Sideward Movement

Jairo Nunes

Assuming the general framework of the Minimalist Program of Chomsky 1995, this article argues that Move is not a primitive operation of the computational system, but rather the output of the interaction among the independent operations Copy, Merge, Form Chain, and Chain Reduction (deletion of chain links for purposes of linearization). The crucial aspect of this alternative model is that it permits constrained instances of *sideward movement,* whereby a given constituent ''moves'' from a syntactic object K to an independent syntactic object L. This version of the copy theory of movement (a) provides an explanation for why (some) traces must be deleted in the phonological component, (b) provides a cyclic analysis for standard instances of noncyclic movement, and (c) accounts for the main properties of parasitic gap and across-the-board extraction constructions.

Keywords: sideward movement, copy theory, Linear Correspondence Axiom, traces, parasitic gaps, across-the-board movement

1 Introduction

A fundamental property of human languages is that elements may be interpreted in positions different from the ones where they are phonetically realized. Within the principles-and-parameters framework (see Chomsky 1981, Chomsky and Lasnik 1993), this ''displacement property'' is captured by means of a movement operation relating structural positions in a phrase marker. With the recent developments of the principles-and-parameters framework that have culminated in the Minimalist Program (see Chomsky 1993, 1994, 1995), the operation *Move* is specifically described as follows (see Chomsky 1994:fn. 13, 1995:250): given the syntactic object Σ with constituents K and α, Move targets K, raises α, and merges α with K, forming Σ'; the operation is cyclic if $\Sigma = K$ and noncyclic otherwise. Σ' differs from Σ in that K is replaced by $L = \{\gamma, \{\alpha, K\}\}$ or $L = \{\langle\gamma, \gamma\rangle, \{\alpha, K\}\}$, depending on whether movement proceeds by substitution or adjunction. Move also forms the chain $CH = (\alpha, t)$, a two-element pair where t (the trace of α) is a copy

of α that is deleted in the phonological component in the case of overt movement, but remains available for interpretation at LF (see Chomsky 1993:35). Under this view, the ''displacement property'' of human languages is thus taken to involve (a) copying, (b) merger, (c) chain formation, and (d) deletion of traces (lower copies) for PF purposes.

As discussed by Chomsky (1993), the interpretation of movement operations in terms of copying accords well with the general conceptual concerns of the Minimalist Program in that it allows binding theory to be stated solely at LF without resorting to noninterface levels, it provides the basis for the interpretation of displaced idiom chunks at LF, and it eliminates reconstruction as an additional operation of the computational system. However, the analysis of Move as an operation as complex as described above has some conceptual problems.

The most obvious one concerns the lack of motivation for deletion of traces (lower copies) in the phonological component.[1] If traces are true copies, why can they not be phonetically realized, behaving like the head of the chain? Another conceptual problem with the computational system as proposed in Chomsky 1994, 1995, 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 (see Gärtner 1997). 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 that contains both notions is redundant.

The overall complexity of the operation Move can be seen as a historical residue of the description of generalized transformations as binary or singular operations: concatenative binary transformations such as Merge target two disconnected syntactic objects, whereas singular transformations such as Move target two constituents of a single syntactic object (see Chomsky 1975, 1993, Kitahara 1995). Assuming the general framework outlined in Chomsky 1995, I propose here that Move should be understood not as a primitive singular operation of the computational system, but as merely the description of the interaction of the independent operations *Copy, Merge, Form Chain,* and *Chain Reduction* (deletion of copies for purposes of linearization). Under this view, ''movement'' operations will always involve the concatenation of two independent syntactic objects. I refer to this approach as the *Copy + Merge theory* of movement.

By decomposing Move into more basic independent operations, the Copy + Merge theory of movement yields a much simpler system that not only overcomes the conceptual problems raised above, but also has a broader empirical coverage than the versions of the copy theory explored in Chomsky 1993, 1994, 1995. In particular, the Copy + Merge theory allows constrained instances of what I have called *sideward movement* (see Nunes 1995),[2] where the computa-

---

[1] This problem arises regardless of whether deletion of traces is conceived of as part of the inner workings of Move or as an independent operation.

[2] The sequence of derivational steps described in (1) is referred to as an *interarboreal operation* by Bobaljik and Brown (1997) and as *paracyclic movement* by Uriagereka (1998).

tional system copies a given constituent α of a syntactic object K and merges α with a syntactic object L, which has been independently assembled and is unconnected to K, as illustrated in (1).[3]

(1) a. $[_K \ldots \alpha^i \ldots]$ α^i $\xleftarrow{\textbf{Merge}}$ $[_L \ldots]$
              $\underline{\qquad\textbf{Copy}\qquad}$

     b. $[_K \ldots \alpha^i \ldots]$ $[_M \alpha^i [_L \ldots]]$
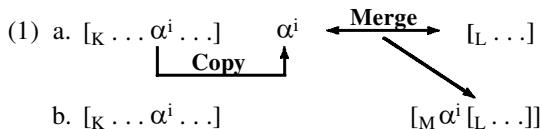
The article is organized as follows. In section 2, I summarize my earlier (1995, 1996, 1999) proposal that deletion of traces is the result of linearization of chains in the phonological component; the conditions regulating chain formation and chain linearization discussed in this section prove to be decisive in allowing or ruling out particular instances of sideward movement. In section 3, I show that if movement is broken down into the basic operations mentioned above, noncyclic movement operations can be eliminated from the grammar. In sections 4 and 5, I show how the movement theory proposed here accounts for the major properties associated with parasitic gap and across-the-board extraction constructions. Finally, in section 6, I present a brief conclusion.

2 Deletion of Traces

2.1 The Puzzles

Any analysis assuming the copy theory of movement has to face two questions. First, why is it the case that a nontrivial chain (in general) cannot have all of its links phonetically realized? That is, why can the structure in (2), for instance, not be realized as in (3c)? And second, why is it the case that (in general) traces and not heads of chains are the links that are deleted ((3a) vs. (3b))?

(2) $[_{TP} \text{John}^i [_{T'} \text{T} [_{VP} \text{was} [_{VP} \text{kissed John}^i]]]]$

(3) a. John was kissed.
     b. *Was kissed John.
     c. *John was kissed John.

The unacceptability of (3c) resulting from the derivation in (2) raises an additional puzzle within the minimalist framework. The derivations of (3a) and (3b) from (2) involve an operation eliminating one of the copies of *John,* whereas no such operation is invoked in (3c). Thus, were the derivations of (3a) and (3b), on the one hand, and the derivation of (3c), on the other, to be compared for economy purposes, the derivation of (3c) should be preferred over the other two because it involves fewer operations, therefore being more economical. Since (3c) is unacceptable,

---

[3] With the elimination of D-Structure in the Minimalist Program (see Chomsky 1993:sec. 3), assembly of unconnected phrase markers is independently required for any phrase involving a complex specifier or a complex adjunct.

Throughout the article, I use superscripted indices to annotate copies.

the logic of the system leads us to expect its derivation from (2) to either crash or be canceled, thereby becoming irrelevant for economy computations.[4]

2.2 Deletion of Traces as Optimal Linearization of Chains

Once the copy theory of movement is assumed, the system must be able to determine whether two elements with the same set of features are to be interpreted as copies or distinct elements.[5] The structure in (2), for instance, should be prevented from converging at LF if the two occurrences of *John* are not copies. In this article, I adopt Chomsky's (1995:227) proposal that two lexical items *l* and *l′* are marked as distinct for the computational system if they are formed by distinct applications of Select accessing the same lexical item of a numeration; hence, the structure in (4a), for example, is taken to represent a derivation built from a numeration with two instances of *John,* as indicated by the indices.

(4) a. $[_{TP}$ John^i T $[_{vP}$ John^i $[_{v′}$ said $[_{CP}$ that $[_{TP}$ John^k was $[_{vP}$ kissed John^k]]]]]]
    b. John said that John was kissed.

   As for phrasal objects, I assume that their labels encode the relevant piece of information regarding distinctiveness; the phrasal object *the man,* for instance, is to be represented as K = {the^i, {the^i, man^k }}. Given that the Copy operation just replicates the targeted material and does not alter distinctiveness markings, it is possible to determine, at any given point in a given derivation, whether two terms with identical sets of features are to be interpreted as copies or distinct elements.[6]

   Given that this notion of (non)distinctiveness is based on purely syntactic computations, it certainly plays no role at the PF level and at the articulatory-perceptual (A-P) interface. However,

---

[4] A derivation that satisfies Full Interpretation at a given syntactic level of representation is said to converge at that level; otherwise, it is said to crash at that level. A derivation is said to be canceled if an illegitimate operation is performed during the computation, if the pair (LF object, PF object) is not formed, or if the numeration is not exhausted (see Chomsky 1995:219–220, 225–226).

   One reviewer asks why we may not simply say that phonetic realization of more than one link, as in (3c), is impossible because chains are not PF objects. There are two reasons for not following this approach. First, as will be discussed in section 2.2, under certain special circumstances there can actually be cases in which more than one chain link is phonetically realized (see Nunes 1999, 2000, for more detailed discussion). Second, the reviewer's suggestion would still require an independent explanation for why the phonological component does not allow phonetic realization of all of the links by treating each link as if it were not part of a chain.

   [5] This is true regardless of whether the system is representational or derivational (see Brody 1995:136, Chomsky 1995:227).

   [6] One reviewer suggests that ''Copy need not be separate from selecting from the lexicon the same thing twice.'' A potential problem for this suggestion is that the Copy operation applies not only to lexical items, but also to phrasal objects built from lexical items; that is, Copy can duplicate the syntactic object $[_{NP}$ the man], but Select cannot have access to such an object in the lexicon. For such cases, the reviewer further suggests that Select could apply to the lexical subparts of phrasal objects.

   The problem I see with such an approach is that the two ''selections'' are treated differently by the computational system with respect to derivational cost. As Chomsky (1995, 2000) argues, if lexical insertion and movement (in the reviewer's suggestion, selection from the lexicon and selection from the available structure) were equally economical, the derivations in (ia) and (ib) should yield equal results since they share the same initial array and the same derivational path up to the point where the structure in (ic) is assembled; the fact that (ia) blocks (ib) indicates that movement is more

the fact that the only convergent output for (4a) is (4b), where the trace of each chain is deleted, indicates that the phonological component does take (non)distinctiveness into account when performing deletion, rather than randomly deleting terms with identical sets of features. Let us then make the plausible assumption that the notion of (non)distinctiveness is available for early computations of the phonological component when the syntactic structure that was shipped by Spell-Out is still available. More precisely, let us assume that the information regarding (non)distinctiveness is present at the point when the computational system applies the operation *Linearize,* which I take to be the operation of the phonological component that applies to the output of Morphology (see Chomsky 1995:340), and converts a syntactic structure into a sequence of X^0 elements in consonance with the Linear Correspondence Axiom (LCA) in (5) (from Kayne 1994: 33).[7]

> (5) *Linear Correspondence Axiom (LCA)*
> Let X, Y be nonterminals and x, y terminals such that X dominates x and Y dominates y. Then if X asymmetrically c-commands Y, x precedes y.

This being so, the two copies of *John* in (2), repeated in (6), should count as nondistinct for purposes of linearization. Let us then consider the asymmetric c-command relations in (6) to determine how this structure should be linearized in accordance with the LCA.[8]

> (6) $[_{TP}$ John^i $[_{T'}$ T $[_{VP}$ was $[_{VP}$ kissed John^i]]]]

Take the relation between the copies of *John* and the copula *was,* for instance. Since the upper copy of *John* asymmetrically c-commands *was,* we should obtain the order ⟨*John, was*⟩,

---

costly than lexical insertion, which in the system that I am arguing for amounts to saying that Copy is more costly than Select.

> (i) a. [there$_i$ seem t$_i$ to be cats in the garden]
> b. *[there seem cats$_i$ to be t$_i$ in the garden]
> c. [to be cats in the garden]

Other cases leading to the same conclusion within a sideward movement approach are discussed in detail in Hornstein and Nunes 1999. The authors argue, for instance, that if Copy were not a costly operation, *how* could undergo sideward movement from K to L in (ii), yielding the unacceptable parasitic gap construction in (iii).

> (ii) a. K = [after cooking the chicken how]
> b. L = [cook the pork]

> (iii) *[how$_i$ did you [cook the pork t$_i$] [after cooking the chicken PG$_i$]]

In the absence of an independent explanation for the difference between Select and Copy in terms of derivational cost, I will continue to treat them as distinct operations of the computational system and will adopt Chomsky's (1995: 227) technical implementation of Select described in the text. I would like to thank Diane Lillo-Martin for helpful discussion of this issue.

[7] In Chomsky 1995, 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 1995:337). I will avoid this ambiguity and employ the term *Linearize* to refer to the relevant operation.

[8] I put aside the question of how two heads in a mutual c-command relation such as *kissed* and *John* in (6) can be linearized in compliance with the LCA (see Kayne 1994:10, Chomsky 1995:337), which is orthogonal to the issue under discussion. Also orthogonal for current purposes is how (5) should be revised in order to be compatible with bare X-bar theory (see Uriagereka 1998 for relevant discussion).

according to the LCA; likewise, since the copula asymmetrically c-commands the lower copy of *John,* the order ⟨*was, John*⟩ should be derived. Combining these two results, we should obtain the partial sequence σ = ⟨*John, was, John*⟩. Were the two instances of *John* distinct, σ would be a well-formed linear order, with the copula following an occurrence of *John* and preceding a different occurrence of *John,* like *said* and the two distinct instances of *John* in (4b), for example. However, since the two instances of *John* in (6) are nondistinct, *was* is predicted to precede and be preceded by the same element, namely, *John.* σ is therefore not a linear order because it lacks asymmetry, which is a defining property of a (strict) linear order (if α precedes β, then it must be the case that β does not precede α). The structure in (6) also violates the irreflexivity condition on (strict) linear order (if α precedes β, then it must be the case that α ≠ β). 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 (6) are nondistinct, that would amount to saying that *John* should precede (and follow) itself.

We thus have a straightforward explanation for why the structure in (6) cannot surface as (3c), repeated in (7c): as is, (6) cannot be linearized and the attempted derivation of (7c) is canceled because no PF object can be formed (see footnote 4). This is a welcome result. Recall that the derivation of (7c) from (6) should be prevented from being compared with the derivation of (7a) for purposes of economy, even if they have the same reference set; otherwise, the derivation of (7c), which employs no application of chain link deletion, would wrongly rule out the derivation of (7a), which employs one such application (see section 2.1). If the derivation of (7c) is canceled because it cannot be linearized, it cannot be compared with the derivation of (7a), because only convergent derivations can be compared for economy purposes (see Chomsky 1995).

(7) a. John was kissed.
    b. *Was kissed John.
    c. *John was kissed John.

Following earlier work (Nunes 1995, 1996, 1999), I assume that in order to linearize a given structure containing a nontrivial chain, the phonological component resorts to the operation Chain Reduction, described in (8).

(8) *Chain Reduction*
    Delete the minimal number of constituents of a nontrivial chain CH that suffices for CH to be mapped into a linear order in accordance with the LCA.

