On Asymmetries between Parasitic Gap and Across-the-Board Constructions

Norbert Hornstein and Jairo Nunes

Abstract. Assuming that parasitic gaps (PGs) and across-the-board (ATB) gaps are derived through sideward movement (Nunes 1995, 2001 and Hornstein 2001), this paper aims at explaining why ATB constructions are in general more permissive than PG constructions (Postal 1993). We argue that sideward movement is licensed only by Last Resort in PG constructions, but is licensed but either Last Resort or the Parallelism Constraint in ATB constructions. The additional legibility condition that coordinate structures have to satisfy ends up permitting instances of sideward movement that would otherwise be impossible.

1. Introduction

As originally observed by Ross (1967), parasitic gap (PG) and across-the-board (ATB) constructions are similar in that they appear to involve extraction of a single element from more than one position at once:

\begin{align*}
1 & \quad (1) \quad \text{(a. [ which paper ]}_i \text{ did you read } e_i \text{ after Mary recommended } e_i) \\
& \quad \text{b. [ which paper ]}_i \text{ did } e_i \text{ you read } e_i \text{ and Mary } e_i \text{ recommend } e_i
\end{align*}

This similarity has provoked attempts to assimilate one construction to the other. Haïk (1985) and Williams (1990), for instance, have proposed that PG constructions are to be treated in terms of ATB extraction, whereas Munn (1992) proposes that ATB constructions involve movement of a null operator like the one proposed by Chomsky (1986a) to account for PGs.

There are conceptual problems with both approaches. The reduction in the direction of ATB crucially relies on a construction-specific ATB mechanism. It must further assume that every structure that allows a parasitic gap is optionally coordinative (see Munn 1992...
and Postal 1993 for detailed discussion of these points). Munn's null operator analysis of ATB extraction, on the other hand, is not general enough to account for ATB movement of categories other than DPs; under standard assumptions, the ATB extraction of the auxiliary in (1b), for instance, is not amenable to a null operator analysis. In addition to these conceptual problems, Postal (1993) has presented an extensive empirical documentation of restrictions that show up in PG, but not in ATB constructions. As Postal (p. 736) puts it, "those English constituents that can be ATB gaps are a vast superset of the class that can form [PGs]." Postal takes these restrictions as evidence against the account of PGs in terms of ATB extraction, but they clearly pose questions to the reduction of ATB in the direction of PGs, as well.

In this paper, we subscribe to the (unoriginal) view that multiple gaps in PG and ATB constructions arise in a similar manner, framing the discussion in the broad context of Chomsky’s (1995) Minimalist Program. From a minimalist perspective, the postulation of distinct ATB and 0-operator grammatical processes is methodologically undesirable. We therefore explore an approach under which the general processes that yield “standard” gaps also derive ATB and PG gaps. More specifically, we assume that all of these gaps arise through the applications of the operations Copy, Merge, and Delete. After the computational system copies a given constituent $\alpha$ of a syntactic object $K$, as illustrated in (2) below, for instance, it has two options: it can merge the copy of $\alpha$ either with $K$, as represented in (3), or with a syntactic object $L$, which has been independently assembled and is unconnected to $K$, as represented in (4). The first option yields an instance of “standard” upward movement and the second option, an instance of “sideward movement”, which we assume to be crucial in the derivation of PG and ATB gaps.\

\begin{align*}
\text{(2) } & \quad [L \ldots ] \quad \alpha \quad [K \ldots \alpha \ldots ] \\
& \quad \text{Copy} \\
\text{(3) a. } & \quad [L \ldots ] \quad \alpha \quad [K \ldots \alpha \ldots ]
\end{align*}
It should be emphasized that sideward movement along the lines of (2) and (4) is in consonance with the general methodological guidelines of the Minimalist Program in that it is not a new operation that is being added to the set of theoretical primitives. Quite the opposite. The stipulation that a given copy must be merged with the syntactic object that contained the “original” is actually a residue of D-Structure. If a unique single-rooted object is presented to the computational system, a moved element is bound to merge with a syntactic object that contains its trace. If no D-Structure is assumed and the computational system resorts to generalized transformations to build phrasal objects, a copied element may have choices other than a phrase containing the “original” to merge with. In other words, in a system that may operate with more than one single-rooted syntactic object at once, as in Chomsky 1995, only brute force would force movement to always target the same tree.

Assuming that PG and ATB constructions are both derived through sideward movement (see Nunes 1995, 2001 and Hornstein 2001), the aim of this paper is to provide a minimalist perspective on these constructions that might explain why they exhibit the asymmetries discussed by Postal (1993), despite their similar derivational histories. We believe that the conceptual resources of minimalism offer a distinctive approach to the fact that ATB movement is more expansive than that found within PG constructions. The novelty can be briefly described as follows. Within the GB model, the natural approach is to take the ATB constructions as manifesting the fundamental features of the two constructions because they permit a wider range of licit movements. PG constructions are then analyzed as ATB plus additional conditions that serve to limit the applicability of the
relevant operations in their full generality, by restricting the kinds of extractable expressions. We aim to show that a minimalist perspective suggests the opposite analysis: rather than PGs being restricted ATB gaps, on this view ATB gaps are "loosened" PGs. In other words, an interface requirement special to ATBs leads to a greater permissiveness in extraction. Thus, in contrast to GB-style theories, where constraints reduce the outputs of the computational system, in a minimalist perspective extra output conditions may serve to free the computational system, allowing it to generate structures that would otherwise be illicit.

The paper is organized as follows. In section 2, we show how sideward movement can yield multiple gaps in PG and ATB constructions. In section 3, we argue that ATB constructions are in a sense more permissive than PG constructions because the Parallelism Requirement may also license instances of (sideward) movement. Section 4 addresses the issue of how the Parallelism Requirement and sideward movement can be computed locally. Section 5 discusses further the role of Case is licensing sideward movement and section 6 presents its consequences as far as the directionality of sideward movement is concerned. Finally, a brief conclusion is presented in section 7.

2. Sideward Movement and Multiple Gap Constructions
Exploring Chomsky's (1995:chap. 3) revival of the copy theory, Nunes (1995, 2001) argues that Move is not a primitive operation of the computational system, but is rather the reflex of the interaction among the more basic operations Copy, Merge, and Delete (see Hornstein 2001 for discussion and further developments).\(^2\) The derivation of the sentence in (5), for instance, proceeds along the lines of (6).

(5) John was arrested.

(6) \(\begin{align*}
\text{a. } K &= \text{ [ was [ arrested John\textsuperscript{i} ] ]} \\
\text{b. } L &= \text{ John\textsuperscript{i}} \\
\text{c. } M &= \text{ [ John\textsuperscript{i} [ was [ arrested John\textsuperscript{i} ] ] ]}
\end{align*}\)
d. \( M' = [\text{John}^1 \ [\text{was} \ [\text{arrested} \ \text{John}^1] \ ]] \)

Given the syntactic object \( K \) in (6a), the computational system makes a copy of \( \text{John} \), as shown in (6b), and merges it with \( K \), yielding \( M \) in (6c) (copies will henceforth be annotated with superscripted indices). Assuming that copies count as the same element for linearization purposes because they refer to the same lexical items of the initial numeration (see Nunes 1995, 1999), no linear order can be established for (6c) as it is; according to Kayne’s (1994) LCA, the copula \( \text{was} \), for instance, should precede and be preceded by \( \text{John} \), a contradictory requirement. In order for \( M \) to be properly linearized, the phonological component then deletes the lower copy of \( \text{John} \), as shown in (6d), yielding the sentence in (5).³

In (6), the copy of \( \text{John} \) merged with a category that dominated the "original", namely, \( K \). But if Copy and Merge are distinct operations, that need not be the case, as abstractly illustrated by sideward movement in (4). Below we discuss concrete instances of sideward movement by examining derivations of PG and ATB constructions.

### 2.1. Parasitic Gaps

Let us consider the (simplified) derivation of the PG construction in (1a), repeated below in (7a) (see Nunes 1995, 2001 and Hornstein 2001 for details), assuming that it starts with numeration in (7b) (irrelevant items are omitted).

