FIVE REASONS FOR BUILDING PHRASE STRUCTURES TOP-DOWN FROM LEFT TO RIGHT *

Cristiano Chesi
University of Siena

1. Introduction: Structure Building Operations and Directionality

In this paper I will summarize five arguments supporting an explicit formalization of a minimalist grammar which is *derivational* and *directional*: *derivational* since constituents and dependencies are built dynamically, piecemeal, using structure building operations such as merge and move in a phase-tailored computation; *directional* in the sense that these structure building operations operate strictly from left-to-right as proposed by Colin Phillips (1996, 2003) and top-down (Chesi 2004). This directionality issue is obviously the main difference with respect to standard minimalist approach[1].

The issues I will summarize are both formal and empirical and are tightly related to the revised definition of merge and move: first of all I will go trough Phillips' (2003) arguments, showing how merge should be redefined in a left-to-right manner, *merge-right*, based on the evidence that intermediate constituents can be targets for coordination. This will allow us to capture interesting conflicting results among constituency and c-command tests (section 3, Phillips 2003) in a cross-linguistic perspective (Choi and Yoon 2006). Two formal arguments will be presented, then, in order to introduce the directionality shift with respect to the move operation and the notion of derivation by phase: unmotivated intermediate steps for movement (section 4, Chesi 2004a) and growing complexity with respect to cross-phasal dependencies (section 6, Chesi 2004b) suggest that top-down, left-to-right derivations are interesting solutions to be explored systematically: the complexity argument, in fact, isolates a natural class of constituents which exactly correspond to strong islands; left-to-right derivations are able to provide an unified account of such constituents, as in Huang (1982), without loosing the ability to discriminate important empirical differences (section 5, Chesi 2004a). This is especially true for parasitic gaps constructions: we will be able to recast the connectedness effect (Kayne 1983) in derivational terms without stipulating any complex-NP constraint (section 7, Bianchi and Chesi 2005).

---

* I am grateful to the audience of the Joint Workshop on Linguistic Theory and Language Acquisition at Nanzan for suggestions and remarks. Special thanks go to Valentina Bianchi, Tomohiro Fujii, Keiko Murasugi, Mamoru Saito and Kensuke Takita for their comments, patience and support.

[1] In this paper I will not stress the similarity of this proposal with respect to other influential frameworks (such us Dynamic Syntax, Cann and al. 2005); notice however that this assumption is not completely original.

Nanzan Linguistics: Special Issue 3, Vol. 1, 71-105
© 2007 Cristiano Chesi

2. Rationale Behind the Arguments

As stressed by Ross (1967), a language can be thought of as a set of constraints on how we can make infinite use of finite means given a, powerful enough, recursive procedure. The main problem didn't change that much in the last 40 years: move α and the projection principle before (Chomsky 1965), then merge, move and agree now (Chomsky 2005a), are very general and (too) powerful devices that allow us to figure out very important universal principles, but which require a big effort so as to be able to define precisely the upper bound of our language set: it is a matter of fact that these operations are defined in such a general way to be practically insufficient to constrain many relevant empirical phenomena.

Within the spirit of the minimalist enterprise, I will try to show that a more restrictive definition of merge, move and the notion of phase (deeply inspired by Kayne's theory of antisimmetry), can be successfully rephrased in top-down, left-to-right terms, attaining superior results in terms of simplicity and empirical adequacy, at least with respect to a relevant subset of key phenomena, without loosing computational tractability. This solution departs from the simplest assumptions proposed within the standard minimalist approach (Chomsky 1995-2005b) that, in fact, lead to solutions which are only seemingly simpler and more essential than the ones proposed within the government and binding framework: theoretical complexity is mainly projected onto (often underspecified) lexical features, recasting potential universal properties to language specific properties (much as in transformational rule-based approach, Chomsky 1965); this way UG specification would become very complex and problematic in terms of explanatory adequacy.

3. Argument 1: Intermediate Constituency

A standard assumption within the minimalist mainstream is that the *merge* operation is the first source of recursion in natural language:

(1) merge (Y , merge (X, ... merge (C, merge (A, B))...))