Applying to the NP chain in (6), Chain Reduction deletes either the upper or the lower copy of *John* (see (7a–b)), eliminating the asymmetry and irreflexivity problems mentioned above and allowing either resulting structure to be linearized in accordance with the LCA. The choice between these two derivations will then depend on which of the copies is in some sense optimal, as will be discussed below.

Two things are worth mentioning before we address the contrast between (7a) and (7b). First, although I will assume the formulation in (8) for purposes of presentation, it is not necessary to specify that Chain Reduction delete the *minimal number* of constituents (each application of

deletion targeting a given constituent).[9] Economy considerations regarding the length of a deriva-
tion may indirectly determine the number of elements to be deleted, by enforcing the minimal
number of applications of a given operation. All things being equal, a short derivation should
block a longer derivation (see Chomsky 1995:314, 357); hence, a derivation in which elements
are unnecessarily deleted is longer, therefore less economical, than a competing derivation where
no such deletion occurs. Under this view, recoverability of deletion follows from economy consid-
erations; deletion of the two links of (6) yielding (9), for instance, allows the structure to be
linearized but is less economical than the derivations of (7a) and (7b), which involve a single
application of deletion.[10]

    (9) *Was kissed.

    The second point to be borne in mind is that, as stated in (8), deletion of chain links is
triggered by linearization considerations as regulated by the LCA. This predicts that if a given
chain link is invisible to the LCA, it will not induce violations of the asymmetry and irreflexivity
conditions on linear order with respect to the other links and the relevant chain will be allowed
to surface with more than one link phonetically realized. For instance, let us assume, following
Chomsky (1995:337), that the LCA does not apply word-internally;[11] if so, a given chain link
that is below the word level will not be computed by the LCA and should be pronounced regardless
of the other links.

    That this prediction is correct is suggested by clitic duplication in some dialects of Argentin-
ean Spanish, for instance (see Nunes 1999, 2000, for further details and discussion).[12]

---

[9] Similar considerations extend, mutatis mutandis, to FF-Elimination (see (17)) and Chain Uniformization (see
footnote 18).

[10] One reviewer asks why deletion of copies must make reference to chains and not simply operate up to recoverability.
As will be discussed in detail in the next sections, free deletion of copies (up to recoverability) actually leads to overgenera-
tion, by ruling in unwanted instances of sideward movement. On the other hand, the requirement that deletion of copies
only target chains, as formulated in (8), restricts sideward movement, yielding the right results: if the copies produced
in the course of the computation fail to form a chain, deletion under Chain Reduction is inapplicable and the relevant
structure cannot be linearized.

[11] Kayne (1994:sec. 4.5) actually proposes that the LCA should also apply below the word level. However, the data
to be discussed below in the text indicate that (at least) the internal structure of words resulting from morphological
reanalysis is not subject to the LCA. Independent evidence that the LCA does not determine the order of morphemes of
words formed in the course of the derivation is provided by the European Portuguese sentence in (ia), which is arguably
derived from (ib) after the preposition *de* adjoins to the auxiliary *hei* and the resulting X^0 adjoins to C^0. If the LCA
applied word-internally, we would incorrectly predict that in (ia), *de* should precede rather than follow *hei,* given that
the preposition asymmetrically c-commands the auxiliary in the final structure. Thanks to Ana Maria Martins for helpful
discussion.

  (i) a. O que hei-de   eu fazer?
        what have.1SG-to I   do
        'What can I do?'
   b. [$_{CP}$ C^0 [$_{TP}$ eu [hei [de [fazer [o que]]]]]]

[12] Clitic duplication is also found in other varieties of South American Spanish, with a very interesting dialectal
variation. I am thankful to Monica Zoppi-Fontana and Marcela Depiante for discussion of the data in (10)–(11).

(10) a. *Nos* vamos acostumbrando   a este pais   poco a   poco.
          us$_{CL}$ go.1PL getting.accustomed to this country little by little

   b. Vamos acostumbrándo*nos*   a este pais   poco a   poco.
          go.1PL getting.accustomed-us$_{CL}$ to this country little by little

   c. Vámo*nos*   acostumbrándo*nos*   a este pais   poco a   poco.
          go.1PL-us$_{CL}$ getting.accustomed-us$_{CL}$ to this country little by little

   d. *Nos* vamos acostumbrándo*nos*   a este pais   poco a   poco.
          us$_{CL}$ go.1PL getting.accustomed-us$_{CL}$ to this country little by little
          'We are getting accustomed to this country little by little.'

The data in (10a–b) illustrate the general paradigm of clitic placement in Spanish, with the object clitic preceding a finite form or following a nonfinite form. What is relevant for our discussion is that in the dialects under consideration, clitic duplication may be allowed, but only if the higher copy is enclitic ((10c) vs. (10d)). Interestingly, this pattern does not change even if the higher verb is in the subjunctive mood, which generally requires proclisis.

(11) a. para que vayámo*nos*   acostumbrándo*nos*   a este pais
          for   that go.SUBJ.1PL-us$_{CL}$ getting.accustomed-us$_{CL}$ to this country

   b. *para que *nos* vayamos   acostumbrándo*nos*   a este pais
          for   that us$_{CL}$ go.SUBJ.1PL getting.accustomed-us$_{CL}$ to this country
          'in order for us to get accustomed to this country'

Let us assume that when the clitic climbs, it adjoins to the left of a functional category F with the verb adjoined to it, as represented in (12) (see, e.g., Kayne 1991, Uriagereka 1995). As is, the structure in (12) cannot be linearized because the two copies of the clitic induce violations of the asymmetry and irreflexivity conditions on linear order; hence the unacceptability of sentences such as (10d) and (11b).

(12)
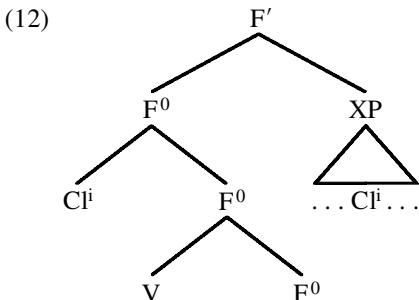
The correlation between exceptional enclisis and clitic duplication in (10) and (11) can be accounted for if we assume that in these dialects, Morphology can reanalyze the three-segment category in (12) as a word, rendering the adjoined clitic invisible to the LCA. Taking enclisis to be the reflex of such morphological restructuring, we would expect clitic duplication to always cooccur with exceptional enclisis: after the three-segment F^0 in (12) is restructured as a word, the only copy of the clitic that is visible to the LCA is the lower one, and it need not (therefore

must not) be deleted by Chain Reduction; hence the contrast between (10c) and (10d) and between (11a) and (11b).[13]

The proposal that chain links below the word level are invisible to the LCA also provides an account for why some languages allow an intermediate *wh*-trace in C to be phonetically realized if it is monomorphemic or has an incorporated preposition, but not if it is a full *wh*-phrase, as illustrated by the German examples in (13)–(15) (from Fanselow and Mahajan 1996).[14]

(13) *Wen*   meinst du   *wen*   sie liebt?
     whom believe you whom she loves
     'Who do you think that she loves?'

(14) *Wovon* glaubst du   *wovon* sie träumt?
     of-what believe you of-what she dreams
     'What do you believe that she dreams of?'

(15) *Welchen Mann glaubst du   *welchen Mann* sie liebt?
      which   man believe you which   man   she loves
     'Which man do you believe that she loves?'

Suppose that in these languages, successive-cyclic *wh*-movement may proceed by adjunction to a $[-\text{wh}]$ C^0, as schematically represented in (16),[15] and that Morphology may reanalyze the adjunction structure $[_C \text{wh}^i [_C \circ C^0]]$ into a word, along the lines of the reanalysis of F^0 in (12).[16]

(16) $[_{CP} \text{wh}^i [_{C'} Q [_{TP} \ldots V [_{CP}[_{C^\circ} \text{wh}^i [_{C^\circ} C^0]] [_{TP} \ldots V \text{wh}^i]]]]]$

---

[13] Two things are worth mentioning. First, I am not claiming that the output of morphological reanalysis always involves enclisis. See Nunes 2000, for instance, for cases of clitic duplication involving proclisis to the finite verb that are arguably subject to the same analysis proposed here. Second, I am not claiming that every head adjunction leads to morphological reanalysis; otherwise, standard verb movement to T, for example, would necessarily involve verb duplication (phonetic realization of both the moved verb and its trace). The fact that clitic duplication in (10c) and (11a) does not allow concomitant verb duplication indicates that the moved verb is still visible to the LCA after restructuring. Three possibilities come to mind that would derive the correct results: (a) the clitic and the verb are adjoined to different functional categories, (b) the clitic adjoins to V and the two-segment V category is the one that is restructured, and (c) the category resulting from restructuring the three-segment F^0 structure in (12) is actually V, rather than F. I leave the choice between these alternatives open, pending further research.

[14] Besides German (see, e.g., McDaniel 1986, Fanselow and Mahajan 1995), this phenomenon is found in Afrikaans (see Du Plessis 1977), Frisian (see Hiemstra 1986), Romani (see McDaniel 1986), and English child grammar (see Thornton 1990).

[15] On the optimality considerations for checking configurations that may block movement to specifiers in favor of adjunction to heads, see Nunes 1998.

[16] Fanselow and Mahajan (1995) propose that in German, intermediate *wh*-traces may cliticize to C^0 in the phonological component, thereby becoming invisible to deletion. Although this proposal can explain why full *wh*-traces cannot be phonetically realized in intermediate trace positions (see (15)), it does not have a principled explanation for why in the general case, traces cannot be phonetically realized. The morphological restructuring proposed in the text is clearly seen in Frisian, where the complementizer appears cliticized to the *wh*-element, as illustrated in (i) (from Hiemstra 1986).

(i) Wa tinke jo wa't   ik sjoen haw?
    who think you who-that I seen have
    'Who do you think that I have seen?'

After Morphology restructures $[_{C^0}$ wh^i $[_{C^0}$ C$^0]]$ in (16) as C^0, the intermediate *wh*-copy becomes invisible to the LCA and need not (therefore must not) be deleted by Chain Reduction. The only copies that are relevant for deletion for purposes of linearization are therefore the head and the tail of the *wh*-chain, which should fall under the same analysis as (6), requiring deletion of the lower copy (see below). The contrast between (13) and (14), on the one hand, and (15), on the other, then follows if Morphology cannot deal with nonminimal maximal projections adjoined to heads (see Chomsky 1995:319). That is, if the ''morphological heaviness'' of a full *wh*-phrase prevents the adjoined structure from being reanalyzed as a word, a chain involving a full *wh*-phrase will always have all of its links visible to the LCA and Chain Reduction will have to delete all but one link; hence the unacceptability of (15). Under this view, the impossibility of a morphological reanalysis involving the higher complementizer in (16) is attributed to its [+wh] feature, which makes it morphologically heavy in the relevant sense; hence, only [−wh] C^{0}s are light enough to permit reanalysis and render an adjoined *wh*-element invisible to the LCA.[17]

Let us return to the issue of why deletion (in general) targets traces and not heads of chains. Elsewhere (Nunes 1995, 1996, 1999), I have proposed that the optimality of deletion for a given chain link is contingent on the elimination of formal features in the phonological component. Although formal features are relevant for morphological computations, they are not interpretable at PF (only phonological features are); thus, as noted by Chomsky (1995:230–231), an operation of the phonological component applying after Morphology must eliminate formal features that are visible at PF. For purposes of discussion, I will assume the technical implementation of this operation given in (17) (from Nunes 1995:291).

---

[17] Two remarks are in order here. First, to the extent that clitic duplication and *wh*-copying are adequately accounted for, the analysis of deletion of chain links in terms of the LCA is empirically more successful than other recent approaches such as Brody's (1995:106) Transparency Principle (''If all chain members c-commanded by the contentive element are copies of the contentive, then it must be the case that only the highest member of such a set of copies (i.e., the contentive itself) is visible for SPELLOUT''), Bobaljik's (1995b:350) Speak Up (''Pronounce the topmost/leftmost copy of each element''), Groat and O'Neil's (1996:135) Form Chain (''Forming a chain results in copying all syntactic features of the category moved, but does not copy the category's phonological matrix: it either moves it to the new position or fails to move it''), or Pesetsky's (1997:153) Silent Trace (''Don't pronounce the traces of a moved constituent'').

Second, I am assuming here that legibility requirements of the A-P system demand that the syntactic structure shipped to the phonological component by Spell-Out be mapped into a sequence of terminals, and that the LCA optimizes such mapping by resorting to the linear order that has already been established by asymmetric c-command; hence, the formulation of Chain Reduction in (8) takes the LCA into consideration. As an alternative, the reviewers suggest that deletion of copies need not make reference to the LCA and that its applications may be directly regulated by the irreflexivity and asymmetry conditions on precedence. It seems to me that the computations can proceed in a more local fashion if Chain Reduction can use the information regarding asymmetric c-command that is already available, without looking ahead to the precedence relations among the terminals yet to be established. However, if it turns out that the relevant mapping from syntactic structure to precedence is not governed by the LCA and that Chain Reduction takes into consideration the precedence relation itself, the only adjustment in the analysis proposed above that would need to be made in the face of empirical facts such as the ones in (10)–(11) and (13)–(15) (see also Nunes 1999, 2000) would be to say that elements below the word level are not computed with respect to the relevant precedence relation. I thank the reviewers for having brought this potential alternative to my attention. For concreteness, I will continue to assume that the general proposal underlying the LCA is correct and keep the formulation of Chain Reduction given in (8).

(17) *Formal Feature Elimination (FF-Elimination)*

Given the sequence of pairs $\sigma = \langle(F, P)_1, (F, P)_2, \ldots, (F, P)_n\rangle$ such that σ 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 formal features in order for σ to satisfy Full Interpretation at PF.

The difference between the head of a chain and its traces in terms of optimality now follows from the number of checking relations a given copy is associated with. Extending Chomsky's (1995:sec. 4.5.2) checking theory, I have proposed (Nunes 1995) that a formal feature becomes invisible at PF after being checked, regardless of its interpretability at LF. Thus, a checked feature need not (therefore must not) be eliminated by FF-Elimination, because it has already been rendered invisible at PF by a checking operation (see Nunes 1995, 1996, 1999, for details and discussion).

Bearing these considerations in mind, let us examine the Case feature of *John* in the course of the derivation of (6), for instance, as shown in (18). The Case feature of the upper copy of *John* becomes invisible at both LF and PF after being checked against the finite T head, as represented by the subscript. After (18) undergoes Chain Reduction for purposes of linearization, we obtain either (19a) or (19b), depending on which copy of *John* is deleted. In order to converge, the derivation operating with the structure in (19b) still requires an application of FF-Elimination targeting the unchecked Case feature, whereas no such application is required in (19a), because its Case feature became invisible at PF after being checked. The derivation in which Chain Reduction deletes the trace thus ends up being more economical than the one in which the head of the chain is deleted; hence the contrast in (20).[18]

(18) [John^i-$_{\text{CASE}}$ [was [arrested John^i-**CASE**]]]

---

[18] Assuming that a checked feature cannot enter into further checking relations regardless of the interface level at which it is rendered invisible, an overt checking operation will render a given interpretable feature F invisible at PF only if this does not cause the derivation to crash or to be canceled. The categorial feature of the intermediate copy of *John* in (i), for example, cannot be checked for PF purposes; otherwise, it will not be able to enter into a checking relation with the strong feature of the matrix T head (see Nunes 1995, 1999, for further discussion). Similar considerations apply to successive-cyclic *wh*-movement.