(7)  
   a. Which book did you read after Mary recommended?
   b. \( N = \{\text{which}_1, \text{book}_1, \text{did}_1, \text{Q}_1, \text{you}_1, \text{read}_1, \text{after}_1, \text{Mary}_1, \text{T}_1, \text{recommended}_1\} \)

After the numeration \( N \) in (7b) has been reduced to \( N' \) in (8) below and the two syntactic objects in (9) have been assembled, the verb \( \text{read} \) has two \( \theta \)-roles to assign, but there is only one possible \( \theta \)-role bearer in \( N' \), namely, \( \text{you} \). A convergent continuation for this derivational stage can however be obtained if the computational system copies \( \text{which} \)
book from K in (9a) and merges it with read in (9b), yielding M in (10) (an instance of sideward movement).  

\[(8) \quad N' = \{ \text{which}_0, \text{book}_0, \text{did}_1, Q_1, \text{you}_1, \text{read}_0, \text{after}_1, \text{Mary}_0, T_0, \text{recommended}_0 \}\]

\[(9) \quad \begin{align*}
  a. \quad & K = [ \text{Mary recommended [ which book ] } ] \\
  b. \quad & L = \text{read}
\end{align*}\]

\[(10) \quad \begin{align*}
  a. \quad & K = [ \text{Mary recommended [ which book ] } ] \\
  b. \quad & M = [ \text{read [ which book ] } ]
\end{align*}\]

Further computations exhaust N' and yield the (simplified) structure in (11a) below. In the following steps, the computational system makes another copy of which book and merges it with (11a) to check the strong wh-feature of the interrogative complementizer Q, yielding the CP in (11b). The highest wh-copy in (11b) then establishes a different chain with each of the lower copies. Finally, in order for the whole structure to be linearized, the phonological component deletes the traces of each wh-chain, as shown in (11c), and the PF output associated with (7a) is derived.

\[(11) \quad \begin{align*}
  a. \quad & [CP \text{ did+Q } TP \text{ you } [VP \text{ read [ which book ] } ] ] ] ] ] \\
  b. \quad & \begin{align*}
    & \text{CP} \\
    & [ \text{ which book ] } ] \\
    & \text{C'} \\
    & \text{did+Q} \\
    & \text{TP} \\
    & \text{you} \\
    & \text{T'} \\
    & \text{T} \\
    & \text{VP} \\
    & [\text{VP read [ which book ] } ] \\
    & \text{[PP after Mary recommended [ which book ] ] } ]
  \end{align*}
\]
To illustrate how sideward movement is adequately constrained and does not lead to overgeneration, let us consider the anti-c-command requirement on PGs (see Taraldsen 1981, Engdahl 1983, among others). As is well known, PGs cannot be c-commanded by “real” gaps; hence, the unacceptability of (12), for example, is attributed to the fact that the PG is c-commanded by the trace of who in [Spec, TP].

(12) *I wonder [CP who [TP ti [VP ti called Mary ] [ after she visited ei ] ] ]

Under the approach sketched above, there is nothing wrong with the sideward movement from K to L in (13) below, which would be required to derive the PG construction in (12). The problem arises after the entire (simplified) structure in (14) is built (numbered copies were used for purposes of presentation).

(13)  a. K = [ she visited who1 ]
      b. L = [ who1 [ called Mary ] ]

(14)  [ I wonder [CP who1 [TP who2 [VP who3 called Mary ] [ after she visited who4 ] ] ] ]

The copies who1, who2, and who3 form the (linked) chain CH = (who1, who2, who3), but who4 cannot form a chain with any other copy. It cannot form a chain with who1, because neither copy c-commands the other; in addition, it cannot for a chain with who2 or who1 due to minimality: she intervenes between who2 and who4 and who2 intervenes between who1 and who4. Thus, the operation Delete can only target the lower links of CH (see fn. 3), yielding the structure in (15) below. The two remaining copies of who in (15) then prevent
the structure from being linearized (Mary, for instance, would be required to follow and precede who) and the derivation is canceled.

(15) \[ \text{I wonder } [\text{CP who}^1 \ [\text{TP who}^2] \ [\text{vP who}^3 \text{ called Mary } ] \ [\text{after she visited who}^4] ] ] \]

If the PG is not c-commanded by a real gap, as exemplified in (16) below, minimality is satisfied and we obtain an acceptable PG construction. Under the approach reviewed here, the relevant structure of (16) is (17a), which is formed by sideward movement of which paper from the object of reading to the thematic position associated with interesting. Given (17a), the computational system deletes the lower links of the chains \( \text{CH}_1 = ([\text{which paper}]^1, [\text{which paper}]^2, [\text{which paper}]^3, [\text{which paper}]^4) \) and \( \text{CH}_2 = ([\text{which paper}]^1, [\text{which paper}]^5) \), yielding the linearizable structure in (17b).

(16) \[ \text{[ [ which paper ] did John [vP [vP say [CP t_i \ [TP t_i \text{ was t_i interesting } ] ] ] ] [ after reading e_i ] ] ]} \]

(17) a. \[ \text{[ [ which paper ] did John [vP [vP say [CP [which paper ]^2 \ [TP [which paper ]^3 \text{ was [which paper ]^4 interesting } ] ] ] [ after reading [which paper ]^5 ] ] ] } \]

b. \[ \text{[ [ which paper ] did John [vP [vP say [CP [which paper ]^2 \ [TP [which paper ]^3 \text{ was [which paper ]^4 interesting } ] ] ] [ after reading [which paper ]^5 ] ] ] } \]

The sideward movement analysis of PGs is therefore able to derive anti-c-command effects without expanding the theoretical apparatus or resorting to construction specific mechanisms. The reasoning used to account for deletion of copies in constructions involving sideward movement was the same as that used in the case of upward movement.\(^6\)
2.2. ATB Gaps

The derivation of the ATB construction in (1b), repeated below in (18a), proceeds along the same lines of the derivation of the PG construction in (7a). Assuming that the (simplified) numeration underlying the derivation is the one given in (18b) and that the coordinating conjunction heads its own projection (see Munn 1987 for an early proposal), the computational system selects lexical items of N and forms the two objects in (19).

\[(18) \quad \text{a. Which book did you read and Mary recommend?}
\]
\[(18) \quad \text{b. N = \{which}_1, \text{book}_1, \text{did}_1, \text{Q}_1, \text{you}_1, \text{read}_1, \text{and}_1, \text{Mary}_1, \text{recommend}_1 \}
\]

\[(19) \quad \text{a. K = [andP and [Mary did recommend [which book]]]}
\]
\[(19) \quad \text{b. L = read}
\]

As seen earlier, the derivation can only converge if \textit{which book} is copied from K and merged with \textit{read}. The only relevant difference in the present case is that the unique Tense element of the numeration, namely, \textit{did}, has already been used to build K in (19a); thus, a convergent continuation of the stage in (19) will also require sideward movement of \textit{did}. After these two instances of sideward movement and further computations, the structure in (20a) below is formed. The strong features of Q then trigger the creation of additional copies of \textit{did} and \textit{which book}, and the CP in (20b) is assembled. As before, the highest copy of \textit{which book} and \textit{did} establishes a distinct chain with each of their lower copies. Finally, the phonological component deletes the traces of each chain involving \textit{did} and \textit{which book} in order for the CP to be linearized, as shown in (20c), yielding the sentence in (18a).

\[(20) \quad \text{a. [CP Q [andP [you did}^k \text{read [which book]_i] [and}^i \text{and [Mary did}^k \text{recommend [which book]_i]]]]}
\]
\[(20) \quad \text{b. CP}
\]
To sum up, once Move is analyzed as a taxonomic notion and not an operation of
the computational system, the existence of multiple gaps such as the ones found in PG and
ATB constructions does not come out as a surprise; they arise in the same manner gaps of
standard movement do: as a result of the interaction among the independent operations of
Copy, Merge, and Delete. The reader is referred to Nunes 1995, 2001 and Hornstein 2001
for general discussion and further evidence for sideward movement. From now on we focus
our attention on the differences between PG and ATB constructions pointed out by Postal
(1993).