A, B, C, X and Y are either lexical elements or results of other merge operations; in this second case, merge can be reasonably thought of as a recursive operation:

(1') merge(*John* , merge(*said*, ... merge(*Mary*, merge(*bought*, merge(*the*, *book*)))...))

The bottom-to-top directionality of the derivation guarantees that at any application of this structure building operation a (piece of a) constituent is built, either satisfying some selectional requirement or complying with adjunction. In any case, last merged elements will always (derivationally, in the sense of Epstein and al. 1998) C-command previously merged ones, according to the no-tampering condition (Chomsky 2005a). Since constituents and C-command relations are both built by merge (internal and/or external), results for tests involving constituency and C-command relations should not conflict as shown below (Phillips 2003):

(2) a. Gromit [likes cheese] and [hates cats]. (*coordination*)

b. Gromit [likes cheese] and Wallace does too. (*deletion/ellipsis*)

c. [Like cheese] though Gromit does, he can't stand Brie. (*movement*)

d. Wallace and Gromit like each other. (*reciprocal binding*)

e. *Each other like Wallace and Gromit. (*illicit reciprocal binding*)

This prediction seems to be correct for a relevant subset of constituents but not for all:

(3) a. Wallace gave [Gromit a biscuit] and [Shawn some cheese] for breakfast.

b. [Wallace designed] and [Gromit built] an enormous tin moon-rocket.

c. Wallace gave Gromit [a biscuit in the morning] and [some cheese just before bedtime]

d. * [Gromit a biscuit] Wallace gave for breakfast.

e. Alice [knew that Fred wanted to talk] and [hoped that he wanted to argue] with the president.

f. *Alice [knew that Fred wanted to talk] with the queen and Ethel did _ with the president.

Coordination (3.c) Vs. topicalization (3.d), for instance, would suggest that the cluster composed by the direct and the indirect object, in a double-object construction, gives rise to conflicting predictions about constituency. Coordination (3.e) conflicts also with ellipsis (3.f) if we look at the cluster [verb + embedded clause subpart].

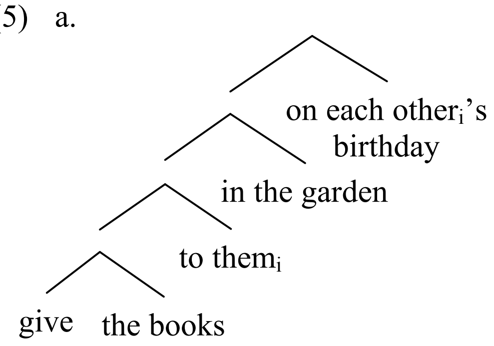
In the end, also the branching directionality of the structure is questioned looking at C-commanding vs. fronting examples:

(4) a. John wanted to give books to them in the garden, and [give the books to them_i in the garden] he did on each other_i's birthdays.

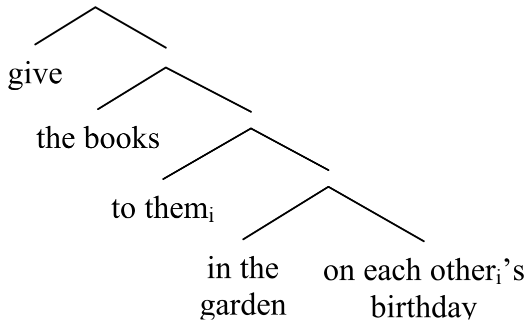
b. John wanted to give books to them, and [give the books to them_i] he did in the garden on each other_i's birthdays.

In (4), while anaphor binding would predict the standard right-branching structure (5.b), VP-fronting conspires for a left-branching structure as in (5.a), where a VP-subconstituent can be moved, leaving lower PPs stranded (Phillips 2003):

(5) a.



b.



3.1 Standard Solutions: Test-targets and Cascades/layers

In order to account for the contrasts presented above, various solutions have been provided in literature. Two main classes of proposals can be isolated: a selective target solution and a solution based on flexible constituency.

As for the first class, the crucial idea is that C-command (e.g. binding) and constituency (e.g. movement, ellipsis, coordination) tests all target specific and idiosyncratic structural properties: in a transformational based approach (e.g. Ross 1967) coordination is a special rule involving a certain kind of constituents[2] while movement apply to another cluster of constituents.