(i) [$_{TP}$ John^i-$_{\text{N-CASE}}$ [$_{T'}$ is likely [$_{TP}$ John^i-**N-CASE** to [$_{VP}$ be elected John^i-**N-CASE**]]]]

Questions arise about how the unchecked Case feature of the lower copy of *John* in (18) circumvents a Full Interpretation violation at LF. This problem is reminiscent of the problem that a sentence such as (ia), for instance, presents for Chomsky's (1995) system, where ''the features of a chain are considered a unit: if one is affected by an operation, all are'' (see Chomsky 1995:381, n. 12); after the formal features of the lower copy of *what* in (ib) raise in the covert component, a checking operation will obliterate the Case features of both links of the newly formed chain, but not the Case feature of the copy of *what* in [Spec, CP], which is part of the chain formed earlier in the overt syntax. Noting this problem, Chomsky (1995:303) further adds that ''a convention is then needed requiring erasure of F throughout the array of chains containing F, so that no −Interpretable feature remains in the operator position.''

(i) a. What did John see?
     b. [$_{CP}$ what^i-**CASE** did + Q [$_{TP}$ John see what^i-**CASE**]]

Assuming that traces are unaffected by the operations affecting heads of chains (see discussion in section 2.2), a single account of (18) and (ib) can be provided (see Nunes 1995) by implementing the convention suggested by Chomsky

(19) a. [John^i-$_{CASE}$ [was [arrested ~~John^i-CASE~~]]]
      b. [~~John^i-$_{CASE}$~~ [was [arrested John^i-**CASE**]]]

(20) a.   John was arrested.
      b. *Was arrested John.

2.3 Summary and Consequences

According to the proposal reviewed above, a syntactic object containing a nontrivial chain CH with links above the word level cannot be linearized in accordance with the LCA; since these links are nondistinct, they induce violations of the asymmetry and irreflexivity conditions on linear order, canceling the derivation because no PF object can be formed. To prevent this state of affairs, the phonological component resorts to the operation Chain Reduction, which in the standard case deletes all but one link of CH.[19]

In order for Full Interpretation to be satisfied at PF, FF-Elimination must then delete the unchecked formal features (if any) of the link that survives Chain Reduction. Since a given head only checks the relevant features of the chain link that is in its checking domain, the head of CH will always have fewer unchecked formal features (if any) to be deleted by FF-Elimination than the lower links. 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; the derivation in which the head of the chain survives Chain Reduction requires fewer (if any) applications of FF-Elimination than a derivation in which a trace survives.[20]

If deletion of traces reflects optimal applications of Chain Reduction in the phonological component, as argued above, the view of Move as a complex singulary transformation is substantially weakened. Among the four mechanisms involved in movement, two of them (Chain Reduction and Merge) can be taken to be independent operations. In addition, we still have the problem

---

in terms of the condition in (ii) and the operation in (iii) (see Nunes 1995, 1996, 1999, for detailed discussion of the effects of (ii) and (iii)).

   (ii) *Feature Uniformity Condition*
      Given a chain CH = (α_1, . . ., α_n), every α_i ($1 < i < n$) must have the same set of features visible at LF.

   (iii) *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.

[19] Another logical possibility to circumvent the linearization problem would be to somehow make the copies distinct, so that they would not induce violations of the conditions on linear order. On this possibility, see Hornstein 2000, where it is proposed that (resumptive) pronouns are copies disguised to comply with the linearization requirements.

[20] Given that the choice of the link to be phonetically realized is determined by economy, we predict that if the phonetic realization of the head of the chain violates other well-formedness conditions of the phonological component, a trace rather than the head of the chain will be the optimal link for phonetic realization. This is precisely what has been argued to be the case in apparently exceptional instances of clitic placement (see Franks 1998) and *wh*-in-situ (see Bošković 1999, 2000) in Serbo-Croatian. See Nunes 2000 for other cases and further discussion.

that the notions of movement and chain formation appear to be redundant in capturing the same types of relations (see Brody 1995).

These conceptual problems of course disappear if the operations Copy, Merge, Form Chain, and Chain Reduction do not necessarily cluster together as suboperations of Move, but apply independently of one another, as in the Copy + Merge theory of movement outlined in the introduction (see Nunes 1995). Thus, Merge is always taken to be a full operation of the computational system, and there may arise some instances in which the operations Copy and Merge are dissociated from Form Chain and Chain Reduction. In particular, when sideward movement takes place, as sketched in (21) (= (1)), no chain can be formed between the two copies of α if we make the standard assumption that the links of a chain must be in a c-command relation.[21]

(21) a. $[_K \ldots \alpha^i \ldots]$ α^i $\xleftarrow{\textbf{Merge}}$ $[_L \ldots]$
       $\lfloor$ __Copy__ $\rfloor$

    b. $[_K \ldots \alpha^i \ldots]$ $[_M \alpha^i [_L \ldots]]$
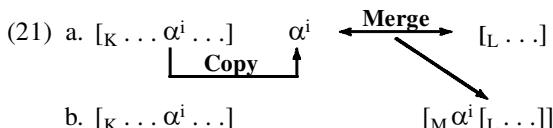
If evidence for postulating sideward movement can be provided and if sideward movement is constrained enough so that it does not overgenerate, we then have reasonable grounds to abandon the standard view of movement as a complex singulary operation. This is the approach I will pursue in the following sections, relying on the analysis of linearization of chains reviewed in section 2.2 to rule out unwanted instances of sideward movement.

3 Cyclicity

In the following subsections, I will first briefly review the exceptional instances of noncyclic movement in Chomsky's (1993, 1995) system and then show that all of these exceptional cases are actually amenable to a cyclic analysis under the Copy + Merge theory of movement. Such gain in conceptual elegance will later be combined with gains in empirical coverage to be discussed in sections 4 and 5.

3.1 Noncyclic Movement in Chomsky 1993, 1995

Chomsky (1993:22) introduces cyclicity in the Minimalist Program by proposing that generalized transformations should be subject to an ''extension requirement,'' according to which if a given syntactic object K is targeted by either Merge or Move, the resulting syntactic object should include K as a proper part. Head movement, certain instances of relative clause adjunction, and substitution movement in the covert component are however exceptional in this system in that they do not target a root syntactic object.

Such an approach is maintained in its essentials with the development of bare phrase structure (see Chomsky 1994). Chomsky (1995:248) argues that a noncyclic application of Merge embed-

---

[21] Actually, the constraints that are taken to hold of Move such as the C-Command Condition, Last Resort, and the Minimal Link Condition (see Chomsky 1995) should be understood in the context of the Copy + Merge theory of movement as constraints that govern chain formation between nondistinct copies.

ding α within some structure β already formed would introduce a serious complication in the system and would require strong empirical motivation; in the absence of such motivation, Chomsky assumes that ''Merge always applies in the simplest possible form: at the root'' (Chomsky 1995:248). As for movement, he takes cyclicity to be a property of strong features (see Chomsky 1995:233); hence, only overt movement must proceed cyclically. The only relevant change from the system developed in Chomsky 1993 is that covert movement by substitution is reanalyzed as adjunction of a set of formal features to a given head (see Chomsky 1995:sec. 4.4.4). With this revision, the distinction between cyclic and noncyclic movement correlates with the difference between movement by substitution and movement by adjunction, the latter not being subject to the extension requirement.

This state of affairs is not satisfactory, however. Move should be subject to the same condition of conceptual simplicity applied to Merge; that is, noncyclic movement should require strong empirical support. Under the Copy + Merge theory of movement, this is even more so, given that Move is understood not as an operation of the computational system, but as a reflex of the interactions among the operations Copy, Merge, Form Chain, and Chain Reduction. Let us then examine each of the exceptional cases of noncyclic movement mentioned above and see how they can receive a cyclic analysis under the Copy + Merge theory.

3.2 Eliminating Noncyclic Movement from the Grammar

3.2.1 Relative Clause Adjunction   The sentences in (22), discussed in Chomsky 1993:36, illustrate the well-known fact that noun complement clauses and relative clauses do not pattern alike as far as reconstruction effects are concerned (see, e.g., Freidin 1986, Lebeaux 1988).

(22) a. *Which claim that John$_i$ was asleep was he$_i$ willing to discuss?
     b. Which claim that John$_i$ made was he$_i$ willing to discuss?

Building on work by Lebeaux (1988), Chomsky (1993) analyzes contrasts such as the one in (22) in terms of the distinction between complements and adjuncts: ''The extension property for substitution entails that complements can only be introduced cyclically, hence before *wh*-extraction, while adjuncts can be introduced noncyclically, hence adjoined to the *wh*-phrase after raising to [Spec, CP]'' (Chomsky 1993:37).[22] According to this reasoning, the only possible (convergent) derivation for (22a) is the cyclic one represented in (23), where the complement clause merges with *claim,* before the whole *wh*-phrase is copied and moves to [Spec, CP]; (22b), on the other hand, may have either the cyclic derivation represented in (24a) or the noncyclic derivation represented in (24b), where the relative clause is adjoined after *which claim* moves to [Spec, CP]. Although both structures in (23) and (24a) arguably violate Principle C of binding

---

[22] As pointed out by a reviewer, although the reconstruction asymmetry between adjuncts and complements in Lebeaux's (1988) original proposal was independently couched in terms of selection, in Chomsky's (1993) system the correlation between being a complement or an adjunct and the need to satisfy the extension requirement is stipulated.

theory, the same is not true of (24b), because *John* is not c-commanded by *he;* hence, the additional derivation in (24b) is the source for the acceptability of (22b), in contrast with (22a).[23]

(23) [[which claim that John was asleep] was he willing to discuss [which claim that John was asleep]]

(24) a. [[[which claim] [Op$_j$ that John made t$_j$]] was he willing to discuss [[which claim] [Op$_j$ that John made t$_j$]]]
   b. [[[which claim] [Op$_j$ that John made t$_j$]] was he willing to discuss [which claim]]

Within the Copy + Merge theory of movement, the contrast between (22a) and (22b) is captured straightforwardly, with no need to resort to noncyclic movement. At some point in the (convergent) derivation of (22b) with the relevant interpretation, we have the two unconnected phrase structures in (25), which have been independently assembled. The phrase *which claim* is then copied, but instead of merging with K in (25a), it adjoins to the relative clause in (25b) (an instance of sideward movement), as illustrated in (26); finally, K and M in (26) merge, yielding the structure in (27) and allowing the interrogative complementizer Q to check its strong *wh*-feature.[24]

(25) a. K = [$_{CP_1}$ was + Q [he willing to discuss [which claim]^k]]
   b. L = [$_{CP_2}$ Op$_j$ that John made t$_j$]

(26) a. K = [$_{CP_1}$ was + Q [he willing to discuss [which claim]^k]]
   b. M = [$_{CP_2}$[which claim]^k [$_{CP_2}$ Op$_j$ that John made t$_j$]]

(27) [$_{CP_1}$[$_{CP_2}$[which claim]^k [$_{CP_2}$ Op$_j$ that John made t$_j$]] [$_{C'}$ was + Q [he willing to discuss [which claim]^k]]]

In (27), the upper copy of *which claim* participates in a checking relation, satisfying Last Resort, and c-commands the lower copy (the upper copy is contained but not dominated by CP$_2$); the Minimal Link Condition is also satisfied because there is no intervening element between the two copies that could participate in the same checking relation as the upper *wh*-copy. Once all conditions on Form Chain are satisfied (see footnote 21), the *wh*-copies form the chain CH = ([which claim]^k, [which claim]^k). After the structure in (27) is shipped to the phonological component by Spell-Out, CH must undergo Chain Reduction; otherwise, the whole structure cannot be

---

[23] However, semantic restrictions may rule out structures resulting from noncyclic adjunction. As discussed by Heycock (1995:558), the ''nonreferential'' *wh*-phrase in (ia) requires reconstruction, as opposed to the ''referential'' *wh*-phrase in (ib). Since my purpose here is only to provide a cyclic analysis for the cases where reconstruction is not required, I will not discuss cases such as (ia), which arguably involve a derivation along the lines of (24a).

(i) a. *[how many lies aimed at exonerating Clifford$_i$]$_j$ is he$_i$ planning to come up with t$_j$
   b. [which lies aimed at exonerating Clifford$_i$]$_j$ did he$_i$ expect t$_j$ to be effective

[24] In Chomsky's (1993) system, an element adjoined to a maximal projection HP or to the specifier of HP is in the checking domain of the head of HP (see Chomsky 1993:11–19 for definitions and discussion). In the case of (27), the higher *wh*-phrase is adjoined to the relative clause, which is itself in the specifier of the matrix clause; thus, the *wh*-phrase is in the checking domain of both the head of the relative clause and the head of the matrix clause. See Kato and Nunes 1998 for further discussion.

linearized (see section 2.2). The optimal reduction of CH is the one that deletes its lower link, thereby minimizing the number of applications of FF-Elimination (see section 2.2); hence, (27) ends up surfacing as (22b).

For presentation purposes, I have tacitly assumed a null operator approach to relative clauses in the above discussion. However, the cyclic analysis of (22b) within the Copy + Merge theory of movement is also compatible with a raising analysis of relative clauses (see, e.g., Vergnaud 1974, Kayne 1994). A possible implementation of a sideward movement account of (22b) under a raising analysis could proceed as follows.[25] After the ''relative'' CP in (28a) has been formed through raising and adjunction of *which claim,* the computational system selects the verb *discuss* from the numeration. If there is no convergent derivation that merges L in (28b) with the remaining lexical items of the numeration, the computational system makes a copy of *which claim* and merges it with L (an instance of sideward movement) to satisfy the selectional/thematic properties of *discuss* (see Chomsky 2000). After further computations, we reach the derivational step sketched in (29); K and M then merge, forming the structure in (30), which for the purposes of binding theory is identical to (27) in that it allows coreference between *John* and *he.*

(28) a. K = $[_{CP_1}$[which claim]^k $[_{CP_1}$ that John made [which claim]^k]]
      b. L = [discuss]

(29) a. K = $[_{CP_1}$[which claim]^k $[_{CP_1}$ that John made [which claim]^k]]
      b. M = $[_{CP_2}$ was + Q [he willing to discuss [which claim]^k]]

(30) $[_{CP_2}[_{CP_1}$[which claim]^k $[_{CP_1}$ that John made [which claim]^k]] $[_{C'}$ was + Q [he willing to discuss [which claim]^k]]]

The topmost copy of *which claim* in (30) may form a different chain with each of the lower copies: it participates in a checking relation (see footnote 24), it c-commands the lower copies, and there is no intervening element that could check the strong *wh*-feature of Q; crucially, given that neither of the lower copies c-commands the other, one is not closer to the highest copy than the other is (see Chomsky 1995:356). Finally, assuming that the θ-Criterion is to be interpreted within the minimalist framework as a condition on θ-role assignment (see Chomsky 1993:n. 23), the highest *wh*-phrase in (30) is allowed to be part of both chains, regardless of the θ-role of each chain (see Brody 1995:86).