3. Sideward Movement, Last Resort, and the Parallelism Requirement

The PG and ATB constructions in (21) and (22) illustrate Postal's (1993) observation that
PGs are more restricted than ATB gaps: PGs cannot be licensed by AdvPs, APs, PPs,
nonreferential NPs, or auxiliaries, whereas any of these categories can undergo ATB
extraction.7

(21) a. *how, did Deborah cook the pork e, after cooking the chicken e,
b. *[ how sick ], did John look e, without actually feeling e,
c. *this is a topic [ about which ], you should think e, before
talking e,
d. *[ how many weeks ], did he spend e, in Berlin without wanting
to spend e₁ in London

e. *did₁ John e₁ call Mary after Paul e₁ call Sue

(22) a. how₁ did Deborah cook the pork e₁ and Jane cook the chicken e₁
b. [ how sick ]₁ did John look e₁ and Betty say he actually felt e₁
c. this is a topic [ about which ]₁ you should think e₁ and I should talk e₁
d. [ how many weeks ]₁ did you spend e₁ in Berlin but want to spend e₁ in
   London
e. did₁ John e₁ call Mary and Paul e₁ call Sue

If PG and ATB constructions are both derived along the lines presented in section 2, the question that arises is why sideward movement yields acceptable results in (22), but not in (21). Let us start the discussion by considering how sideward movement in acceptable PG constructions satisfies Chomsky's 1995 Last Resort condition, according to which every movement operation must be properly motivated. Given the decomposition of Move in the more basic operations Copy, Merge, and Delete discussed in section 2, we accordingly reinterpret Last Resort as a condition on the operation of Copy. Following Hornstein 1998, 2001 (see also Boškovic’ 1994, Lasnik 1995, and Boškovic’ and Takahashi 1998, among others), we further assume that Last Resort is satisfied by formal feature checking, as well as θ-role assignment/checking. In (10), repeated here as (23), for instance, sideward movement complies with Last Resort, for the copying of which book is triggered by the θ-requirements of the verb read.

(23) a. \( K = [ \text{Mary recommended} [\text{which book}]^1 ] \)
b. \( M = [ \text{read} [\text{which book}]^1 ] \)

If this reasoning is on the right track, we have an explanation for the unacceptability of the PG constructions in (21): none of the instances of sideward movement that are
required to derive the relevant sentences are licensed by Last Resort. Let us examine each of them, starting with the more transparent case in (24).

(24)  
   a. K = [PRO cooking the chicken how ]
   b. L = [ cook the pork ]

The copying of how from K to L in (24) is not triggered by either formal feature checking or a θ-relation (how would be an adjunct in L). Once this copying is not licensed by Last Resort, the sideward movement of how in (24) does not take place and the PG construction in (21a) cannot be generated.

Similar considerations apply to the instances of sideward movement of the elements in bold in (25)-(27) from K to L, which should be required in the derivation of the PG constructions in (21b-d).

(25)  
   a. K = [ PRO actually feeling [ how sick ] ]
   b. L = look

(26)  
   a. K = [ PRO talking [ about which ] ]
   b. L = think

(27)  
   a. K = [ PRO spend [ how many weeks ] in London ]
   b. L = spend

Although the verbs in L in (25)-(27) respectively select for an AP, a PP and a measure NP, they arguably do not assign a θ-role to these elements; therefore, they do not trigger the copying of these elements from K. Once the required instances of (sideward) movement are not licensed by Last Resort, the PG constructions in (21b-d) cannot be generated.9

Under this approach, the paradigm in (28) below ((28a-b) are from Postal 1993), whose sentences are minimally different, is accounted for as follows. In (28a), θ-role
assignment by the preposition *in* in the matrix clause triggers the copying of *what city* and licenses its (sideward) movement from the object of *in* inside the adjunct clause. In (28b) and (28c), by contrast, there is no \( \theta \)-relation or feature checking associated with the verb *work* that could license the copying of *where* and *in what city*; consequently, the derivations of the PG constructions in (28b) and (28c) are ruled out because the required sideward movements are not licensed by Last Resort.\(^{10}\)

(28)  
\[
\begin{align*}
\text{a. } & \text{[ what city ], did Elaine work in } e_i \text{ without ever living in } e_i \\
\text{b. } & \text{*where, did Elaine work } e_i \text{ without ever living } e_i \\
\text{c. } & \text{*[ in what city ], did Elaine work } e_i \text{ without ever living } e_i
\end{align*}
\]

It should be pointed out that the contrast between true arguments and non-\( \theta \)-marked selected material is not restricted to cases of sideward movement, but also appears in standard instances of upward movement. As extensively discussed in the GB literature (see Cinque 1990, for instance), the elements in bold in (25)-(27), although selected, actually behave like adjuncts in being unable to move across a weak island, as illustrated in (29).

(29)  
\[
\begin{align*}
\text{a. } & \text{*how, do you wonder whether Jane cooked the chicken } e_i \\
\text{b. } & \text{*[ how sick ], do you wonder whether John felt } e_i \\
\text{c. } & \text{*this is a topic [ about which ], I wonder whether you want to talk } e_i \\
\text{d. } & \text{*[ how many weeks ], do you wonder whether he spent } e_i \text{ in Berlin}
\end{align*}
\]

In the framework of Chomsky (1981, 1986a), for instance, it must be assumed that the verbs of L in (25)-(27) do not \( \theta \)-govern the selected constituents; otherwise, the relevant traces in (29b-d) would satisfy the ECP and the sentences would be incorrectly ruled in.\(^{11}\) It is not our purpose in this paper to attempt to provide an explanation for why the system treats selection and \( \theta \)-role assignment/checking differently; rather, we will rely on this traditional distinction as a reliable diagnosis for the kind of licensing condition a given operation is subject to.\(^{12}\)
Let us finally consider the sideward movement of the auxiliary in (30), which would be necessary to derive the PG construction in (21e).

\[(30)\]
\[
\begin{align*}
\text{a. } K &= \text{ [ Paul did call Sue ]} \\
\text{b. } L &= \text{ [ VP John call Mary ]}
\end{align*}
\]

Leaving aside the issue of whether or not the selectional relation between the auxiliary \textit{did} and vP in (30b) involves some sort of \(\theta\)-role (see Chomsky 1986a, for instance), it is arguably the case that T takes VP as its argument and not vice versa. Thus, the VP in L does not trigger the copying of \textit{did} and its sideward movement is not licensed by Last Resort; hence, the PG construction in (21e) cannot be generated.

Let us assume that this approach to PGs is (at least roughly) on the right track. If it is, we must now explain why the instances of sideward movement required to generate the ATB constructions in (22) yield acceptable results despite the fact that they are not licensed by Last Resort. One possible reason is that these instances of sideward movement are licensed by a condition that applies to ATB but not to PG constructions. There is an obvious candidate: the Parallelism Requirement manifested by coordinate structures. In the specific case of movement operations, the Parallelism Requirement demands that movement apply to all the conjuncts if it applies to any (Ross's 1967 Coordinate Structure Constraint). In this one respect, extraction from within conjuncts is indeed less permissive than extraction from within adjuncts, as illustrated in (31).

\[(31)\]
\[
\begin{align*}
\text{a. } \text{[ which paper ]}, \text{ did you read } e_i \text{ and Mary recommend } e_i/\ast [ \text{ this book} ] \\
\text{b. } \text{[ which paper ]}, \text{ did you read } e_i \text{ after Mary recommended } e_i/[ \text{ this book} ]
\end{align*}
\]

In standard ATB formats, the Parallelism Requirement is packed into the ATB formalism itself (see Williams 1978, for instance). In other analyses, it is treated as a semantic condition on well-formed coordinate structures. Munn (1993), for example, argues that the imposed parallelism does not require parallelism of overt gaps but semantic
function. We assume here that Munn (1993) is basically correct and that coordinate structures come with a requirement that their conjuncts be semantically "similar" (see section 4.1 below for further discussion). In a minimalist scheme, the Parallelism Requirement may thus be treated as a bare output condition (viz. a legibility condition imposed by the Conceptual-Intentional interface) on the interpretation of coordinate structures.