On the other hand, with respect to the second class of solutions, the central idea is that a sentence might have multiple parallel structural representation. A representative proposal can be considered the one involving layered Vs. cascade structures, suggested by Pesetsky (1995): the *layered structure* (5.a) is the locus where phrase extractions are computed, while the *cascade structure* (5.b) is the locus of *binding*, *negative polarity items* interpretation and so on.

We can conclude, with Phillips (2003:41) that "although the syntactic literature contains many excellent characterizations of the range of constituents picked out by individual structural diagnostics, there is no general theory of why particular tests yield the results that they do. As a result, discrepancies between the results of constituency tests have typically not been very informative".

3.2 Inadequacies: A Cross-linguistic Perspective on Argument Cluster Coordination

A general lack of understanding and the stipulative flavour of the solutions just presented seem to emerge as soon as we look at the data from a cross-linguistic perspective. As noticed by Koisumi (2000) and Choi & Yoon (2006), Japanese (and Korean) Vs. English, allow for specific argument cluster coordination:

[2] Or better, specially constrained variables (rather than constituents) representing (sub-)constituents (e.g. the clausal-coordination rule would involve a variable corresponding to the cluster [subject-verb] as in (3.b)).

(6) a.[[Mary-ga ringo-o 2-tu] to [Nancy-ga banana-o3-bon]] tabeta (koto).
[[M.-NOM apple-ACC 2-CL] and [N.-NOM banana-ACC 3-CL]] ate

Lit. '[Mary two apples] and [Nancy three bananas] ate.'
(Mary ate two apples, and Nancy three bananas.)

b. [[Mary-ga John-ni ringo-o 2-tu] to [Nancy-ga Bob-ni banana-o 3-bon]] ageta (koto).
[[M.-NOM J.-DAT apple-ACC 2-CL] and [N.-NOM B.-DAT ban.-ACC 3-CL]] gave

Lit. '[Mary two apples to John] and [Nancy three bananas to Bob] gave.'
(Mary gave two apples to John, and Nancy gave three bananas to Bob.)

Moreover, constituency tests in Japanese (and Korean) Vs. English do not conflict:

(7) a. Wallace gave [Gromit a biscuit] and [John some cheese] for breakfast (*coordination*)

b. *[Wallace a biscuit] and [Gromit some cheese] gave to John (*coordination*)

c. *[Gromit a biscuit] Wallace gave ____ for breakfast (*fronting*)

d. *What Wallace gave is [Gromit a biscuit] ((*pseudo-)cleft*)

(8) a. Mary-ga [[John-ni ringo-o 2-tu] to [Bob-ni banana-o 3 hon]] ageta (koto) (*coord.*)
M-Nom J-Dat apple-Acc 2-Cl and B-Dat banana-Acc 3-Cl gave (fact)

'Mary gave two apples to John and three bananas to Bob'

b. [[John-ni ringo-o 2-tu] to [Bob-ni banana-o 3-bon]] Mary-ga _ ageta (koto) (*scram.*)
J-Dat apple-Acc 2-Cl and B-Dat banana-Acc 3-Cl M-Nom gave (fact)

Lit. '(the fact that) [two apples to John and three bananas to Bob] Mary gave'

c. Mary-ga age-ta no-wa [John-ni ringo-o 3-tu] da ((*pseudo-)cleft*)
M-Nom give-Pst Nm-Top J-Dat apple-Acc 3-Cl be