Once these two *wh*-chains are formed before Spell-Out, each of them must undergo Chain Reduction in the phonological component in order for the structure in (30) to be linearized (see section 2.2). The output of the optimal reduction of the chain involving the object of *made* is given in (31a), and the one involving the object of *discuss* is given in (31b).[26] Further computations of the phonological component applying to (31b) finally yield the PF output associated with the sentence (22b).

---

[25] Since an adequate discussion of the correct analysis of relative clauses is beyond the scope of this article, I leave for another occasion the choice between the two traditional analyses and a full exploration of the raising analysis to be suggested below.

[26] For the purposes of the present discussion, the order in which chains are reduced is irrelevant.

(31) a. [$_{CP_1}$[$_{CP_2}$[which claim]^k [$_{CP_2}$ that John made ~~[which claim]~~*]] [$_{C'}$ was + Q [he willing to discuss [which claim]^k]]]

   b. [$_{CP_1}$[$_{CP_2}$[which claim]^k [$_{CP_2}$ that John made ~~[which claim]~~*]] [$_{C'}$ was + Q [he willing to discuss ~~[which claim]~~*]]]

Given the successful cyclic derivation of relative clause adjunction in cases such as (22b), we have to make sure that a sideward movement analysis does not extend to noun complement clauses such as (22a); otherwise, we would incorrectly allow *he* and *John* to be coreferential. Under the proposal explored here that every instance of ''movement'' (actually Merge in the Copy + Merge theory of movement) is cyclic, the derivational steps sketched in (32) and (33) are excluded by definition: in order for the CP in (32b) to become the complement of the copy of *claim,* as shown in (33b), a noncyclic merger between L and M is required.

(32) a. K = [was + Q [he willing to discuss [which^i claim^k]]]
   b. L = [that John was asleep]
   c. M = [which^i claim^k]

(33) a. K = [was + Q [he willing to discuss [which^i claim^k]]]
   b. N = [which^i [claim^k [$_L$ that John was asleep]]]

Suppose, on the other hand, that this problem can be circumvented by having the derivation proceed cyclically, along the lines of (34)–(37). That is, *claim* is copied from K in (34a) and merges with L in (34b), forming N in (35b), which then merges with a copy of *which,* forming P in (36b); finally, K and P in (36) merge, yielding the structure in (37).

(34) a. K = [was + Q [he willing to discuss [which^i claim^k]]]
   b. L = [that John was asleep]
   c. M = [claim^k]

(35) a. K = [was + Q [he willing to discuss [which^i claim^k]]]
   b. N = [claim^k [$_L$ that John was asleep]]
   c. O = [which^i]

(36) a. K = [was + Q [he willing to discuss [which^i claim^k]]]
   b. P = [which^i [claim^k [that John was asleep]]]

(37) [[$_P$ which^i [claim^k [that John was asleep]]] [$_K$ was + Q [he willing to discuss [which^i claim^k]]]]

The problem with the derivation outlined in (34)–(37) is that no chain involving the two copies of *claim* or the two copies of *which* can be formed: a potential chain CH_1 = (which^i claim^k, which^i claim^k) cannot be formed because the first link is not a constituent in (37); the other potential chains CH_2 = (claim^k, claim^k) and CH_3 = (which^i, which^i) cannot be formed either, because there is no c-command between their links. The nondistinct copies of *claim* and *which* in (37) thus induce violations of the asymmetry and irreflexivity conditions on linear order (see section 2.2), but they cannot be deleted by Chain Reduction. As stated in (8), Chain Reduction

crucially operates with nontrivial chains, not with multiple occurrences of nondistinct constituents. Thus, the structure in (37) cannot be linearized, and the derivation is canceled (see footnote 10).

The only other relevant cyclic derivation for (22a) that would allow coreference between *John* and *he* is outlined in (38)–(40), where a copy of *which claim* merges with L in (38b) and the resulting syntactic object M in (39b) merges with K, forming the structure in (40).

(38) a. K = [was + Q [he willing to discuss [which claim]^i]]
      b. L = [that John was asleep]

(39) a. K = [was + Q [he willing to discuss [which claim]^i]]
      b. M = [[which claim]^i [$_L$ that John was asleep]]

(40) [[$_M$ [which claim]^i [$_L$ that John was asleep]] [$_K$ was + Q [he willing to discuss [which claim]^i]]]

In order for a chain involving the two copies of *which claim* in (40) to be formed, the upper copy must have merged with L in (39b) by adjunction; otherwise, the upper copy would not be able to c-command the lower one. If this is the case, however, L cannot be interpreted as the complement of *claim* in (40), because it does not fall within the internal domain of *claim* (see Chomsky 1993:12). The derivation involving the steps in (38)–(40), if convergent at all, should receive a deviant interpretation at the interface.

To sum up, the Copy + Merge theory of movement is able to account for the differences between relative clauses and noun complement clauses as far as reconstruction effects are concerned, without resorting to noncyclic movement. Unwanted instances of sideward movement in the derivations involving noun complement clauses were shown to be excluded by the independently motivated conditions on linearization of chains in the phonological component.[27]

3.2.2 Head Movement   As mentioned in section 3.1, head movement is exceptional in Chomsky's (1993, 1995) systems in not being subject to the extension requirement. It should be noted that

---

[27] One reviewer observes that the cyclic analysis of relative clause adjunction developed here cannot account for the possibility of coreference in a sentence such as (i), which involves adjunction within a complement clause.

   (i) Which claim that a woman that John$_i$ met was asleep was he$_i$ willing to discuss?

The amelioration effect noted by the reviewer also shows up when a noun complement clause is embedded in another noun complement clause as in (iia–b), or when the relevant nominal expression is within the specifier of a complement as in (iic) (see, e.g., Safir 1998 for relevant discussion).

   (ii) a. Whose claim that the rumors that John$_i$ acted illegally are correct was he$_i$ willing to discuss?
        b. Whose attempts to prove the accusation that John$_i$ was guilty was he$_i$ unaware of?
        c. Which picture of John$_i$'s mother did he$_i$ like?

What the data in (i)–(ii) appear to show is that the level of embedding of the relevant nominal expression is an extra factor in allowing (marginal) coreference with the pronoun. At this point, I have nothing to say on whether this amelioration effect should be captured in grammatical terms by somehow imposing lack of reconstruction in the most embedded position in cases of multiple embedding or whether it should be explained by parsing considerations. I would like to point out, however, that if it turns out that the complement/relative clause reconstruction asymmetry is illusory (see Lasnik 1998) or that a relative clause cannot modify a single chain link for independent reasons, the main point of the article remains essentially unaltered. The reasoning used here to provide a sideward movement analysis for relative but not for complement clauses will be the same employed to exclude unwanted instances of sideward movement of heads (see sections 3.2.2 and 5.2) and derive the core properties of parasitic gap constructions (see section 4).

this assumption is motivated only for theory-internal reasons, having to do with the view of Move as a complex singulary transformation. If it is assumed that Move must relate constituents of a single phrase marker, a verb can adjoin to a tense head T, for instance, only after a projection of T containing the verb has been formed.

Once the displacement property of human languages is interpreted as the interaction of the independent primitive operations Copy, Merge, Form Chain, and Chain Reduction, overt head movement may proceed in a cyclic fashion without any problems. Verb movement, for instance, can proceed as illustrated in (41)–(43) (see Bobaljik 1995a, Nunes 1995, Bobaljik and Brown 1997, Uriagereka 1998).

(41) a. $K = [_{VP} \ldots V \ldots]$
    b. $L = T$

(42) a. $K = [_{VP} \ldots V^i \ldots]$
    b. $M = [_T \circ V^i [_T \circ T]]$

(43) $[_{TP} [_T \circ V^i [_T \circ T]] [_{VP} \ldots V^i \ldots]]$

Given the VP and the tense head T in (41), the computational system makes a copy of the verb and adjoins it to T (an instance of sideward movement) to check the strong V-feature of T, as shown in (42); after K and M in (42) merge, as shown in (43), the two copies of the verb form the chain $CH = (V^i, V^i)$. In the phonological component, CH undergoes Chain Reduction in order for the structure containing (43) to be linearized, and the copy of the verb inside the VP is deleted.[28]

Again, we have to make sure that this cyclic derivation of head movement does not overgenerate. The system should be able to rule out derivations such as the one sketched in (44)–(47), for instance, where the preposition *to* undergoes sideward movement from (44) to (45) and one of the copies of the preposition is deleted in (47), yielding the unacceptable sentences in (48).

(44) $K = $ [and John flew to^i São Paulo]
    $L = $ [Rio]

(45) $K = $ [and John flew to^i São Paulo]
    $M = $ [to^i Rio]

(46) $K = $ [and John flew to^i São Paulo]
    $N = $ [Mary drove to^i Rio]

(47) $[[_N$ Mary drove to^i Rio] $[_K$ and John flew to^i São Paulo]]

---

[28] One reviewer observes that the structure resulting from cyclic adjunction in (42b) violates Chomsky's (1995:253) Uniformity Condition, because the adjoined V is computed as a maximal projection. Although the reviewer's observation is strictly accurate, the alleged problem in fact also arises for standard (noncyclic) analyses of head movement in Chomsky's (1995) system. This was noticed by Chomsky himself, who then proposed that ''at LF, wordlike elements are 'immune' to the algorithm that determines phrase structure status'' (Chomsky 1995:322). For arguments that the Uniformity Condition should indeed be dispensed with, see Nunes 1998 and Brody 1998.

(48) a. *Mary drove to Rio and John flew São Paulo.
      b. *Mary drove Rio and John flew to São Paulo.

In what concerns the applications of Copy and Merge, there is nothing wrong with the derivational steps in (44)–(47). However, problems arise for the computations in the phonological component, after the structure in (47) is spelled out. Since the two copies of the preposition cannot form a chain (they do not enter into a c-command relation), neither copy can be deleted by Chain Reduction and the sentences in (48) cannot be derived. These copies then induce violations of the asymmetry and irreflexivity conditions on linear order, preventing the linearization of (47) and canceling the derivation.

The Copy + Merge theory of movement, therefore, not only is able to provide a cyclic derivation for standard overt head movement, but also can exclude unwanted instances of ''sideward head movement,'' based on the reasoning independently required to explain why chains in general cannot have more than one link phonetically realized (see section 2.2).

3.2.3 Covert Movement   Covert movement is the last instance of noncyclic movement required in Chomsky's (1993) system. If overt movement can only be triggered by strong features (see Chomsky 1993:30, 1995:233), movement in the covert component to check weak uninterpretable features will invariably target a nonroot syntactic object. This is also true of Chomsky's (1995) system, where covert movement is reinterpreted as movement of formal features.

Recent approaches to the copy theory have however attempted to eliminate covert movement, while still maintaining a derivational model (see, e.g., Bobaljik 1995b, Groat and O'Neil 1996, Gärtner 1997). The common traits among these various analyses are the following assumptions: (a) every movement operation takes place before Spell-Out, and (b) the standard difference between overt and covert movement is expressed in terms of whether the head or the tail of the chain is phonetically realized, in accordance with some notion of optimality of copies or positions (see footnote 17). If any of these approaches is on the right track, operations traditionally analyzed as covert movement of phrases may proceed cyclically without any problems (the same is true of Kayne 1998). On the other hand, if covert movement actually involves movement of sets of formal features in the covert component, as proposed by Chomsky (1995), it is bound to proceed noncyclically.

Since my goal in this section is only to show that noncyclic movement can be eliminated from the grammar if sideward movement is assumed, I will not attempt to choose among these various derivational ways of dealing with ''covert movement.''[29] Rather, I will show that even

---

[29] But see Aoun and Nunes 1997 for evidence in favor of Chomsky's (1995) Move F and against overt categorial movement (see Bobaljik 1995b, Groat and O'Neil 1996) or expletive-argument chains (see Brody 1995), based on vehicle change effects (in the sense of Fiengo and May 1994). The gist of Aoun and Nunes's proposal is that what is copied in the second conjunct of (i), for instance, is not the whole predicate of the first conjunct, but the verb with the adjoined FF(*John*), which has moved for Case-checking purposes. Assuming that for the purposes of binding theory, FF(*John*) is equivalent to the pronoun *him*, copying *admire* + FF(*John*) to the second conjunct induces a Principle B effect in (ia), but not in (ib).

(i) a. *Mary admires John$_i$, but he$_i$ doesn't.
     b.   Mary admires John$_i$, but he$_i$ doesn't think Susan does.

if Chomsky's Move F approach proves to be the correct one, it is still amenable to a cyclic analysis. For the sake of argument, let us assume that every movement takes place overtly and that strong features trigger overt movement of categories, whereas weak features trigger overt movement of sets of formal features.[30]

    We have seen in section 3.2.2 that overt head movement can proceed cyclically under the Copy + Merge theory, which allows restricted instances of sideward movement. Given that movement of any set of formal features FF to check a feature of the head H must involve adjunction of FF to H, as argued by Chomsky (1995:271), ''covert feature movement'' can be reanalyzed as overt sideward movement of FFs. Given the derivational step in (41) in a language where T has a weak V-feature, for instance, the computational system copies the formal features of the verb and adjoins them to T (an instance of sideward movement), as shown in (49), and later merges the VP and the complex T head, as shown in (50), allowing the chain CH $= (FF(V^i),$ $FF(V^i))$ to be formed.[31]

(49) a. $K = [_{VP} \ldots V^i \ldots]$
     b. $M = [_T \circ FF(V^i) [_T \circ T]]$

(50) $[_{TP} [_T \circ FF(V^i) [_T \circ T]] [_{VP} \ldots V^i \ldots]]$

    Assuming the null hypothesis that reduction of FF chains such as CH $= (FF(V^i), FF(V^i))$ in the phonological component does not differ from reduction of category chains (see section 2.2), the same linearization considerations that ruled out sideward movement in the derivation of (44)–(47) should also rule out corresponding derivations where nondistinct copies of FFs do not form a chain and, consequently, are not subject to Chain Reduction. By relying on sideward movement, the Copy + Merge theory is therefore able to provide a cyclic analysis of Chomsky's Move F approach, paving the way for the elimination of noncyclic movement from the grammar.

3.3 Summary

There are basically three motivations for postulating noncyclic movement in Chomsky's (1993, 1995) system. The first one is theory-internal. If it is assumed that Move is a complex singulary

---

[30] A similar proposal was independently developed by Oishi (1997), who also observes that under this view, the effects of Procrastinate regarding covert versus overt movement are locally derived from feature strength.

