such differences then provide an economy criterion for the phonological component to choose among the copies to be deleted.

I have proposed that nondistinct terms induce violations of the irreflexivity and asymmetry conditions on linear order, preventing the structure from being linearized and canceling the derivation due to the fact that no PF object can be formed. Deletion of all but one chain link is then forced upon nontrivial chains in order to allow these chains to be linearized in accordance with Kayne's (1994) LCA. Given that the head of a chain enters in more checking relations than its trace(s), it will always have fewer (if any)
visible formal features to be eliminated in the phonological component. Deleting traces for linearization purposes is therefore more economical than the deletion of heads of chains, in that applications of FF-Elimination are minimized. By accounting for the lack of phonetic realization of traces, this proposal about optimal linearization of chains allows us to dispense with the notion of a trace as a primitive of linguistic theory.

Another focus of this dissertation has been the movement operation itself. I have proposed that Move is not an operation of the computational system. It is rather a description of the interaction between the operations Copy, Merge, Form Chain, and Chain Reduction. Crucial evidence for this proposal was provided by instances of sideward movement, where applications of Copy and Merge were dissociated from applications of Form Chain and Chain Reduction. It was shown that under this conception of movement operations, the core properties of parasitic gap and across-the-board extraction constructions can be subsumed under the properties of standard movement, without the postulation of construction-specific operations or non-interface level conditions.

I would like to point out that more important than being THE correct approach to movement operations, which certainly is not the case, the line of reasoning pursued in this dissertation seems to me to be of right TYPE of approach, in that it sticks to Minimalist assumptions in exploring the simplest possible copy theory of movement. To the extent that something
along the lines described above may be on the right track, we have given some steps towards a better understanding of movement operations and chains, which lie at the heart of the language faculty.

This kind of approach seems to have large consequences for the system as a whole. As suggested in the appendix of chapter III (see also Nunes (1995b)), an extension of the analysis developed here may account for why null operators are phonologically null, for instance. Also, if it turns out to be true that a chain is not interpreted as a whole at the C-I interface, but only one of its links receives the appropriate interpretation (see Hornstein (1995:chap. 8)), the proposal about linearization of chains may perhaps extend to LF as well. In fact, the defense version of this dissertation included a chapter (cited here as Nunes (1995a)), where I explored this possibility and proposed that reconstruction effects can be derived from linearization of nontrivial chains at LF. Time constraints prevented me from revising this chapter for the present version. Since (unfortunately) one does get tenure for being a graduate student, I will surrender to the deadline pressure and leave a full discussion of these and other potential implications of the Copy + Merge theory of movement and the hypothesis on linearization of chains at LF for future work.
References