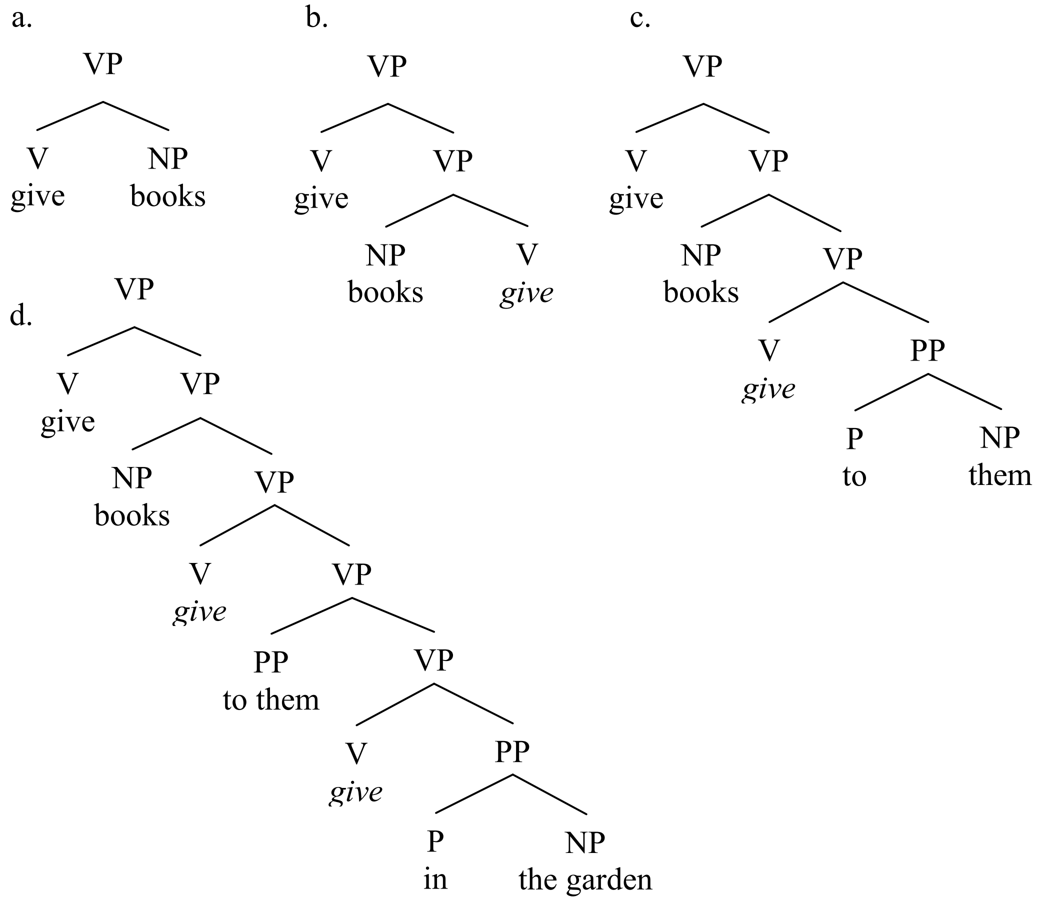
Lit. 'It is [three apples to John] that Mary gave'

This contrast is unexplained under the above mentioned theories.

3.3 A Possible Solution: Merge Right and Incremental Structure Building

Phillips proposes that a slight modification of (5.b) would be sufficient to account for any apparent contrast that seems to involve the *left-branching* structure proposed in (5.a). His solution is based on a *left-to-right* structure building procedure as illustrated in the simplified example below:

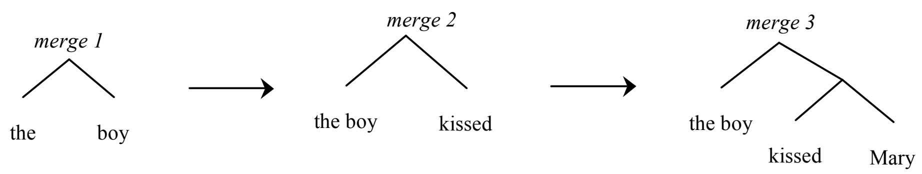
(9)



The derivation in (9) (reminiscent of Larson's 1988 *VP shells*) yields a solution to the *constituency problem*: in fact, both (9.b) and (9.c) represents *temporary constituents* (later destroyed by further merge operations) that can be targeted by specific operation at specific point of the derivation. This is the cornerstone of the *left-to-right* structure building procedure Phillips proposes: *temporary constituency* is the property, created during the derivation, we could use to account for apparent paradoxes. Crucially, a left-to-right processing procedure creates temporary constituents that are different from the ones built by standard *bottom-to-top*[3] derivation proposed by Chomsky (1995):