[31] Both Chomsky's (1995:sec. 4.4.4) covert movement of FFs and its reinterpretation above in terms of overt sideward movement of FFs need to assume (a) that the set of formal features of a given lexical item is accessible to the Copy operation, and (b) that the computational system may form a chain CH $= (FF^i, FF^i)$ relating two copies of a given set of formal features. If Copy and Form Chain can only operate with terms (constituents), which are defined in (i) (from Chomsky 1995:247), a lexical item LI should then have the format LI $= \{\gamma, \{FF, SF, PF\}\}$, where FF is a set of formal features, SF is a set of semantic features, PF is a set of phonological features, and γ is a label specifying the relevant properties of LI much like the label of phrasal syntactic objects (see Chomsky 1995:243). Notice that although SF and PF also end up being terms under this view, Last Resort correctly prevents them from moving outside LI, because they do not participate in checking relations.

   (i) *Term*
       For any structure K, (a) K is a term of K; and (b) if L is a term of K, then the members of the members of L are terms of K.

operation that targets constituents of a single phrase marker, head movement can proceed only after the projections of the landing site and the target of movement are constituents of the same phrase marker; hence, head movement is bound to be noncyclic. The second motivation, also theory-internal, is related to the format of the computational system. If it is assumed that movement operations may proceed before and after Spell-Out, covert movement invariably targets a part of the whole syntactic object formed overtly. Finally, the postulation of noncyclic movement is also empirically motivated by some reconstruction effects associated with relative clauses, in contrast with noun complement clauses (but see footnote 27).

In section 3.2.3, I showed that, under the assumption that every movement operation takes place overtly, Chomsky's (1995:sec. 4.4.4) proposal regarding covert movement can be interpreted as overt adjunction of sets of formal features to the relevant heads. The lack of cyclicity of ''covert movement'' inherent to the Move F approach should then be reduced to the lack of cyclicity in overt head movement. Such a reduction allows us to dispense with noncyclic movement motivated by theory-internal reasons, for the Copy + Merge theory of movement was shown to be successful in providing a cyclic analysis for head movement in terms of sideward movement (see section 3.2.2). A cyclic analysis in terms of sideward movement for the relevant cases of relative clause adjunction (see section 3.2.1) also proved tenable. Moreover, unwanted instances of sideward movement were prevented by the same considerations that were independently required to explain why a chain in general does not have all of its links phonetically realized (see section 2.2): a syntactic object with instances of nondistinct copies cannot be linearized in accordance with the LCA.

To the extent that the Copy + Merge theory of movement makes it possible to eliminate noncyclic movement operations and constrain sideward movement without enriching the theoretical apparatus, it constitutes a substantial conceptual improvement over an approach that takes Move as a complex singulary operation, such as the one in Chomsky 1993, 1994, 1995. In sections 4 and 5 below, the Copy + Merge theory will be shown to be empirically superior as well, in that it allows a simple account of parasitic gap and across-the-board constructions.

4 Parasitic Gaps and Sideward Movement

After the first systematic studies of parasitic gap constructions such as (51) by Taraldsen (1981), Chomsky (1982), and Engdahl (1983), the literature on the subject has grown continuously (see, e.g., Kayne 1983, 1984, Contreras 1984, Aoun and Clark 1985, Haïk 1985, É. Kiss 1985, Chomsky 1986, Browning 1987, Safir 1987, Weinberg 1988, Cinque 1990, Frampton 1990, Williams 1989-1990, Lasnik and Stowell 1991, Postal 1993, 1994, Manzini 1994, Munn 1994, Brody 1995).

(51) a. Which paper did you file without reading?
     b. [[which paper]$_i$ did you file t$_i$ without reading PG$_i$]

Several different analyses have been put forward to account for constructions such as (51). Parasitic gaps have been analyzed as traces resulting from movement of null operators (see, e.g., Chomsky 1986), traces of across-the-board extraction (see, e.g., Williams 1989-1990), traces of *wh*-phrases that are not the result of movement (see, e.g., Frampton 1990), or null resumptive

pronouns (see, e.g., Cinque 1990). For reasons of space, I will not provide a comprehensive review of the parasitic gap literature in this article. I will limit myself to showing how parasitic gaps can be accounted for under the Copy + Merge theory of movement and highlighting the large conceptual differences between previous approaches and the sideward movement approach proposed here. I will specifically address the properties of parasitic gaps listed in (52).[32]

(52) a. Parasitic gaps are ''selectively'' sensitive to islands; that is, although they typically occur within CED (Condition on Extraction Domain) islands in the sense of Huang 1982, they cannot be separated from their licenser by more than one island (see, e.g., Kayne 1983, 1984, Chomsky 1986).

b. Parasitic gaps are licensed at S-Structure (see, e.g., Chomsky 1982).

c. Parasitic gaps cannot be c-commanded by the ''real'' gap (see, e.g., Taraldsen 1981, Engdahl 1983).

d. Parasitic gaps cannot be licensed by A-chains (see, e.g., Chomsky 1982).

4.1 General Approach

Let us examine step by step how the parasitic gap construction given in (51a) can be derived in the Copy + Merge theory of movement. Suppose that the computational system starts with the initial numeration N in (53a) and operates until the derivational stage where N has been reduced as N' in (53b) and the two syntactic objects K and L in (54) have been assembled.

(53) a. N = {which$_1$, paper$_1$, Q$_1$, you$_1$, did$_1$, v$_2$, file$_1$, without$_1$, C$_1$, PRO$_1$, T$_1$, reading$_1$}

b. N' = {which$_0$, paper$_0$, Q$_1$, you$_1$, did$_1$, v$_1$, file$_0$, without$_1$, C$_0$, PRO$_0$, T$_0$, reading$_0$}

(54) a. K = [$_{CP}$ C [$_{TP}$ PRO$_j$ [$_{T'}$ T [$_{vP}$ t$_j$ [$_{v'}$ v [$_{VP}$ reading [which paper]^i]]]]]]

b. L = file

Possible continuations of (54) involving merger of L with K or any object formed from the lexical items still available in N' do not converge: the θ-roles of *file* and the remaining v in N' still have to be discharged, and there is only one element in N' that could bear a θ-role, namely, *you*. The derivational step sketched in (54) may however lead to a convergent result if the computational system makes a copy of *which paper* and merges it with L (an instance of sideward movement) to satisfy the selectional/thematic properties of *file* (see Chomsky 2000), as illustrated in (55).

(55) a. K = [$_{CP}$ C [$_{TP}$ PRO$_j$ [$_{T'}$ T [$_{vP}$ t$_j$ [$_{v'}$ v [$_{VP}$ reading [which paper]^i]]]]]]

b. M = [$_{VP}$ file [which paper]^i]

---

[32] Other properties not directly relevant to sideward movement are given in (i). The fact that the restrictions in (i) are also found in constructions involving *wh*-phrases in situ suggests that they should be viewed as general conditions on operator-variable constructions, rather than specific properties of parasitic gap constructions (for relevant discussion, see Kayne 1983, 1984, Brody 1995, Nunes 1995, Hornstein and Nunes 1999).

(i) a. Parasitic gaps can only be NPs (see, e.g., Aoun and Clark 1985, Chomsky 1986).

b. Parasitic gaps cannot be licensed by nonreferential NPs (see, e.g., Cinque 1990, Postal 1993).

As the derivation proceeds, other lexical items are pulled out from N′ in (53b) and merge with K and M in (55), yielding the objects P and Q in (56).

(56) a. P = [$_{PP}$ without [$_{CP}$ C [$_{TP}$ PRO$_j$ [$_{T'}$ T [$_{vP}$ t$_j$ [$_{v'}$ v [$_{VP}$ reading [which paper]^i]]]]]]]]
   b. Q = [$_{vP}$ you [$_{v'}$ v [$_{VP}$ file [which paper]^i]]]

(57) represents the next derivational step, where P adjoins to Q. In (57), no chain formation between the nondistinct copies of *which paper* can take place, since they are not in a c-command relation.

(57)   vP

[$_{vP}$ you [$_{v'}$ v [$_{VP}$ file   [$_{PP}$ without [$_{CP}$ C [$_{TP}$ PRO$_j$
[which paper]^i]]]   [$_{T'}$ T [$_{vP}$ t$_j$ [$_{v'}$ v [$_{vP}$ reading
                            [which paper]^i]]]]]]]]
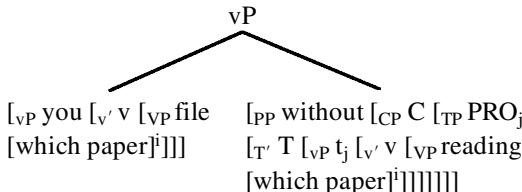
After the remaining lexical items of N′ in (53b) are pulled out and merge with the structure in (57) and after *you* and *did* move, the structure in (58) is derived (copies irrelevant for the discussion are omitted). Under the assumption that the interrogative complementizer Q in (58) has a strong *wh*-feature, another copy of *which paper* is made and merges with the structure in (58), forming (59).

(58) [$_{CP}$ did + Q [$_{TP}$ you [$_{vP}$[$_{vP}$ file [which paper]^i] [$_{PP}$ without PRO reading [which paper]^i]]]]

(59) [$_{CP}$ [which paper]^i did + Q [$_{TP}$ you [$_{vP}$[$_{vP}$ file [which paper]^i] [$_{PP}$ without PRO reading [which paper]^i]]]]

In (59), the copy of *which paper* in [Spec, CP] can form a chain with either of the lower copies (see, e.g., Frampton 1990, Brody 1995), in compliance with the Minimal Link Condition: given that neither of the lower copies in (59) c-commands the other, one is not closer to the copy in [Spec, CP] than the other is. If neither chain is formed, Chain Reduction will be inapplicable and the copies of *which paper* will induce violations of the irreflexivity and asymmetry conditions on linear order (see section 2.2), preventing (59) from being linearized. Once no linear order is established, no PF object is formed and the derivation is canceled.

Similar considerations apply to the situation in which only one chain is formed: the link that survives Chain Reduction and the copy that has not been part of a chain prevent the whole structure from being linearized. Thus, the structure in (59) can only yield a PF object if the copy of *which paper* in [Spec, CP] actually forms a chain CH$_1$ with the object of *reading* and a chain CH$_2$ with the object of *file* (see the discussion of (30)). If this happens, the optimal reduction of CH$_1$ and CH$_2$ in the phonological component involves the deletion of their lower link, as respectively shown in (60a) and (60b) (see section 2.2). After (60b) is linearized, further computations in the phonological component finally yield the PF output corresponding to the sentence (51a).

(60) a. [$_{CP}$[which paper]^i did + Q [$_{TP}$ you [$_{vP}$[$_{vP}$ file [which paper]^i] [$_{PP}$ without PRO reading ~~[which paper]~~*]]]]]

b. [$_{CP}$[which paper]^i did + Q [$_{TP}$ you [$_{vP}$[$_{vP}$ file ~~[which paper]~~*] [$_{PP}$ without PRO reading ~~[which paper]~~*]]]]]

The Copy + Merge theory of movement, which allows constrained instances of sideward movement, 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. In fact, the reasoning used here was the same one used in the raising analysis of relative clauses suggested in section 3.2.1. A chain formed before Spell-Out that 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 copy before Spell-Out, it will not be deleted by Chain Reduction and will induce violations of the irreflexivity and asymmetry conditions on linear order, preventing the structure from being linearized and canceling the derivation.

Let us now examine how this analysis of parasitic gaps in terms of sideward movement and linearization of chains is able to account for the properties listed in (52).

4.2 Parasitic Gaps and Islandhood: An Argument for Derivations

Perhaps the most interesting questions posed by parasitic gap constructions are related to island effects (especially CED effects, since other island effects may be captured by the Minimal Link Condition or Subjacency). If the copy inside the adjunct clause can form a chain with the copy in the matrix [Spec, CP] in (59), for instance, why does regular extraction out of an adjunct island, as illustrated in (61), yield an unacceptable sentence? Besides, why do island effects show up if the parasitic gap is further embedded in another (CED) island, as illustrated in (62) (see, e.g., Kayne 1983, 1984, Chomsky 1986)?

(61) a. *Which book did you review this paper without reading?

b. [$_{CP}$[which book]^i did + Q [$_{TP}$ you [$_{vP}$[$_{vP}$ review this paper] [$_{PP}$ without reading [which book]^i]]]]]

(62) a. *Which book did you borrow after leaving the bookstore without finding?

b. [$_{CP}$[which book]^i did + Q [$_{TP}$ you [$_{vP}$[$_{vP}$ borrow [which book]^i] [$_{PP}$ after [$_{CP}$ PRO [$_{vP}$[$_{vP}$ leaving the bookstore] [$_{PP}$ without PRO finding [which book]^i]]]]]]]]

From a representational point of view, there seems to be no way to capture the distinction between the *wh*-chain involving the copy inside the adjunct in (59) and the one in (61b), without additional provisos regarding the ''parasitic'' chain in (59) (see Brody 1995, for a representational approach to parasitic chains). By contrast, a derivational approach can provide a uniform account of (59), on the one hand, and (61) and (62), on the other, by exploring their different derivational

histories. Let us examine these differences by first comparing the movement steps that were involved in the derivation of (59) and (61b).

Assuming that Merge only proceeds cyclically (see section 3), the derivational stage immediately preceding the *wh*-movement in (61b) is represented in (63).

(63) $[_{CP}$ did + Q $[_{TP}$ you $[_{vP}[_{vP}$ review this paper] $[_{PP}$ without reading [which book]^i]]]]

As for the derivation of (59), the first movement of *which paper* (an instance of sideward movement) took place in the derivational step represented in (55), repeated in (64), when the gerundive clause was still an independent syntactic object and not an adjunct. After the structure in (58), repeated in (65), was assembled, another copy of *which paper* had to be created in order to check the strong *wh*-feature of Q. Regardless of how the adjunct islands are to be captured in minimalist terms, it is reasonable to take the copy of *which paper* inside the adjunct clause of (65) to behave like the *wh*-phrase inside the adjunct in (63) in being unavailable for the computational system. The crucial difference between these structures is that in (65), the *wh*-copy in the object of *file* is clearly not inside any island and could undergo standard *wh*-movement without any problems. To put it differently, the derivation of (59) does not exhibit a CED effect because *which paper* moved prior to the derivational step where the gerundive clause became an island; therefore, it was free to undergo additional movement, regardless of the status of the gerundive clause in later steps of the derivation.