It is worth pointing out that what is at issue here is not whether to adopt a Parallelism Requirement on coordinates but how to implement this idea. To our knowledge, every approach to coordinates imposes some kind of parallelism restriction on these structures. We have no idea why it is that coordinate structures should require parallelism. However, there is plenty of empirical evidence suggesting that they do. Thus, we adopt the common wisdom that such constructions do indeed impose requirements on the conjuncts. What is (somewhat) distinctive here is the proposal that this parallelism be interpreted as a C-I interface requirement, a bare output condition on the legibility of coordinate structures (as suggested in Fox 1998). If the Parallelism Requirement is a bare output condition that applies to coordinates (and to them alone), the reason why more types of constituents can undergo sideward movement in ATB constructions is that the Parallelism Requirement may also license movement in order to ensure legibility at the C-I interface. In other words, the Parallelism Requirement, like Last Resort, may also license applications of the operation Copy.

Consider, for the sake of illustration, the two instances of sideward movement shown in (33), which are required in the derivation of the ATB construction in (22a), repeated as (32).

(32) \( \text{how}_i \text{ did}_k \text{ Deborah}_k \text{ cook the pork}_e, \text{ and Jane}_e \text{ cook the chicken}_e, \)

(33)  

a. \( K = [ \text{ and } [\text{TP} \text{ Jane } \text{did}^k \text{ cook the chicken } \text{how}^i] ] \)  

b. \( L = [\text{TP} \text{ Deborah } \text{did}^k \text{ cook the pork } \text{how}^i] \)
As discussed in the case of analogous PG constructions, the copying of *how* or *did* from K in (33) is not licensed by Last Resort. The copying of *how* is however licensed by the Parallelism Requirement in that it renders the two VPs parallel, by providing the VP of (33b) with a logical variable playing the same semantic function as the one in (33a). The Parallelism Requirement also licenses the copying of *did* by making possible the coordination of two like-categories (the two TPs). Similar considerations apply to the other ATB constructions in (22b-e).\textsuperscript{14,15}

We would like to acknowledge that at this point, we do not have an explanation for all the different kinds of asymmetries between PG and ATB constructions that Postal (1993) documents; neither do we have an account of the variation among speakers regarding some of these asymmetries.\textsuperscript{16} What we are proposing is a general approach within which it is possible to capture the similarity between PG and ATB constructions (their multiple gaps arise through sideward movement), while at the same time couching their differences on an independently motivated constraint on coordinate structures (the Parallelism Requirement). The directionality of the predictions of the present approach is however clear: the Parallelism Requirement may license applications of operations of the computational system that would not be licensed by Last Resort. However, this does not mean that the Parallelism Requirement may overrule other bare output conditions (see section 5 below for an example); all this means is that, as derivations unfold, the Parallelism Requirement functions as an enabling condition *widening* the derivational options within coordinate structures. In this way, the approach we are suggesting is qualitatively consistent with the observed superset/subset relation that Postal documents between ATB and PG constructions.

Independent evidence for the approach pursued here can be found in instances of (apparent) violations of Ross’s 1967 Coordinate Structure, as illustrated in (34) below (from Postal 1998:66). Whether sentences like (34) involve true coordination or some sort of adjunction is controversial (see Lakoff 1986, Postal 1998, and Levine, Hukari, and Calgano 2001 for detailed discussion). However, what is relevant for our discussion is the contrast between (34) and (35), noted by Postal (1998:66).
(34)  [ [ which student ], did Nora go to the drugstore, come home, and talk to e, for an hour ]

(35)  

a. *[ [ to which student ], did Nora go to the drugstore, come home, and talk e, for an hour ]

b. *[ how long ], did Nora go there, come home, and talk to that student e, ]

c. *[ [ very fat ], though Nora went to Italy, ate a lot, and became e, ]

The unacceptability of the sentences in (35) shows that if (for some reason) the Parallelism Requirement is inapplicable, as seen in (34), extraction from within conjuncts ceases to be free of restrictions and starts to mimic the restrictions found in PG constructions with respect to the kinds of constituents that can license a PG. The contrast between (34) and (35) thus provides empirical support for taking ATB constructions to be “loosened” PG constructions, that is, PG constructions with extra movement possibilities made available by the Parallelism Requirement.

To sum up, the Copy operation must be licensed either by Last Resort or by the Parallelism Requirement. If licensed by Last Resort, it allows a checking relation involving formal features or \( \theta \)-roles to be established; if licensed by the Parallelism Requirement, it allows coordinate structures to become semantically similar. Asymmetries between PG and ATB constructions are due to the fact that sideward movement is licensed only by Last Resort in the former, but is licensed by either Last Resort or the Parallelism Requirement in the latter. The additional legibility condition that coordinate structures have to meet thus ends up permitting instances of sideward movement that would otherwise be impossible.

4. Sideward Movement, the Parallelism Requirement, and Local Economy

4.1. Computing the Parallelism Requirement in a Local Fashion

At first sight, the licensing of sideward movement in terms of the Parallelism Requirement appears to necessarily require global computations, as opposed to licensing by Last Resort,
which may be invoked in every step of the derivation. However these kinds of motivations for movement are not as different as this suggests. Last Resort is ultimately the reflex of interpretability at the interface. In other words, in standard cases, the interface requirements are factored into a series of local checking relations in which otherwise uninterpretable features get progressively checked with every application of merger and movement. It is actually this translation from a general interface legibility condition into a more local requirement that seems to be lacking in the Parallelism Requirement.

It is likely possible to construe the licensing imposed by the Parallelism Requirement in an analogous manner if one is inclined to adopt a set of features to enforce it. However, it is not clear whether so proceeding will shed much light on the basic facts or analysis. We rather propose that the computational system relies on the content of coordinating heads in order to ensure that the Parallelism Requirement be computed in a more local fashion. That is, given that the Parallelism Requirement holds of coordinated structures, it rather natural to take coordinating heads to be the locally determinable elements that are able to define the relevant templates that the computational system must take into consideration when building parallel structures. So, the computational system need not look ahead at LF in order to make a decision as to whether or not to copy a given constituent in compliance with the Parallelism Requirement. Rather, after a coordinating head merges with a given constituent X, it signals that the computational system should proceed to build a constituent Y parallel to X, with the lexical items available at the relevant derivational step. If the available lexical material does not yield a (semantically) parallel structure, then the Parallelism Requirement, locally enforced by the coordinating head (or by the label of the syntactic object it heads), licenses the copying of constituents of X in order to build Y.

Evidence for this local computation of the Parallelism Requirement is provided by coordination involving PPs in Romance. Consider the Portuguese sentences in (36) (from Nunes 2001), for instance.

(36) a. Eu conversei com o João e com a Maria.
Both sentences in (36) allow an interpretation where there is a single event with three participants, but only (36a) permits a reading under which there are two events with two participants each (the subject and one of the objects). Let us consider why (36a) is ambiguous, assuming that its (simplified) initial numeration is the one given in (37) (English words are used for the sake of exposition).

(37) \( N = \{ I_1, v_1, talked_1, to_2, João_1, and_1, Maria_1 \} \)

Suppose that the computational system reduces the numeration \( N \) in (37) to \( N' \) in (38a) and build the syntactic object \( K \) in (38b). Once \( and \) has merged with a PP in (38b), the computational system proceeds to building another PP in order to merge it with \( K \), in accordance with the Parallelism Requirement. Further computations then yield the structure in (39), which is interpreted as the one-event reading.