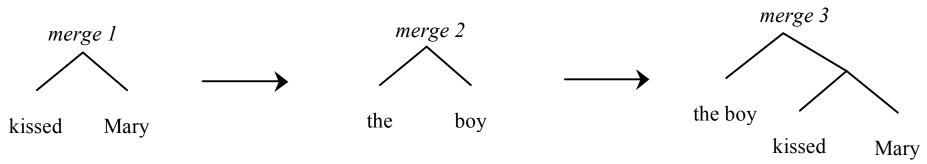
(10) "The boy kissed Mary"

a. left-to-right derivation (Phillips' proposal):



[3] That is quite different from *bottom-up*, which simply means starting from input data rather than from structure projections. On the other hand, *Bottom-to-top* means starting from the inner verbal shell, then adding higher and higher layers piecemeal.

b. bottom-up derivation (standard minimalist proposal):



For instance, only (10.a) creates the temporary constituent [the boy kissed] that can be targeted for coordination[4]. The directionality of the process could be sufficient to account for the cross-linguistic asymmetry between English on one hand and Japanese-Korean on the other: subject-object cluster never creates a temporary constituent in English while it does in Japanese-Korean[5]. Since these kinds of constituents are relevant for phenomena such as those described in (6), (7), and (8), Phillips takes this to be a proof of the fact that the grammar incorporates a simple structure-building procedure, *merge right*:

(11) **Merge right** (Phillips 1996:18)
new items must be introduced at the right edge of the structure[6]

This allows Phillips to make a series of predictions which turn out to be correct in a cross-linguistic perspective:

Prediction 1: A constituency test may refer to only those strings that are constituents at the point in the incremental derivation when the test applies (e.g. subject-object cluster coordination in Japanese-Korean Vs. subject-verb cluster coordination in English, (6) Vs. (7))

Prediction 2: Contradictions between constituency tests can only arise when those tests apply at different stages in the incremental derivation of a sentence (e.g. coordination targets intermediate constituents while movement and ellipsis only target maximal constituents, (3))

Prediction 3: Constituency changes during the course of a derivation, but most c-command relations do not. Therefore, tests involving c-command relations should not conflict with one another (2.d-e)).

---

[4] Notice that Phillips' account too has to stipulate the selective nature of the target for constituency tests: coordination can target temporary constituents; movement only targets maximal phrasal constituents.

[5] Choi and Yoon's (2006) story is more complex and requires a modification of the selection relation between the verbal complex and its arguments (we will discuss this proposal in section 5). for the purposes of this section, the directionality shift is however sufficient to account for the asymmetries between Japanese-Korean and English.

[6] This definition is not sufficient to produce the right branching structures we would need; Phillips introduces a "branch right" bias that I will implicitly adopt without discussion; this bias forces the merge operation to "select the attachment that uses the shortest path from the last item in the input to the current input item" (Phillips 1996:19).

4. Argument 2: Unmotivated Intermediate Steps in Successive Cyclic Movement

A second source of recursion in natural language is A'-movement[7]:

(12) merge (move (A), ... merge (move (A), ... merge (A, B)...)...))

A classical example is successive cyclic *wh*-movement in head-initial languages:

(12') *Who*$_i$ do you believe [t_i that Mary think [t_i that ... [t_i that everybody admires t_i]...]] ?

Independently of the specific devices used to account for displacement (probe-goal, edge features etc.) every intermediate step, necessary to escape the phase impenetrability condition (or subjacency or barrier-hood, depending on the adopted framework) in any bottom-to-top derivation, has to be triggered by some feature: in the classical minimalist understanding of *wh*-movement (Chomsky 1995), for instance, crucially movement could not blindly alter the relevant *wh*- feature that triggers the last step of the *wh*- chain. If this could happen, either we would have a non-deterministic operation that at any step could leave the element undisplaced or move it at will (leading, in any case, to a grammatical result) or else preventing the element from moving further (then stranding it in a ungrammatical position). A pertinent (recurrent) question, then, is related to what triggers intermediate steps in successive cyclic movement.