(64) a. K = $[_{CP}$ C $[_{TP}$ PRO$_j$ $[_{T'}$ T $[_{vP}$ t$_j$ $[_{v'}$ v $[_{VP}$ reading [which paper]^i]]]]]]
    b. $[_{VP}$ file [which paper]^i]

(65) $[_{CP}$ did + Q $[_{TP}$ you $[_{vP}[_{vP}$ file [which paper]^i] $[_{PP}$ without PRO reading [which paper]^i]]]]

As for (62b), the relevant derivational step required to yield the parasitic gap construction in (62a) is represented in (66). At the step where the selectional/thematic requirements of *borrow* would license sideward movement of *which book* from K to L, the PP has already become an (adjunct) island, and the *wh*-phrase, like the one in (63), is unavailable for the computational system. Thus, from this derivational perspective, the structures in (61b) and (62b) simply cannot be generated.[33]

(66) a. K = $[_{CP}$ PRO $[_{vP}[_{vP}$ leaving the bookstore] $[_{PP}$ without PRO finding [which book]^i]]]
    b. L = borrow

---

[33] The approach sketched above still requires an explanation for why syntactic objects within CED islands cannot be copied, which is of course tied to the thorny issue of how CED islands are to be accounted for in the Minimalist Program. Because of space limitations, I will not expand on this issue here. For concreteness, I will assume Uriagereka's (1999) Multiple Spell-Out approach as implemented in Nunes and Uriagereka 2000, according to which complex adjuncts and complex specifiers must be spelled out before they are merged in the relevant structure; hence, the constituents of spelled-out subjects and adjuncts are simply inaccessible to the computational system and cannot be copied (see also Hornstein 2000 for a similar approach and relevant discussion).

The reader may have noted that a crucial aspect of this proposal is that (sideward) movement always proceeds from a subordinated to a subordinating domain, which is exactly what we should expect in a strongly cyclic derivational system.[34] If the computational system could first start building the matrix derivational workspace before building an embedded derivational workspace, the sentence in (62a), for instance, would be incorrectly ruled in by a derivation where sideward movement proceeded from the object of *borrow* to the object of *finding*. This directionality of movement from more to less embedding domains is motivated not only by theory-internal reasons, but also by empirical considerations. Take the contrast between the sentences in (67), for instance (based on Postal 1993). The relevant structural difference that seems to underlie the contrast is that in (67a) *read* is in the most embedded domain and *give* is in the subordinating domain, but in (67b) their positions are reversed. Under the analysis developed here, the parasitic gap construc-tions of (67) should involve the (simplified) derivational steps in (68) and (69), respectively, followed by sideward movement of *which*.

(67) a.   This is the book which I was given by Ted after reading.
     b. *This is the book which I read before being given by Ted.

(68) a. K $=$ [$_{CP}$ PRO reading which]
     b. L $=$ given

(69) a. P $=$ [$_{CP}$ PRO being given which by Ted]
     b. Q $=$ read

Assuming with Lasnik (1995) and Chomsky (2000) that what makes an element visible for A-movement is its unchecked *structural* Case, Hornstein and Nunes (1999) argue that sideward movement in parasitic gap constructions must be launched from a position associated with struc-tural, rather than inherent, Case. Contrasts such as the one in (67) are then accounted for, under the assumption that the theme of *give* in double object constructions is associated with inherent Case (see Larson 1988 for relevant discussion); hence, *which* is visible for copying in (68), but is inert in (69). Glossing over the details of Hornstein and Nunes's analysis, what is relevant for our current discussion is that if *which* could freely move from the object position of *read* to the object position of *give* in the derivation of (67b), we would incorrectly predict that it should be as acceptable as (67a); however, if movement must proceed from more to less embedded domains in a cyclic fashion, this scenario does not arise and (67b) is correctly excluded, because *which* is inert for A-movement in (69), the derivational step that would be required in order for (67b) to be generated.

To sum up, the derivational approach explored here accounts for the lack of CED island effects in acceptable parasitic gap constructions and for asymmetries between the ''real'' and the ''parasitic'' gap by relying on the accessibility of the relevant syntactic objects at specific deriva-

---

[34] See Nunes and Uriagereka 2000, where this directionality in the structure-building process is captured in terms of access to subarrays of the numeration (see Chomsky 2000).

tional steps; thus, there is no need to postulate special licensing requirements on parasitic gaps.[35] To the extent that it succeeds, this analysis provides a compelling argument for derivations them- selves, for it is not immediately obvious how a representational alternative can fare as well, without special provisos.

4.3 ''S-Structure Effects''

In pre-minimalist analyses of parasitic gaps, the contrast between (70a) and (70b), for instance, was taken to show that a parasitic gap must be licensed at S-Structure (see, e.g., Chomsky 1982); otherwise, the parasitic gap in (70b) could be licensed at LF after *which report* moved in the covert component. Given that the only syntactic levels of representation within the minimalist framework are the interface levels LF and PF, the contrast in (70) presents a challenge for any analysis of parasitic gaps within the Minimalist Program. I will discuss two potential scenarios in which the contrast in (70) should be explained.

(70) a.   Which report did you file without reading?
     b.   *Who filed which report without reading?

Following standard analyses of (70), let us first suppose for the sake of the argument that in the covert component, the *wh*-in-situ in (70b) adjoins to the *wh*-phrase in [Spec, CP], from which position it should be able to form a chain with the parasitic gap. Such covert movement in no way undermines the analysis of parasitic gaps presented in section 4.1. Under the relevant reading, (70b) is to be derived by sideward movement of *which report* from the object of *reading* to the object of *file* (see section 4.2), and the (simplified) structure in (71) is the one shipped to the phonological component by Spell-Out.

(71) [who [[filed [which report]^i] [without PRO reading [which report]^i]]]

The pair ([which report]^i, [which report]^i) in (71) cannot form a chain, because it does not satisfy the c-command condition on Form Chain. Recall that Chain Reduction deals with nontrivial chains and not with multiple occurrences of nondistinct constituents (see footnote 10). Thus, after (71) 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. Since neither of the *wh*-copies in (71) is deleted in the phonological component, they induce violations of the irreflexivity and asymmetry

---

[35] Chomsky (1986), citing Kearney 1983, takes the absence of reconstruction in the parasitic gap position in construc- tions such as (i) as evidence for the parasitic nature of the empty category in the object of *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 and that other factors are responsible for the pattern exhibited below (see Munn 1994 for discussion).

(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?

conditions on linear order, making it impossible for (70) to be linearized and canceling the deriva-tion.[36] Sideward movement is therefore constrained enough so that it does not incorrectly rule (70b) in.

Let us now assume, in consonance with the discussion in section 3.2.3, that every movement takes place overtly. The derivation that concerns us here is the one in which the two copies (or their sets of formal features) overtly move to the position where they can check their Case. For the sake of discussion, let us examine the situation where the position to which the object of *filed* moves c-commands the other copies in the adjunct clause, as represented in (72).

(72) [who [[which report]^1-$_{CASE}$ [$_{vP}$[$_{vP}$ filed [which report]^2-**CASE**] [$_{PP}$ without PRO [[which report]^3-$_{CASE}$ reading [which report]^4-**CASE**]]]]]

In (72), the only chains that can be formed are $CH_1 = $ (copy^1, copy^2) and $CH_2 = $ (copy^3, copy^4) (I refer to each individual copy by its superscripted number). Crucially, copy^1 cannot form a chain with copy^3 even if the former c-commands the latter; Last Resort prevents multiple Case checking in a single chain, and the intervention of PRO induces a Minimal Link Condition violation. After the structure in (72) is shipped to the phonological component and both CH_1 and CH_2 are reduced, as represented in (73), the links that survive Chain Reduction induce violations of asymmetry and irreflexivity, preventing the structure from being linearized and canceling the derivation. The same reasoning holds if the overt object movement in (72) actually involves overt movement of sets of formal features, as discussed in section 3.2.3. Again, the sentence in (70b) cannot be generated through sideward movement.

(73) [who [[which report]^1-$_{CASE}$ [$_{vP}$[$_{vP}$ filed ~~[which report]^2-**CASE**~~ [$_{PP}$ without PRO [[which report]^3-$_{CASE}$ reading ~~[which report]^4-**CASE**~~]]]]]

While the analyses of the contrast in (70) in terms of resumptive pronouns or traces of null operators, for instance, must rely on a construction-specific licensing at a noninterface level (see (52b)), the analysis developed here accounts for (70) on the basis of general considerations regard-ing linearization in the phonological component. To the extent that the analysis of parasitic gaps in terms of sideward movement and linearization of chains sketched in section 4.1 derives ''S-Structure effects'' without postulating a noninterface 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 constrain sideward movement of heads (see section 3.2.2), for instance, derives the ''S-Structure effects'' on parasitic gap licensing.

4.4 Structural Requirements

The licensing of parasitic gaps is also dependent on the structural position of the ''real'' gap (see Taraldsen 1981, Engdahl 1983). As shown in (74), a parasitic gap cannot be c-commanded by

---

[36] Obviously, this explanation remains intact if Last Resort actually prevents covert movement of the *wh*-in-situ in (71) to [Spec, CP], as argued by Chomsky (1995:291).

the licensing gap. (75) further shows that the unacceptability of (74) is not due to some Case incompatibility between the two gaps (see Barss 1986).[37]

(74) a. *I wonder which man called you before you met.
   b. *[I wonder [$_{CP}$[which man]$_i$ [$_{TP}$ t$_i$ [$_{vP}$[$_{vP}$ t$_i$ called you] [$_{PP}$ before you met PG$_i$]]]]]

(75) a. I wonder which papers John said were unavailable before reading.
   b. [I wonder [$_{CP}$[which papers]$_i$ [$_{TP}$ John [$_{vP}$[$_{vP}$ said t$_i$ were t$_i$ unavailable] [$_{PP}$ before reading PG$_i$]]]]]

In this section, I show that the anti-c-command restriction on the distribution of parasitic gaps also follows from the interaction between chain formation and linearization in the phonological component. Let us start by examining the derivation of (74a) under the assumption that the temporal PP is adjoined to the light vP shell.[38] Given the simplified derivational step in (76), the computational system may copy *which man* from K and merge it with L, if there is no other element in the numeration that could receive the external θ-role of the light verb in L. In such a case, further computations would then yield the structure in (77), where two other copies of *which man* are made to check the EPP-feature of T and the strong *wh*-feature of the embedded interrogative complementizer Q.[39]

(76) a. K = [$_{CP}$ you met [which man]^i]
   b. L = [$_{vP}$ v [$_{VP}$ called you]]

---

(77)

```
              vP
            /    \
          I       v′
                /    \
           wonder     CP
                     /    \
            [which          TP
             man]⁴        /    \
                  [which        T′
                   man]³      /    \
                         T        vP
                               /       \
                  [vP [which man]² v    [PP before you met
                   [vP called you]]      [which man]¹]
```

Let us examine which copies of *which man* in (77) can form chains. As in regular *wh*-movement of a subject, copy² can form a chain with copy³, which in turn can form a chain with copy⁴; assuming that each of these chains is formed, we obtain the linked chain (in the sense of Chomsky and Lasnik 1993:563) CH = (copy⁴, copy³, copy²). As for copy¹, it cannot form a chain with copy² because the c-command condition is not satisfied. Furthermore, since copy³ checks Case and there are some nonlocal Case-checking positions and Case-bearing elements between copy³ and copy¹, the Minimal Link Condition prevents these two copies from forming a chain. Finally, the Minimal Link Condition also prevents copy¹ from forming a chain with copy⁴, which checks the strong *wh*-feature of Q: copy³ has a *wh*-feature and is closer to copy⁴ than copy¹ is.

Since the only nontrivial *wh*-chain formed in (77) is CH = (copy⁴, copy³, copy²), its optimal reduction in the phonological component will involve the deletion of copy³ and copy², as shown in (78). The two surviving nondistinct copies of *which man* then induce violations of the irreflexivity and asymmetry conditions on linear order, preventing the structure in (78) from being linearized and canceling the derivation. Therefore, there is no convergent source for the sentence in (74a) under a sideward movement approach to parasitic gap constructions.[40]

---

[40] For the sake of completeness, suppose we try to circumvent the Minimal Link Condition effect preventing the pair (copy³, copy¹) from forming a chain in (77), by placing copy¹ in the subject position inside the adjunct clause, so that there would be no Case-checking position intervening between the two copies. This is schematically illustrated in (i) and (ii), with a finite and nonfinite adjunct clause, respectively (the irrelevant *wh*-phrases in [Spec, vP] are omitted).

(i) a. *I wonder which man called you before met you.
    b. *[I wonder [CP[which man]ⁱ Q [TP[which man]ⁱ T [vP[vP called you] [PP before [TP[which man]ⁱ met you]]]]]]

(78) [I wonder [$_{CP}$[which man]^4 Q [$_{TP}$~~[which man]^3~~ T [$_{vP}$[$_{vP}$~~[which man]^2~~ [$_{v'}$ v [$_{VP}$ called you]]] [$_{PP}$ before you met [which man]^1]]]]]

Let us now consider the relevant structure of the acceptable sentence in (75a) given in (79), which is derived through sideward movement of *which papers* from within the PP to the thematic position associated with *unavailable* (see section 4.2) and adjunction of the PP to the vP related to *said*. Copy^5 in (79) is the only copy that c-commands copy^1; since there is no intervening element between these two copies that could check the strong *wh*-feature of Q, the chain CH$_1$ = (copy^5, copy^1) is formed. The computational system also forms the linked chain CH$_2$ = (copy^5, copy^4, copy^3, copy^2). Optimal reduction of CH$_1$ in the phonological component deletes copy^1, yielding the structure in (80), which is then converted to (81) after CH$_2$ is reduced and its three lower links are deleted. After linearization and remaining operations of the phonological component, the PF output associated with (75a) is derived.