(38) a. \( N' = \{ I_1, v_1, talked_1, to_1, João_1, and_0, Maria_0 \} \)
   b. \( K = [ and [ pp to Maria ] ] \)

(39) \([ TP I [ vp talked i + v [ vp talked i [ and p [ pp to João ] and' and [ pp to Maria ] ] ] ] ] \)

Suppose, on the other hand, that the numeration \( N \) in (37) is reduced as \( N'' \) in (40a) below and the syntactic objects \( L \) and \( M \) are formed. Given that \( and \) has now merged with a VP, the Parallelism Requirement should trigger the construction of another VP. However, there is no (main) verb available in \( N'' \). The computational system then copies \( talked \) from
L and merges it with M (an instance of sideward movement), yielding O in (41), which can then merge with L in compliance with the Parallelism Requirement. Further computations involve the creation of an additional copy of talked to check the strong feature of the light verb and the deletion of the trace of each verb chain, as illustrated in (42). Under the plausible assumption that each verb chain is interpreted as an event, (42) gives rise to the two-event reading of (36a).

\[(40)\]
\[
\begin{align*}
    a. \quad N'' &= \{I_1, v_1, talked_0, to_0, João_0, and_0, Maria_0\} \\
    b. \quad L &= [\text{and} [VP talked [PP to Maria ] ] ] \\
    c. \quad M &= [PP to João ]
\end{align*}
\]

\[(41)\]
\[
\begin{align*}
    a. \quad L &= [\text{and} [VP talked\text{\textsuperscript{\textcircled{d}}} [PP to Maria ] ] ] \\
    b. \quad O &= [VP talked\text{\textsuperscript{\textcircled{d}}} [PP to João ] ]
\end{align*}
\]

\[(42)\]
\[
\begin{align*}
\end{align*}
\]

By contrast, a sideward analysis along the lines of (40)-(42) is not available for (36b), because it has only one instance of the preposition in its initial numeration and one of the complement DPs would not have its Case checked (see Nunes 2001 for details). The one-event reading of (36b) is derived from the structure in (43), with a single verb chain.

\[(43)\]
\[
\begin{align*}
    [TP I [VP talked\text{\textsuperscript{\textcircled{d}}} +v [\text{andP} [VP talked\text{\textsuperscript{\textcircled{d}}} [PP to \text{andP} João [\text{and} and Maria ] ] ] ] ] ] ]
\end{align*}
\]

By taking into consideration the active numeration and the structure headed by a coordinating head, it is therefore possible to compute the Parallelism Requirement in a more local fashion, with interesting interpretive consequences depending on the specific derivational step when the coordinating head is merged.\textsuperscript{18}
4.2. Directionality of Sideward Movement

Let us reconsider the derivation of the PG construction in (7a), repeated in (44), which requires the sideward movement from K to L illustrated in (45).

(44) Which book did you read after Mary recommended?

(45) a. K = [ recommended [ which book ]]
    b. L = [ read [ which book ]]

It was tacitly assumed in the discussion of (44) that sideward movement of which book proceeds from the adjunct clause to the main clause, rather than the opposite. Although the matter does not arise in a representational system (see Brody 1995), the directionality of movement may be an issue in a derivational system such as the one we have been exploring. Nunes and Uriagereka (2000) and Hornstein (2001) in fact argue that cyclicity considerations dictate that movement (upward or sideward) always takes place from more to less embedded domains. The empirical basis for the argument has to do with paradigms such as the one in (46), which illustrates the well known fact that although PGs typically show up within islands, they cannot be separated from their licenser by more than one island (see Kayne 1984, Chomsky 1986a, among others).

(46) a. *which documents did John mention that law without checking e,  
    b. which documents did John mention e, without checking e,  
    c. *which documents did John mention e, after signing papers without  
       checking e, 

Assuming that adjuncts are islands for extraction (see Huang 1982), as exemplified in (46a), Nunes and Uriagereka (2000) and Hornstein (2001) show that (46b) is acceptable because (sideward) movement of which documents to the object of mention, as illustrated in (47) below, takes place at a derivational step where K is not an adjunct island (it is still an
independent syntactic object). After K in (47a) becomes part of an adjunct, as shown in (48a), its constituents are no longer extractable, but the copy that underwent sideward movement to the object of mention can move further to the matrix [Spec, CP] without crossing any islands, yielding the PG construction in (48b).

(47)  a. K = [ checking [ which documents ] ]  
b. L = [ mention [ which documents ] ]

(48)  a. [CP did+Q [TP John [vP mention [ which documents ] ] [island without PRO checking [ which documents ] ] ] ]  
b. [CP [ which documents ] did+Q [TP John [vP [vP mention [ which documents ] ] [island without PRO checking [ which documents ] ] ] ]]

As for the PG construction in (46c), if the computational system proceeds cyclically and movement proceeds from more to less embedded domains, it simply cannot be generated. At the derivational step represented in (49) below, where the θ-requirements of mention could license copying of an argument, which documents is already part of the adjunct island and therefore cannot undergo movement; hence, the unacceptability of the PG construction in (46c).

(49)  a. K = [vP [vP signing papers ] [island without PRO checking [ which documents ] ] ]  
b. L = mention

Notice that if the sideward movement could proceed from the matrix to the future adjunct clause, as represented in (50), the PG construction in (46c) would be incorrectly ruled in. From (50), the system could build the syntactic object in (51a), which allows movement of the copy of which documents in the matrix object position to [Spec, CP], as shown in (51b), and (46c) should be acceptable.
The intuition behind the argument reviewed above is that in order to reduce computational complexity, the computational system is required to proceed to building a new clausal constituent, say, only if it has already finished the active clausal constituent being constructed. This idea can be implemented in terms of Chomsky’s (1998) subarrays, as shown by Nunes and Uriagereka (2000). Assuming that each subarray of a given numeration is determined by one instance of a light verb or a complementizer (see Chomsky 1998), the structured numeration that underlies the derivation of (46b), for instance, is given in (52).

\[
\begin{align*}
(52) \quad N &= \{\{A, Q_1, did_1\}, \\
& \quad \{B, John_1, v_1, mention_1, without_1\} \\
& \quad \{C, C_1, T_1\}, \\
& \quad \{D, PRO_1, v_1, checking_1, which_1, documents_1\}\}
\end{align*}
\]

Assuming that a new subarray can be activated only after the elements of the currently active subarray have been integrated in the structure (see Hornstein 2001 for discussion), a convergent derivation for (52) can be obtained if the system first exhaust subarray D, building K in (53a), and then activates subarray B, allowing sideward movement to proceed from K to L, as discussed earlier.\textsuperscript{20}
(53) \[ \begin{align*} K &= [vP \text{ PRO } [v \text{ checking } [ \text{ which documents } ]] ] \\ L &= [vP \text{ mention } [ \text{ which documents } ]] \end{align*} \]

As for (46c), there are two relevant numerations to consider. In the first one, the lexical items *which* and *documents* belong the subarray containing *checking*, as shown in (54).

(54) \[ N = \{ \{ A Q_1, \text{did}_1 \}, \]
\[ \{ B \text{John}_1, v_1, \text{mention}_1, \text{after}_1 \} \]
\[ \{ C C_1, T_1 \}, \]
\[ \{ D \text{ PRO}_1, v_1, \text{signing}_1, \text{papers}_1, \text{without}_1 \} \]
\[ \{ E C_1, T_1 \}, \]
\[ \{ F \text{ PRO}_1, v_1, \text{checking}_1, \text{which}_1, \text{documents}_1 \} \}

Activation of subarrays F, E and D of (54) yields the structure in (55a), where *which documents* is within an island; hence extraction of *which documents* after the array containing *mention* is activated is not allowed and the PG construction in (46c) cannot be generated.

(55) \[ \begin{align*} \text{a. } K &= [vP [vP \text{ PRO } v \text{ signing papers } ] [ \text{island without C PRO T checking} \\ &\text{[ which documents } ]] ] \\ \text{b. } L &= \text{mention} \end{align*} \]

In the other numeration to be considered, *which* and *documents* belong to the subarray containing *mention*, as shown in (56).

(56) \[ N = \{ \{ A Q_1, \text{did}_1 \}, \]
\[ \{ B \text{John}_1, v_1, \text{mention}_1, \text{which}_1, \text{documents}_1, \text{after}_1 \} \]
In this case, no subarray can lead to a licit syntactic object before the computational system activates a new subarray. If the computation starts with subarrays A, C, or D, the only possible merger is between C and T/did; once they merge, forming CP, the derivation is doomed because the future vP complement of T/did cannot be merged cyclically. If the computation starts with subarrays B or D, the preposition cannot be integrated in the structure. Finally; if it starts with subarray F, it will not be possible to cyclically merge an object with checking after the subarray B is activated. Again, the PG construction in (46b) cannot be generated.