(79)

[$_{vP}$ said [$_{CP}$[which papers]^4 C [$_{PP}$ without reading
[$_{TP}$[which papers]^3 T   [which papers]^1]
[$_{VP}$ were unavailable
[which papers]^2]]]]
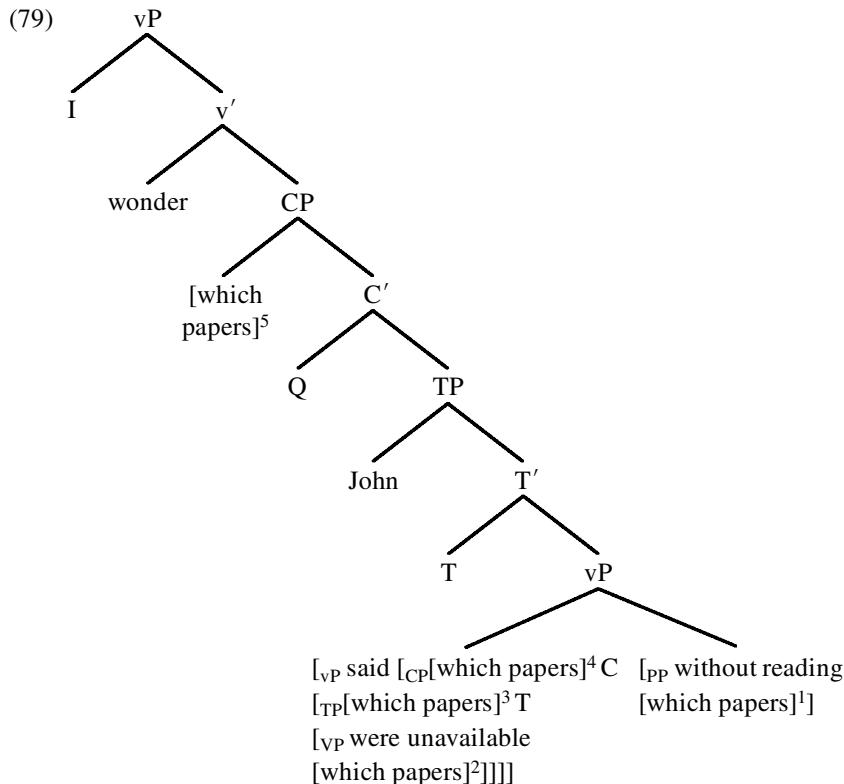
(ii) a. I wonder which man called you before meeting you.
   b. *[I wonder [$_{CP}$[which man]^i Q [$_{TP}$[which man]^i T [$_{vP}$[$_{vP}$ called you] [$_{PP}$ before [$_{TP}$[which man]^i meeting you]]]]]]
   c. [I wonder [$_{CP}$[which man]^i Q [$_{TP}$[which man]^i T [$_{vP}$[$_{vP}$ called you] [$_{PP}$ before [$_{TP}$ PRO meeting you]]]]]]
   In (ib), the Minimal Link Condition does not prevent the ''parasitic gap'' copy in the adjunct clause from forming

(80) [I wonder [$_{CP}$[which papers]^5 Q [$_{TP}$ John [$_{vP}$[$_{vP}$ said [$_{CP}$[which papers]^4 C [$_{TP}$[which papers]^3 T [$_{VP}$ were unavailable [which papers]^2]]]] [$_{PP}$ without reading ~~[which papers]^1~~]]]]]]

(81) [I wonder [$_{CP}$[which papers]^5 Q [$_{TP}$ John [$_{vP}$[$_{vP}$ said [$_{CP}$~~[which papers]^4~~ C [$_{TP}$~~[which papers]^3~~ T [$_{VP}$ were unavailable ~~[which papers]^2~~]]]] [$_{PP}$ without reading ~~[which papers]^1~~]]]]]]

The analysis presented above extends straightforwardly to sentences such as (82a) (from Chomsky and Lasnik 1993), which is taken to show that parasitic gaps cannot be licensed by traces of A-movement (see (52d)). Under the relevant reading, (82a) should be derived by sideward movement of *the book* from the object of *reading* to the object of *filed,* yielding the structure in (82b). The copy inside the adjunct clause can form a chain neither with the copy in the object of *filed,* because of lack of c-command, nor with the copy in the matrix subject position, because of the intervention of *my.* Thus, sideward movement correctly fails to provide a convergent derivation for (82a): copy^1 and the link that survives reduction of the chain CH = (copy^3, copy^2) induce violations of the asymmetry and irreflexivity requirements on linear order, preventing the structure from being linearized.

(82) a. *The book was filed without my reading first.
   b. [[the book]^3 was [$_{vP}$[$_{vP}$ filed [the book]^2] [$_{PP}$ without my reading [the book]^1 first]]]

To summarize, the conditions on chain formation (see footnote 21) prevent a ''parasitic gap'' from being licensed by a member of an A-chain that c-commands it, yielding a situation where two different chains involving nondistinct copies do not have a link in common. The links of these chains that escape Chain Reduction in the phonological component then induce violations of the irreflexivity and asymmetry conditions on linear order, canceling the derivation because no PF object is formed. The analysis of parasitic gap constructions in terms of sideward movement and linearization of chains thus derives the structural requirements on parasitic gaps (see (52c–d)) without introducing any new principle or operation. The same factors that 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 copy in an A-position.

5 Across-the-Board Extraction and Sideward Movement

5.1 Wh-*Phrases and Auxiliaries*

Let us now consider constructions involving across-the-board (ATB) extraction (see, e.g., Ross 1967, Williams 1978) such as the one illustrated in (83).

(83) a. Which paper did John file and Mary read?
      b. [[which paper]$_i$ did$_j$ John t$_j$ file t$_i$ and Mary t$_j$ read t$_i$]

   The similarities between parasitic gap and ATB constructions have led researchers to attempt to assimilate one of the two to the other. For example, Haïk (1985) and Williams (1989-1990) 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 which are the fact that analyses of parasitic gaps as ATB constructions have to assume that every structure that allows a parasitic gap is optionally coordinative, and the fact that these analyses crucially rely on the construction-specific ATB operation.

   The opposite approach is taken by Munn (1993), who proposes that ATB constructions involve movement of a null operator, as in Chomsky's (1986) analysis of parasitic gap constructions. This approach has the conceptual advantage of eliminating the ATB formalism from the grammar. However, as discussed in section 4.2, there are reasons to believe that the island effects in parasitic gap constructions can be accounted for with no need to resort to null operators. Moreover, the analysis of the ATB extraction of auxiliaries as in (83) or of main verbs (see section 5.2) does not appear to be amenable to a null operator approach. I will thus proceed under the assumption that no null operator is required in ATB extraction of *wh*-phrases and show that ATB constructions can also be analyzed in terms of sideward movement and linearization of chains. In other words, despite their apparent differences, ATB and standard movement receive a uniform treatment under the Copy + Merge theory of movement explored here.[41]

   I assume that coordination structures involve hierarchical rather than flat structures and that coordinating conjunctions head their own phrases (see Munn 1987 for an early proposal, among many others). Under this view, *and* in (83b) takes a TP as its complement and another TP as its specifier. Assuming that (84) is the initial numeration underlying (83), the relevant details of the derivation are as follows. At some given derivational step, the syntactic objects K and L in (85) 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* (an instance of sideward movement), and further applications of Select and Merge form the object M in (86).

---

[41] Postal (1993) documents several cases showing that ATB constructions are more permissive than parasitic gap constructions, which could be seen as a serious problem for a unified approach in terms of sideward movement. Hornstein and Nunes (1999) however argue that the more permissive behavior of ATB constructions is actually due to the independent Parallelism Requirement on coordinate structures. To put it in general terms, the Parallelism Requirement functions as an enabling condition that authorizes movement operations that otherwise would not be licensed by Last Resort. See Hornstein and Nunes 1999 for details and discussion.

(84) N = {which$_1$, paper$_1$, did$_1$, Q$_1$, John$_1$, v$_2$, file$_1$, and$_1$, Mary$_1$, read$_1$ }

(85) a. K = [$_{TP}$ did [$_{vP}$ Mary v [$_{VP}$ read [which paper]]]]
 b. L = file

(86) a. K = [$_{TP}$ did [$_{vP}$ Mary v [$_{VP}$ read [which paper]^i]]]
 b. M = [$_{vP}$ John v [$_{VP}$ file [which paper]^i]]

Given that the only Tense head available in the numeration (namely, *did*) has already been used, the computational system makes a copy of *did* and merges it with M (another instance of sideward movement), yielding O in (87).

(87) a. K = [$_{TP}$ did^k [$_{vP}$ Mary v [$_{VP}$ read [which paper]^i]]]
 b. O = [$_{TP}$ did^k [$_{vP}$ John v [$_{VP}$ file [which paper]^i]]]

The subjects of both K and O in (87) move to their respective [Spec, TP] to check the strong feature of each instance of *did,* and the coordinating head *and* merges with one of the resulting structures, forming the (simplified) objects P and R in (88). P and R then merge, and the resulting structure in turn merges with the interrogative complementizer Q, exhausting the numeration and yielding the structure in (89).

(88) a. P = [$_{andP}$ and [$_{TP}$ Mary did^k read [which paper]^i]]
 b. R = [$_{TP}$ John did^k file [which paper]^i]

(89) [$_{CP}$ Q [$_{andP}$[$_{TP}$ John did^k file [which paper]^i] [$_{and'}$ and [$_{TP}$ Mary did^k read [which paper]^i]]]]

Since the numeration has been exhausted and the complementizer Q in (89) has strong features to be checked, the computational system makes a copy of *did* and adjoins it to Q;[42] it then makes another copy of *which paper* and merges it with the resulting structure, yielding the object in (90).

(90) [$_{CP}$[which paper]^3 did^3 +Q [$_{andP}$[$_{TP}$ John did^2 file [which paper]^2] [$_{and'}$ and [$_{TP}$ Mary did^1 read [which paper]^1]]]]

If (90) is spelled out before the *wh-* and T-chains are formed, the nondistinct copies of *which paper* and *did* will induce violations of the irreflexivity and asymmetry conditions on linear order, canceling the derivation. The computational system must therefore form the chains CH$_1$ = ([which paper]^3, [which paper]^1), CH$_2$ = ([which paper]^3, [which paper]^2), CH$_3$ = (did^3, did^1), and CH$_4$ = (did^3, did^2). Each of these chains is then reduced in the phonological component, resulting in the structure in (91), and further computations of the phonological component finally yield the PF output associated with the sentence in (83a).

(91) [$_{CP}$[which paper]^3 did^3 +Q [$_{andP}$[$_{TP}$ John ~~did^2~~ file ~~[which paper]^2~~] [$_{and'}$ and [$_{TP}$ Mary ~~did^1~~ read ~~[which paper]^1~~]]]]

---

[42] To be precise, after the interrogative complementizer Q is selected, the computational system makes a copy of *did* and adjoins it to Q (another instance of sideward movement) before merging Q with *and*P (see section 3.2.2).

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, e.g., Ross 1967, Williams 1978, Munn 1993). 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 the Copy + Merge theory of movement, which takes Move to be the description of the 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 cyclic movement (see section 3) and to dispense with the ATB formalism.

Below, I show how this analysis can be extended to ATB movement of verbs out of VP shells (see Larson 1988) with interesting empirical consequences.

5.2 Main Verbs

In section 3.2.2, we saw that the Copy + Merge theory of movement makes it possible to provide a cyclic analysis of head movement by allowing constrained instances of sideward head movement. In the admissible case discussed there, after a given verb underwent sideward movement and adjoined to T, the resulting category merged with VP, allowing the two copies of the verb to form a chain (see (41)–(43)). Crucially, if the two nondistinct copies could not form a chain, as in instances where one copy did not c-command the other, they caused the derivation to be canceled because the structure containing them could not be linearized (see (44)–(47)).

When more than two copies of a given head are considered, new possibilities arise. As in the case of analysis of across-the-board extraction of *wh*-phrases and auxiliaries discussed in section 5.1, a convergent derivation involving two copies that do not c-command one another may arise if an additional copy forms a different chain with each of the other copies, yielding across-the-board effects. Let us consider one such derivation.

In Portuguese, both sentences in (92) may have an interpretation where there is a single talking event involving three participants. At first sight, the only difference between these sentences seems to be the kind of constituents that are coordinated: PPs in (92a) and DPs in (92b). However, these sentences show an interesting contrast. (92a) also admits a reading under which there are two events with two participants each (the subject and one of the objects), but (92b) does not.[43] This is clear when adverbial expressions modify each of the events, as shown in (93).[44]

---

[43] I am thankful to Ricardo Echepare for helpful discussion about the relevant readings.

[44] At this point, I have no account for why English also allows the two-event reading for constructions such as (92b). The sentences in (ia) (from Pesetsky 1995) and (ib) show that this contrast between English and Portuguese also turns up in constructions involving a verb and a complement.

    (i) a. John gave flowers to her yesterday and (to) him on Sunday.
       b. João deu flores pra ela ontem e *(pra) ele domingo.

(92) a. Eu conversei com o João e   com a   Maria.
        I   talked   with the João and with the Maria
        'I talked to João and to Maria.'

     b. Eu conversei com o João e   a   Maria.
        I   talked   with the João and the Maria
        'I talked to João and Maria.'

(93) a. Eu conversei com o João sábado e com a Maria domingo.
        'I talked to João on Saturday and to Maria on Sunday.'

     b. *Eu conversei com o João sábado e a Maria domingo.
        'I talked to João on Saturday and Maria on Sunday.'

Let us examine the relevant steps of the derivation of (92a) under the two-event reading, assuming that the relevant initial numeration involves only a single instance of the main verb (I use English words for the sake of exposition). If the general requirement that conjuncts be of the same syntactic type is a convergence condition, there is no convergent continuation for the derivational step in (94), if K merges with L. A convergent result can however be obtained if the main verb is copied from K and merges with L (an instance of sideward movement), as illustrated in (95), and the resulting VP merges with K, as shown in (96a); a light verb is then selected from the numeration and its strong V-feature triggers adjunction of *talked* via sideward movement, yielding R in (96b). With the merger of P and R, the structure in (97) is finally derived. The computational system then forms the chains $CH_1 = (copy^3, copy^1)$ and $CH_2 = (copy^3, copy^2)$, whose optimal reduction in the phonological component involves the deletion of the lower copies, as shown in (98), yielding what at the surface seems to be a coordination of PPs (see (92a)).

(94) a. K $=$ $[_{andP}$ and $[_{VP}$ talked^i $[_{PP}$ to Maria]]]
     b. L $=$ $[_{PP}$ to João]

(95) a. K $=$ $[_{andP}$ and $[_{VP}$ talked^i $[_{PP}$ to Maria]]]
     b. M $=$ $[_{VP}$ talked^i $[_{PP}$ to João]]

(96) a. P $=$ $[_{andP}[_{VP}$ talked^i to João] $[_{and'}$ and $[_{VP}$ talked^i to Maria]]]
     b. R $=$ $[_{v}$o talked^i $[_{v}$o v]]

(97) $[_{vP}[_{v}$o talked^3 $[_{v}$o v]] $[_{andP}[_{VP}$ talked^2 to João] $[_{and'}$ and $[_{VP}$ talked^1 to Maria]]]]

(98) $[_{vP}[_{v}$o talked^3 $[_{v}$o v]] $[_{andP}[_{VP}$ ~~talked²~~ to João] $[_{and'}$ and $[_{VP}$ ~~talked¹~~ to Maria]]]]

Notice that a sideward analysis along the lines of (94)–(98) is not available for (92b), which has a single instance of the preposition in its numeration. Copying the verb from a structure such as K in (94a) and merging it with the DP *o João* does not give rise to a well-formed result (*conversei o João* 'talked João'), for Case reasons. Under the plausible assumption that each verb chain in (97) is interpreted as an event, the contrast in (93) follows straightforwardly: only (93a) can have a convergent derivation involving sideward movement along the lines of (94)–(98); hence, each event associated with a verb chain can be independently modified.

6 Conclusion

As discussed in Chomsky 1993, 1994, 1995, one of the key features of the analysis of movement operations in the Minimalist Program is its commitment to the copy theory. The adoption of the copy theory of movement is conceptually justified in that it allows interpretive phenomena having to do with binding theory and idioms to be accounted for without resorting to noninterface levels (see Chomsky 1993); it is also empirically justified by phonetically realized traces such as the ones discussed in section 2.2. This difference aside, the Minimalist Program as formulated in Chomsky 1993, 1994, 1995, essentially follows the traditional analysis of movement as a complex singulary transformation. More specifically, movement under the copy theory is taken to involve the suboperations of copying, merger, and chain formation, coupled with the operation of trace deletion for PF purposes (if the movement is overt).

This analysis of movement faces several conceptual problems: Merge is taken to be either a ''full'' operation or a suboperation of Move (see Gärtner 1997); Move must proceed noncyclically in instances involving head movement, covert movement, and certain cases of relative clause adjunction; movement and chain formation capture the same type of relation (see Brody 1995); and there is no motivation for deletion of traces. The last point also leads to empirical problems in that it is not obvious why some traces can escape deletion.

In this article, I have argued instead that Move is not a primitive operation of the computational system; rather, it is the mere reflex of the interaction among the independent operations Copy, Merge, Form Chain, and Chain Reduction. Under this conception of movement, which I referred to as the Copy + Merge theory of movement, all the conceptual problems are overcome.

Starting with the last one mentioned above, deletion of traces is taken to be triggered by linearization in the phonological component. Assuming that copies count as nondistinct for interpretation at LF, as well as for ''syntactic'' computations in the phonological component, a structure containing copies cannot be linearized in accordance with Kayne's (1994) LCA because the copies induce violations of the asymmetry and irreflexivity conditions on linear order. In order to allow a structure containing a chain to be linearized, the phonological component employs the operation Chain Reduction, which in general has the effect of deleting all but one link of a chain. The choice of the link to be kept is determined by economy considerations regarding the elimination of formal features in the phonological component. In the standard case, that means that only the head of the chain survives Chain Reduction; however, if a given chain link becomes invisible to the LCA in virtue of being reanalyzed as part of a word, it will also survive Chain Reduction, as illustrated by *wh*-copying and clitic reduplication constructions (see section 2.2). The Copy + Merge theory of movement thus meets both conceptual and empirical demands in regard to deletion of traces in the phonological component.

By taking Copy, Merge, Form Chain, and Chain Reduction to be independent operations, the system argued for here crucially differs from the standard analysis of movement within the principles-and-parameters framework in that it allows instances of sideward movement, where a given constituent of a syntactic object is copied and merges with an independent syntactic object. In these cases, applications of Copy and Merge are dissociated from Form Chain and Chain

Reduction, eliminating the redundancy between chain formation and movement that exists in the standard analysis. Once sideward movement is permitted in the system, head movement, adjunction of formal features, and adjunction of relative clauses can always proceed cyclically, paving the way for the elimination of noncyclic movement in the theory of grammar. Thus, Merge is always a ''full'' operation concatenating root syntactic objects.

Unwanted applications of sideward movement are ruled out by the linearization considerations that trigger deletion of traces: if two copies do not form a chain or if the independent chains containing them do not have a link in common, these copies cannot be targeted by Chain Reduction and end up preventing the structure from being linearized by violating the asymmetry and irreflexivity conditions on linear order.

The Copy + Merge theory of movement also broadens the empirical coverage of standard analyses of movement by deriving the core properties of parasitic gap and across-the-board constructions from the analysis independently motivated by instances of ''regular'' movement. In particular, sideward movement is responsible for the possibility of there being multiple gaps, and potential unwanted instances of multiple gaps either do not arise given the cyclic derivational nature of the computation or are excluded by general linearization considerations. To the extent that parasitic gaps and across-the-board extraction are handled without construction-specific operations or the introduction of principles that are not independently needed, they lend strong conceptual and empirical support to the Copy + Merge theory of movement and the proposal regarding linearization of chains.

References

Aoun, Joseph, and Robin Clark. 1985. On non-overt operators. In *Southern California occasional papers in linguistics,* 17–85. Department of Linguistics, University of Southern California, Los Angeles.

Aoun, Joseph, and Jairo Nunes. 1997. Vehicle change and Move-F. Paper presented at the Colloque de Syntaxe et Sémantique de Paris - CSSP 1997, Université de Paris 7, October 16–18.

Barss, Andrew. 1986. Chains and anaphoric dependencies. Doctoral dissertation, MIT, Cambridge, Mass.

Bobaljik, Jonathan. 1995a. In terms of Merge: Copy and head movement. In *MIT working papers in linguistics 27: Papers on minimalist syntax,* 41–64. MITWPL, Department of Linguistics and Philosophy, MIT, Cambridge, Mass.

Bobaljik, Jonathan. 1995b. Morphosyntax: The syntax of verbal inflection. Doctoral dissertation, MIT, Cambridge, Mass.

Bobaljik, Jonathan, and Samuel Brown. 1997. Interarboreal operations: Head movement and the Extension Requirement. *Linguistic Inquiry* 28:345–356.

Bošković, Željko. 1999. What is special about multiple *wh*-fronting. Paper presented at NELS 30, Rutgers University.

Bošković, Željko. 2000. On the nature of the syntax-phonology interface: Cliticization in South Slavic. Ms., University of Connecticut, Storrs.

Brody, Michael. 1995. *Lexico-Logical Form: A radically minimalist theory.* Cambridge, Mass.: MIT Press.

Brody, Michael. 1998. Projections and phrase structure. *Linguistic Inquiry* 29:367–437.

Browning, M. A. 1987. Null operator constructions. Doctoral dissertation, MIT, Cambridge, Mass.

Chomsky, Noam. 1975. *The logical structure of linguistic theory.* New York: Plenum.

Chomsky, Noam. 1981. *Lectures on government and binding.* Dordrecht: Foris.

Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding.* Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1986. *Barriers.* Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger,* ed. Kenneth Hale and Samuel Jay Keyser, 1–52. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1994. Bare phrase structure. MIT Occasional Papers in Linguistics 5. MITWPL, Department of Linguistics and Philosophy, MIT, Cambridge, Mass. [Published in *Government and Binding Theory and the Minimalist Program,* ed. Gert Webelhuth, 383–439. Oxford: Blackwell (1995).]

Chomsky, Noam. 1995. Categories and transformations. In *The Minimalist Program,* 219–394. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik,* ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, Mass.: MIT Press. [Also available as MIT Occasional Paper in Linguistics 15, MITWPL, Department of Linguistics and Philosophy, MIT, Cambridge, Mass.]

Chomsky, Noam, and Howard Lasnik. 1993. The theory of principles and parameters. In *Syntax: An international handbook of contemporary research,* ed. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann, 506–569. Berlin: Walter de Gruyter.

Cinque, Guglielmo. 1990. *Types of Ā-dependencies.* Cambridge, Mass.: MIT Press.

Contreras, Heles. 1984. A note on parasitic gaps. *Linguistic Inquiry* 15:704–713.

Du Plessis, Hans. 1977. *Wh*-movement in Afrikaans. *Linguistic Inquiry* 8:723–726.

Dyła, Stefan. 1984. Across-the-board dependencies and Case in Polish. *Linguistic Inquiry* 15:701–705.

Engdahl, Elisabet. 1983. Parasitic gaps. *Linguistics and Philosophy* 6:5–34.

Fanselow, Gisbert, and Anoop Mahajan. 1995. Partial movement and successive cyclicity. In *Papers on* wh-*scope marking,* ed. Uli Lutz and Gereon Müller, 131–161. Bericht Nr. 76, Arbeitspapiere des Sonderforschungsbereichs 340, Sprachtheoretische Grundlagen für die Computer Linguistik.

Fiengo, Robert, and Robert May. 1994. *Indices and identity.* Cambridge, Mass.: MIT Press.

Frampton, John. 1990. Parasitic gaps and the theory of *wh*-chains. *Linguistic Inquiry* 21:49–77.

Franks, Steven. 1993. On parallelism in across-the-board dependencies. *Linguistic Inquiry* 24:509–529.

Franks, Steven. 1998. Clitics in Slavic. Paper presented at the Comparative Morphosyntax Workshop, Bloomington, Ind.

Freidin, Robert. 1986. Fundamental issues in the theory of binding. In *Studies in the acquisition of anaphora,* vol. 1, ed. Barbara Lust, 151–188. Dordrecht: Reidel.

Gärtner, Hans-Martin. 1997. Generalized transformations and beyond (Reflections on minimalist syntax). Doctoral dissertation, Johann-Wolfgang-Goethe Universität, Frankfurt.

Groat, Erich, and John O'Neil. 1996. Spell-Out at the interface: Achieving a unified syntactic computational system in the minimalist framework. In *Minimal ideas: Syntactic studies in the minimalist framework,* ed. Werner Abraham, Samuel David Epstein, Höskuldur Thráinsson, and C. Jan-Wouter Zwart, 113–139. Amsterdam: John Benjamins.

Haïk, Isabelle. 1985. The syntax of operators. Doctoral dissertation, MIT, Cambridge, Mass.

Heycock, Caroline. 1995. Asymmetries in reconstruction. *Linguistic Inquiry* 26:547–570.

Hiemstra, Inge. 1986. Some aspects of *wh*-questions in Frisian. *NOWELE* 8:97–110.

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

Hornstein, Norbert. 2000. *Move.* Oxford: Blackwell.

Hornstein, Norbert, and Jairo Nunes. 1999. On asymmetries between parasitic gap and across-the-board constructions. Ms., University of Maryland, College Park, and Universidade Estadual de Campinas.

Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. Doctoral dissertation, MIT, Cambridge, Mass.

Kato, Mary, and Jairo Nunes. 1998. Two sources for relative clause formation in Brazilian Portuguese. Paper presented at the Eighth Colloquium on Generative Grammar, Universidade de Lisboa, Palmela (Portugal), April 19–22.

Kayne, Richard. 1983. Connectedness. *Linguistic Inquiry* 14:223–249.

Kayne, Richard. 1984. *Connectedness and binary branching*. Dordrecht: Foris.

Kayne, Richard. 1991. Romance clitics, verb movement, and PRO. *Linguistic Inquiry* 22:647–686.

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

Kayne, Richard. 1998. Overt vs. covert movement. *Syntax* 1:128–191.

Kearney, Kevin. 1983. Governing categories. Ms., University of Connecticut, Storrs.

É. Kiss, Katalin. 1985. Parasitic chains. *The Linguistic Review* 5:41–74.

Kitahara, Hisatsugu. 1995. Target α: Deducing strict cyclicity from derivational economy. *Linguistic Inquiry* 26:47–78.

Larson, Richard. 1988. On the double object construction. *Linguistic Inquiry* 19:335–391.

Lasnik, Howard. 1995. Last Resort and Attract F. In *Papers from the 6th Annual Meeting of the Formal Linguistics Society of Mid-America,* vol. 1, 62–81. Indiana University Linguistics Club, Bloomington.

Lasnik, Howard. 1998. Some reconstruction riddles. In *Proceedings of the 22nd Annual Penn Linguistics Colloquium,* 83–98. University of Pennsylvania Working Papers in Linguistics 5.1. Penn Linguistics Club, University of Pennsylvania, Philadelphia.

Lasnik, Howard, and Tim Stowell. 1991. Weakest crossover. *Linguistic Inquiry* 22:687–720.

Lebeaux, David. 1988. Language acquisition and the form of the grammar. Doctoral dissertation, University of Massachusetts, Amherst.

Manzini, Maria Rita. 1994. Locality, minimality, and parasitic gaps. *Linguistic Inquiry* 25:481–508.

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

McDaniel, Dana. 1986. Conditions on wh-chains. Doctoral dissertation, CUNY.

Munn, Alan. 1987. Coordinate structures and X-bar theory. In *McGill working papers in linguistics 4,* 121–140, Department of Linguistics, McGill University, Montreal, Quebec.

Munn, Alan. 1993. Topics in the syntax and semantics of coordinate structures. Doctoral dissertation, University of Maryland, College Park.

Munn, Alan. 1994. A minimalist account of reconstruction asymmetries. In *NELS 24,* 397–410. GLSA, University of Massachusetts, Amherst.

Nunes, Jairo. 1995. The copy theory of movement and linearization of chains in the Minimalist Program. Doctoral dissertation, University of Maryland, College Park.

Nunes, Jairo. 1996. On why traces are not phonetically realized. In *NELS 26,* 211–225. GLSA, University of Massachusetts, Amherst.

Nunes, Jairo. 1998. Bare X-bar theory and structures formed by movement. *Linguistic Inquiry* 29:160–168.

Nunes, Jairo. 1999. Linearization of chains and phonetic realization of chain links. In *Working minimalism,* ed. Samuel David Epstein and Norbert Hornstein, 217–249. Cambridge, Mass.: MIT Press.

Nunes, Jairo. 2000. Linearization of chains and sideward movement. Ms., Universidade Estadual de Campinas.

Nunes, Jairo, and Juan Uriagereka. 2000. Cyclicity and extraction domains. *Syntax* 3:20–43.

Oishi, Masayuki. 1997. Reducing Procrastinate to feature strength. *Interdisciplinary Information Sciences* 3:65–70.

Pesetsky, David. 1995. *Zero syntax*. Cambridge, Mass.: MIT Press.

Pesetsky, David. 1997. Optimality Theory and syntax: Movement and pronunciation. In *Optimality Theory: An overview,* ed. Diana Archangeli and D. Terence Langendoen, 134–170. Malden, Mass.: Blackwell.

Postal, Paul M. 1993. Parasitic gaps and the across-the-board phenomenon. *Linguistic Inquiry* 24:734–754.

Postal, Paul M. 1994. Parasitic and pseudoparasitic gaps. *Linguistic Inquiry* 25:63–117.

Ross, John R. 1967. Constraints on variables in syntax. Doctoral dissertation, MIT, Cambridge, Mass.

Safir, Ken. 1987. The Anti-C-Command Condition on parasitic gaps. *Linguistic Inquiry* 18:117–171.

Safir, Ken. 1998. Reconstruction and bound anaphora: Copy theory with deletion at LF. Ms., Rutgers University, New Brunswick, N.J.

Taraldsen, Knut Tarald. 1981. The theoretical implications of a class of marked extractions. In *Theory of markedness in generative grammar,* ed. Adriana Belletti, Luciana Brandi, and Luigi Rizzi, 475–516. Scuola Normale Superiore, Pisa.

Thornton, Rosalind. 1990. Adventures in long-distance moving: The acquisition of complex *wh*-questions. Doctoral dissertation, University of Connecticut, Storrs.

Uriagereka, Juan. 1995. Aspects of the syntax of clitic placement in western Romance. *Linguistic Inquiry* 26:79–124.

Uriagereka, Juan. 1998. *Rhyme and reason: An introduction to minimalist syntax.* Cambridge, Mass.: MIT Press.

Uriagereka, Juan. 1999. Multiple Spell-Out. In *Working minimalism,* ed. Samuel David Epstein and Norbert Hornstein, 251–282. Cambridge, Mass.: MIT Press.

Vergnaud, Jean-Roger. 1974. French relative clauses. Doctoral dissertation, MIT, Cambridge, Mass.

Weinberg, Amy. 1988. Locality principles in syntax and parsing. Doctoral dissertation, MIT, Cambridge, Mass.

Williams, Edwin. 1978. Across-the-board rule application. *Linguistic Inquiry* 9:31–43.

Williams, Edwin. 1989-1990. The ATB theory of parasitic gaps. *The Linguistic Review* 6:265–279.

Caixa Postal 6045
Instituto de Estudos da Linguagem
Universidade Estadual de Campinas
13083-970 Campinas, SP
Brazil

nunes@iel.unicamp.br