The reader is referred to Nunes and Uriagereka 2000 and Hornstein 2001 for further details and discussion. What is relevant for our current discussion is that cyclicity considerations impose a given directionality of sideward movement. This directionality in turn not only reduces computational complexity by activating a new subarray only after the currently active subarray has been exhausted, but it also prevents overgeneration with respect to island effects. We will see in section 6 below that some asymmetries between PGs and their licensing gaps provide interesting independent evident for the proposed directionality discussed above. But before we get to that, let us first consider some other asymmetries between PG and ATB constructions noted by Postal (1993).

5. Further Asymmetries between PG and ATB Constructions

The discussion in section 3 has shown that some of the asymmetries catalogued by Postal (1993) can be approached as puzzles that could potentially shed light on the various conditions that movement operations must satisfy, in absence of the Parallelism Requirement. This is also the case of the contrasts below (from Postal 1993).
(57)  a. *[ what color ], did they criticize e, after painting their house e,
      b. [ the color ], that they chose e, yesterday and (Mike) will paint their barn
e, tomorrow is red

(58)  a. *it was Ida who, Bob contacted e, immediately after concluding that it
      would amuse e, to tickle alligators
      b. the kind of people who, Bob might warn e, but it would nonetheless
      amuse e, to tickle alligators

The problem that these sentences present is that sideward movement of what color
from K to L in (59) and sideward movement of who from P to Q in (60), which would be
required to derive the PG constructions in (57a) and (58a), should in principle be allowed.
Since criticize and contacted could assign a θ-role, Last Resort should license the copying
of what color and who; in addition, neither K nor P are islands for extraction at the
derivational step represented in (59) and (60).

(59)  a. K = [ PRO painting their house [ what color ] ]
      b. L = criticize

(60)  a. P = [ amuse who [ PRO to tickle alligators ] ]
      b. Q = contacted

Given that Last Resort is a licensing condition on the Copy operation and that both
instances of sideward movement in (59) and (60) would be licensed if they had taken place,
it should then be the case that what color in (57a) and who in (58a) are somehow inert for
copying at the derivational step where such copying would be licensed by Last Resort. The
challenge is to find a common property that explains why they are inactive. Postal (1993)
has already observed that these elements do not behave like canonical nominal expressions:
what color in (57a) cannot be pronominalized, as shown in (61a), and the object of amuse cannot be passivized, as shown in (61b) (Postal's (30a) and (25b), respectively).

\[(61)\]
\[
\begin{align*}
a. & \text{ Blake painted his house green/*it} \\
b. & *\text{Sonia was amused by it to tickle alligators.}
\end{align*}
\]

We conjecture that the common property between the relevant potential targets for copying in (59a) and (60a) is that neither of them is involved in a structural Case relation. What color in (59a) is a secondary predicate (hence its lack of pronominalization) and is presumably associated with a default Case. In turn, given Burzio's Generalization (see Burzio 1986) and the failure of amuse to undergo passivization, it is arguably the case that who in (60a) is inherently Case-marked (see Chomsky 1986b). Assuming that an element is rendered active for purposes of A-movement by its unchecked structural Case, as proposed by Lasnik (1995) and Chomsky (1998), we now have an account for why sideward movement of what color in (59) and who in (60) does not take place; their Case-features alone are not able to activate them and make them potential targets for copying. Thus, the PG constructions in (57a) and (58a) cannot be generated.

The ATB constructions in (58b) and (59b), on the other hand, are acceptable because the required sideward movements in (62) and (63) below are legitimated by the Parallelism Requirement.\footnote{22} As seen in section 3, after the coordinating head merges with a given constituent X, it in a sense defines X as a template that the computational system must follow in the relevant respects, when it builds parallel structures. By rendering the TPs of (62a) and (63a) as the relevant templates to be followed, the coordinating heads end up activating all the constituents of the TPs; hence, the color in (65a) and who in (66a) become active despite the fact that their Case-features would not by themselves be able to make these elements active for purposes of A-movement. Once they are activated, the color and who can undergo sideward movement to enter into a thematic relation with chose in (62b) and warn in (63b), respectively, and the ATB sentences in (57b) and (58b) are generated after further computations along the lines discussed in section 2.\footnote{23}
(62) a. K = [ and [TP Mike will paint their barn [the color] tomorrow] ]
b. L = [ chose [the color] ]

(63) a. K = [ but [TP it would nonetheless amuse who PRO to tickle alligators] ]
b. L = [ warn who ]

It is worth emphasizing that the Parallelism Requirement functions as an additional enabling condition for movement, as long as no other condition is violated. Consider, for instance, a variation of the sentence in (57b) given in (64).

(64) *[the color], that they chose e, yesterday and (Mike) got unsatisfied after painting his barn e, last week is red

Despite the fact that the coordinating head defines the TP in (65a) as the relevant template, the color is within an island and cannot be extracted out of it. Thus, the ATB construction in (64) cannot be generated, even though it would comply with the Parallelism Constraint.

(65) a. K = [ and [TP Mike [ [ got unsatisfied ] [island after painting their barn [the color] last week] ] ] ]
b. L = chose

Whether or not the analysis sketched above can be extended to the other instances of asymmetry between PG and ATB constructions listed by Postal (1993) remains to be determined. As mentioned above, rather than committing ourselves to specific proposals regarding the unacceptability of certain kinds of PG constructions, our main goal here is to provide general independent grounds to account for the acceptability of analogous ATB constructions. Our aim has been to show that given the way that the Minimalist Program understands the relation between last resort movement and convergence, a treatment of the
Parallelism Requirement as a bare output condition allows us to treat ATB gaps and PGs in a uniform manner and still allows room for the observed asymmetries between the two construction types.²⁴

6. Accessibility to the Computational and Directionality of Sideward Movement

Combining the approach sketched in section 5 with the proposal that sideward movement proceeds from more to less embedded domains (see section 4.2), we may now account for some interesting asymmetries between PGs and licensing gaps. Consider the sentences in (66) and (67), for instance.²⁵

(66)  a. this is the woman who, Frank bothered e, with his marital problems after meeting e, only once
       b. *it was Graham, that everyone who began to bother e, with their marital problems ended up offending e, 
       c. *it was Lucy who, he insulted e, after bothering e, with his marital problems

(67)  a. this is the book which, the students who were supposed to read e, thought they would be given e, (by Ted)
       b. *it was that book which, everyone who was given e, by Ted refused to read e, 
       c. *it was that book which, I read e, before being given e, by Ted

According to the cyclic approach regarding directionality of movement that we are assuming, the sideward movement required to derive the acceptable sentences in (66a) and (67a) should proceed from K to L in (68) and from M to N in (69):

(68)  a. K = [ meeting who only once ]
b. L = [ bothered who ]

(69)  
a. M = [ who were supposed to read which ]
b. N = [ given which ]

These instances of sideward movement in (68) and (69) are licit because (i) the moved elements in (68a) and (69a) are active for purposes of A-movement thanks to their structural Case-features; (ii) movement establishes a thematic relation, satisfying Last Resort; and (iii) K and M are not adjunct islands at these derivational steps. Hence, the acceptability of the PG constructions in (66a) and (67a).

By contrast, the other PG constructions in (66b-c) and (67b-c), which would require the sideward movement of the elements in bold in (70)-(73) from K to L, cannot be generated.

(70)  
a. K = [ PRO bother Graham with their marital problems ]
b. L = offending

(71)  
a. K = [ PRO bothering who with his marital problems ]
b. L = insulted

(72)  
a. K = [ was given who which by Ted ]
b. L = read

(73)  
a. K = [ being given PRO which by Ted ]
b. L = read

Given that the object of bother in these constructions and the theme of give in double object constructions cannot be passivized (see Postal 1993:739, 741), it is reasonable to assume that they are inherently Case-marked. If this is so, the potential targets for Copying in the
syntactic object K in (70)-(73) are not active for purposes of A-movement (see section 5), and the relevant sideward movements do not happen despite the fact that they would be licensed by Last Resort if they had taken place.

As we should expect from the discussion in section 5, if a coordinating head merges with structures analogous to K in (70)-(73), the Parallelism Requirement should then activate the constituents of the relevant structure, including the objects of *bother* and *give*. Hence, the acceptability of the analogous ATB constructions in (74) ((74a) is Postal's (15g); see fn. 22).

\[(74) \quad \begin{align*}
\text{a. } & \text{who}_i \text{ did Tony respect } e_i \text{ and (Arnold) constantly bother } e_i \text{ with his} \\
& \text{marital problems} \\
\text{b. } & \text{it was [ this book ], that everybody asked for } e_i \text{ but only Greg was} \\
& \text{given } e_i
\end{align*}\]

To summarize, the data above presents not only evidence for the specific directionality of sideward movement assumed here (from more to less embedded contexts), but also evidence for derivations themselves, since the accessibility of a given element to the computational system may change as the derivation unfolds.

7. Conclusion
The version of minimalism outlined in Chomsky 1995 takes derivations to be sensitive to convergence conditions. Operations are permitted only if they further the global goal of interface legibility. It is in this sense that operations are last resort. When considering ATBs together with PGs, this general structure provides some analytical options that are precluded in *GB*-style theories. Postal's observations are difficult to approach in *GB* terms, because the only real way of unifying ATBs and PGs in this framework is to treat the former as basic and the latter as involving additional restrictions. As all movement is free, the only resources available for reigning in overgeneration is additional filters. The main problem with this is that it is quite unclear what the relevant filters might be, i.e. what
filters should apply to adjuncts but not to coordinates. Minimalism affords another way of approaching the facts noted by Postal while still maintaining a unified approach to ATBs and PGs. Though there are many details left to be ironed out, we hope to have shown that the perspective afforded by taking the Parallelism Requirement as a C-I legibility condition yields a potentially interesting avenue for further exploration of these intricate phenomena.

References


Norbert Hornstein
1401 Marie Mount Hall
We are thankful to Zeljko Boškovic’, Tim Stowell, and two anonymous reviewers for comments and suggestions on an earlier version of this paper. The first author has received support from NSF (grant SBR-9817569) and the second author from CNPq (grant 300897/96-0) and FAPESP (grants 97/9180-7 and 98/05558-8).

1 The sequence of the derivational steps in (2) and (4) has also been called *interarboreal operation* by Bobaljik and Brown (1997) and *paracyclic movement* by Uriagereka (1998).

2 In the system proposed in Nunes 1995, 1999, 2001, there is actually an additional operation of chain formation and deletion of copies can only take place within a chain (an operation that is referred to as *Chain Reduction*). Hornstein (2001) develops an alternative approach according to which deletion of copies proceeds in a deterministic fashion without resorting to the notion of chain. The conceptual and technical differences between these two approaches are not relevant for the discussion that follows. For concreteness, we will assume that Delete only applies to copies that are chain links.

3 The structure in (6c) could also be linearized if the upper copy of *John* were deleted. The choice of the chain links to be deleted is actually determined by optimality
considerations. Roughly speaking, the head of the chain in general becomes the optimal link with respect to phonetic realization as it participates in more checking relations than the lower links (see Nunes 1995, 2001 and Hornstein 1998, 2001 for a detailed discussion of this issue). For presentation purposes, we will not discuss cases where the head of the chain is deleted (see Franks 1998 and Boškovic’ 2000, among others) and assume that Delete always targets lower chain links.

4 For sake of briefness, we ignore the possibility that there may be additional copies of which book in (10a), which would account for island effects within the adjunct (see Hornstein 2001). For general discussion of sideward movement and island effects, see Nunes and Uriagereka 2000 and Hornstein 2001.

5 See Brody 1995 for a discussion of this kind of "forking" chains from a representational point of view.


7 (21a-c) and (22a-c) are from Postal 1993 and (21d) is from Cinque 1990. Munn (1992) judges (22d) as "*", with which we disagree.

8 For detailed discussion, see Hornstein 1998, 1999, 2001, where it is argued that the Control module can be eliminated from the grammar if movement to θ-positions is allowed.

9 Our account of the impossibility of adjunct PGs is similar to the account of the impossibility of long distance scrambling of adjuncts in Japanese offered in Boškovic’ and Takahashi 1998 in that lack of θ-role assignment/checking prevents the relevant movement of the adjunct. If Boškovic’ and Takahashi’s analysis and our account of (21b-d) are both correct, elements that are selected but not θ-marked should also be unable to undergo long distance scrambling. Although Japanese speakers judge the contrast between short and long distance scrambling involving such constituents as somewhat subtle, the directionality of the judgements conforms with the prediction in that long distance is worse than short distance scrambling, as illustrated in (i) and (ii). Our thanks to Nobuhiro Miyoshi and Masao Ochi for judgements and discussion.
Levine, Hukari, and Calgano (2001; henceforth L, H & C) provide many examples where selected (but not θ-marked) expressions do license parasitic gaps, as illustrated in (i) and (ii).

As L, H & C observe, if it is possible to license PGs with selected material even if it is not a DP, this is problematic for approaches that treat PGs as dependent on null resumptive DPs or moved PROs. However, this possibility does not affect our proposal, which involves
sideward movement. What prevents moving an expression is whether it can be licensed when moving. What is at issue is thus whether such movement can be licensed by selection. If the examples in (i) and (ii) prove to be typical (*contra* Postal), this will in fact streamline our analysis, for we can simply assume that selection is sufficient to license movement.

We suspect that the examples that L, H & C offer indicate that Postal’s descriptive generalizations are likely incorrect. However, given that the main point of this paper is to provide a conceptual framework within which to consider ATBs and PGs as derivationally homogeneous, we proceed in the text on the assumption that Postal’s description is essentially accurate as it offers the greater challenge. What is required in this case is a distinction between selection and $\theta$-marking; the latter licensing greedy movement, the former not.

11 Notice that although it was assumed within GB that the wh-phrases in (29b-d) could not undergo long movement because they were not $\theta$-governed, it was also tacitly assumed that they were appropriately licensed at D-Structure, where the input structure was built. In current terms, that amounts to saying that selection is sufficient to license the structure building operation Merge. For the role of selection in licensing Merge, see Chomsky 1998:50-51.

12 We wish to emphasize that our account does not necessarily require the distinction between selection and $\theta$-marking with respect to the licensing of movement operations. Whether the distinction holds is an empirical matter. What we offer here is a worse case approach: if Postal is right, then we need to assume a distinction between selection and $\theta$-marking. This distinction has been argued for independently and so, we believe, is plausible (see Jackendoff 1972, for instance, for an argument to the effect that selectional restrictions should not be cashed out in syntactic terms; also see Schein 1993:sec. 5 for an additional argument for the distinction between being selected and being an argument). However, if the reported data are incorrect, as L, H & C argue, then we can revert to the simpler theory in which selection, no less than $\theta$-marking, suffices to license movement.
Chomsky (1995) adopts what is arguably a similar view of the Parallelism Requirement, following arguments in Fox 1998. See Munn 1993 and references therein for further discussion.

Recall that similar ATB constructions may also be licensed by Last Resort if L, H & C are correct in their argument that PGs with selected adverbials, prepositions, and predicate nominals are well formed (see fn. 10 and 12).

Munn (2001) has observed that PG constructions do not license a functional reading for the PG, but ATB constructions do license such a reading for a gap in the second conjunct. Thus, (ib) is an appropriate answer for the question in (ia), but not (ic), which involves a sloppy reading for the argument of the function; by contrast, both sentences in (iib) and (iic) can be appropriate answers for the ATB question (iia) (from Munn 2001:182-183).

(i)  
\begin{itemize}
  \item a. [ which poem ]i did every poet throw out e, before her agent read e,
  \item b. Every poet threw out her first poem before her agent could read it.
  \item c. #Every poet threw out her first poem before her agent could read his first poem.
\end{itemize}

(ii)  
\begin{itemize}
  \item a. which restaurant did Bill review e, on Tuesday and Fred review e, on Wednesday
  \item b. Bill reviewed that restaurant on Tuesday and Fred reviewed it on Wednesday
  \item c. Bill, reviewed his first restaurant on Tuesday and Fred, reviewed his second on Wednesday
\end{itemize}

Based on these facts, an anonymous reviewer suggests that “if functional readings are \( \theta \)-marked, then they should allow parasitic gaps. If they are not \( \theta \)-marked, then they should be allowed in the parasitic gap domain but not in the real domain, but [it is] the opposite pattern that holds.”
The contrast between (ia) and (iia) does not seem to hinge on θ-marking, but on the absence of functional readings across islands (see Aoun and Li 1993, among others). In the ATB construction in (iii)a, for instance, the introduction of an island in the second conjunct blocks the functional reading; that is, the (iiic) cannot be an appropriate answer to (iii) a.

(iii) a. ?Which restaurant did Bill review on Tuesday and Fred wonder how to review on Wednesday?
   b. Bill reviewed that restaurant on Tuesday and Fred wondered how to review it on Wednesday.
   c. #Bill\_i reviewed his\_i first restaurant on Tuesday and Fred\_k reviewed wonder how to review his\_k second on Wednesday.

This shows that the contrast between (ia) and (iia) is orthogonal to the issue of whether selection and θ-role assignment differ in licensing applications of Copy. There is no intrinsic difference between PG and ATB constructions with respect to functional readings. Once PG constructions generally involve binding across an island, functional readings will be blocked; but if ATB constructions also happen to involve binding across an island, the same pattern will arise.

16 For instance, Postal observes that, for many speakers, ECM verbs do not license PGs, but license ATB gaps, as illustrated in (i) (Postal’s (49) and (51a)):

(i) a. *[ which candidate \_i ] did the CIA hire e\_i without believing/considering/proving e\_i to be loyal
   b. *[ which candidate \_i ] did the CIA believe/consider/prove e\_i to be loyal but Melvin believe/consider/prove e\_i to be disloyal

17 The contrast becomes clearer when we add adverbial expressions modifying each of the events:
(i) Eu conversei com o João sábado e *(com) a Maria domingo.
'I talked to João on Saturday and (to) Maria on Sunday.'

A reviewer points out the sentence in (i) as a counterexample to our description of the interpretations available in (36), for the two-event reading is available despite the fact that we do not have PP coordination.

(ii) Marília Gabriela falou com o Clinton e o Bush quando eles foram eleitos.
‘Marília Gabriela spoke with Clinton and Bush when they were elected.’

We suspect that the availability of the two-event reading in (ii) as a whole is due to the semantic contribution of the adjunct clause, for the sentence in (iii) may also have a two-event reading. More importantly, if each of the events of (ii) is independently modified by the temporal clause, as illustrated in (iv), we are back to the requirement that PPs be coordinated in such a case.

(iii) Eles/o Clinton e o Bush foram eleitos presidentes.
‘They/Clinton and Bush were elected presidents.’

(iv) Marília Gabriela falou com o Clinton quando ele foi eleito governador e *(com) o Bush quando ele foi eleito presidente.
‘Marília Gabriela spoke with Clinton when he was elected governor and (with) Bush when he was elected president.’

18 The computation may proceed even more locally if lexical access is defined in terms of subarrays and not in terms of the numeration as a whole (see Chomsky 1998). For a discussion of lexical access in terms of subarrays in the derivation of PG constructions, see Nunes and Uriagereka (2000) and section 4.2 below.
Notice that sideward movement can iterate as long as it does not proceed from within an island. A PG construction like (i) (see Kayne 1984 for relevant discussion), for instance, involves sideward movement from the object of *checking* to the object of *signing* in (ii) and from the object of *signing* to the object of *mention* in (iii). In neither case is the extraction site within an island *at the derivational step where it occurs*.

(i) \[ \text{[ which documents ], did John mention e_i after signing e_i without checking e_i} \]

(ii) \[ K = \text{[ PRO checking [ which documents ]^i ]} \]
\[ L = \text{[ signing [ which documents ]^i ]} \]

(iii) \[ M = \text{[vP [vP PRO signing [ which documents ]^i ] [island after PRO checking [ which documents ]^i ]]} \]
\[ N = \text{[ mention [ which documents ]^i ]} \]

Assuming that sideward movement may proceed before the operation Agree (see Chomsky 1998) establishes feature checking, each copy will then be able to check Case in a different derivational workspace. Similar considerations apply to PGs in subject position, as illustrated in (i), which is taken to involve sideward movement from K to L in (ii), before Case checking occurs. See Nunes 1995 for additional discussion of sideward movement in languages that do not allow Case mismatch between the PG and the licensing gap.

(i) \[ \text{[ which student ], did you hire e_i after Mary said e_i impressed the boss} \]

(ii) \[ K = \text{[vP [ which student ]^i impressed the boss ]} \]
\[ L = \text{[ hire [ which student ]^i ]} \]
Based on the sentences in (61), Postal (1993) takes the contrasts in (57) and (58) as showing that PG constructions are subject to pronominalization and passivization constraints.

For presentation purposes, we are assuming a raising analysis of relative clauses in (62), to account for the sideward movement required to derive (57b). Nothing would essentially change if the element that undergoes sideward movement in (62) is a null operator, rather than the color. Similar observations hold for the derivation of the remaining relative clauses.

Another contrast noted in Munn (2001) is that amount relatives license ATB gaps but not PGs, as illustrated in (i) below (Munn’s (40) and (39a)). Given the fact that the amount element can undergo A-movement, as shown in (ii), a reviewer asks why then the amount reading is not available for the PG construction in (ib).

(i)  
   a. it would take us weeks to drink [the wine], that John drank e, and Bill spilled e (amount reading OK)
   b. it was amazing [the wine], Bill drank e, after Fred spilled e, on the floor (amount reading *)

(ii) it was amazing the wine that was drunk at the party (amount reading OK)

In order for the relevant sideward movement from K to L in (iii) below to be allowed, not only the wine (see fn. 22) should be active for the computational system, but also the θ-properties of drank should trigger its copying in order for Last Resort to be satisfied. As shown in (iv), however, amount readings are not available across a weak island, which was the traditional diagnostic we used to independently distinguish θ-marking from mere selection. In other words, being active is a necessary, but not sufficient condition for (sideward) movement to apply to a given element. Further illustration of this fact is provided by (v), which shows a measure phrase being able to undergo A-movement
for purposes of Case checking, but not sideward movement to satisfy the selectional properties of \textit{spend} (see section 3).

(iii)  
\begin{enumerate}
\item a. K = [ Fred spilled [the wine ] on the floor ]
\item b. L = drank
\end{enumerate}

(iv)  
\begin{quote}
\text{it was amazing [ the wine ], we forgot when we drank } e_i \text{ (amount reading *)}
\end{quote}

(v)  
\begin{enumerate}
\item a. Many weeks were spent in the attempt to solve the problem.
\item b. *[ how many weeks ], did he spend } e_i \text{ in Berlin without wanting to spend } e_i \text{ in London}
\end{enumerate}

At any rate, even if sideward movement from K to L in (iii) were permitted, we should not expect the amount reading to be possible, given the presence of the adjunct island in (iib). Even ATB constructions disallow the amount reading if an island intervenes, as illustrated in (vi) with the negative island.

(vi)  
\begin{quote}
\text{it would take us weeks to drink the wine that John drank and Bill didn’kt spill (amount reading *)}
\end{quote}

\footnote{The general analysis outlined here is also consistent with the fact that there are languages such as Polish (see Franks 1993), for instance, which allows ATB but not PG constructions, but to our knowledge there seems to be no language that allows PG, but not ATB constructions. We speculate that in languages like Polish, the relevant syntactic object containing the would-be PG copy is a strong phase (see Chomsky 1999) that must be spelled out before the computational system start building other phases; hence, its constituents are unavailable for copying (for purposes of A-movement) and PGs can not be derived. However, if such syntactic object is merged with a coordinating head, it becomes a template to be appropriately matched and the Parallelism Requirement may license
instances of sideward movement, yielding ATB constructions. We leave a full exploration of this conjecture for future research.

See Postal's (1993) (46a), (13e), (13f), (46b), (18c), and (18d), respectively. The expression this is was added to (66a) and (67b) to form complete sentences. On the relative clause analysis assumed for (66b), see fn. 22.