4.1 Standard Solutions: Formal/edge Features

Within the minimalist mainstream, essentially two influential solutions have been proposed in order to make successive cyclic A'-movement possible:

i. Formal Features (FFs), e.g. Probe-Goal approach described in Chomsky (2000:135)
ii. Edge Features (EFs), Chomsky 2005

The nature of the two solution differs substantially in terms of paired deletion: while FFs are both on the probe and on the goal and deletes against each other causing a freezing effect ((13.a-a'); Rizzi 2004), EFs are present only on the phase head and force internal merge as shown in (13.b-b'):

(13) a. ... [$_{+FF}$ C] everybody admires [$_{-FF\ -WH}$ who]?[8]

a'. ... [$_{\cancel{+FF}}$ [$_{\cancel{-FF}\ -WH}$ who] C] everybody admires <who> ?

b. ... [$_{EF}$ C] everybody admires [$_{WH}$ who]?

b'. ... [$_{\cancel{EF}}$ [$_{WH}$ who] C] everybody admires <who> ?

---

7 This is true at least for head-initial languages. Something has to be said for head-final languages but for sake of compactness I can not address the issue here.

8 The formalism used (+/-FF) is inspired by Stabler (1997) proposal.

FFs describe movement strictly in terms of uninterpretable-interpretable feature checking, while EFs charge the entire burden of the movement operation on the phase-head that forces internal merge to comply with scope/discourse-related "edge requirements".

4.2 Inadequacies: Finitary Nature of the Lexicon and the Non-selectivity of Edge Features

Both solutions leave some theoretical problems open: as for the first solution, the use of FFs would predict any single step to be triggered by a single (different) feature. Unless we would assume that A'-move is indeed a non-recursive operation, successive cyclicity would need an infinite number of formal features on the *wh*-element as shown below[9]:

(14) a. $[_{+WH}$ C] do you think $[_{+FF}$ C] Mary said ... $[_{+FF}$ C] everybody admired $[_{-FF\ -FF\ ....\ -FF\ \ -WH}$ who]?

a'. $[_{\cancel{+wh}}\ [_{\cancel{-WH}}$ who] C] do you think $[_{\cancel{+FF}} <_{\cancel{-FF}\ \ -WH}$ who> C] Mary said ... $[_{\cancel{+FF}} <_{\cancel{-FF}\ -FF\ ....\ -FF\ \ -WH}$ who> C] everybody admired $<_{-FF\ -FF\ ....\ -FF\ \ -WH}$ who> ?

This is in contrast with the finitary nature of the lexicon.

On the other hand (Chomsky 2005b:5,17), EF on a phase head would require movement (re-merge or internal merge) of the relevant *wh*-element to the (relevant) position in (the left periphery of) the phase edge. This would allow, in fact, a successive cyclicity without requiring an infinite number of feature on a lexical item, as shown below:

(15) a. $[_{EF}$ C] do you think $[_{EF}$ C] Mary said ... $[_{EF}$ C] everybody admired [who]?

a'. [who $_{\cancel{EF}}$ C] do you think [<who> $_{\cancel{EF}}$ C] Mary said ... [<who> $_{\cancel{EF}}$ C] everybody admired <who> ?

However a problem with this device derives from the definition of internal merge which is essentially a genuine instance of merge: it is not clear, at least to me, how EFs should force internal Vs. external merge if the element to be displaced is not marked anyway. Why should EF on C not be satisfied by an expletive or, in general, by external merge? Another (related) problem comes from the fact that we should expect a minimal variation of (15) to be grammatical, contrary to fact:

(15') a. John thinks $[_{EF}$ C] Mary said ... $[_{EF}$ C] everybody admired [who]?

a'. *John thinks [who $_{\cancel{EF}}$ C] Mary said ... [<who> $_{\cancel{EF}}$ C] everybody admired <who> ?

The intimate relation between the *wh*-element in its verbal selected position and the very same element in its scopal position (the dual semantics property of the conceptual-intentional

---

9 For the sake of simplicity I did not introduce +FF on the edge of *vP*. This would have been obviously unproblematic.

system, Chomsky 2005b:7) is lost if we do not have any featural mark-up requiring movement both on the *wh*-element and on the relevant phase edge/criterial position.

It is worth to mention a variant of the first mechanism suggested by Luigi Rizzi (2004): Rizzi proposes to eliminate the +/- distinction, marking the *wh*-element and the attracting complementizer head both with the very same feature *F*. In order to account for successive cyclicity Rizzi needs to include formal features *f* (crucially different from the one that triggers the last step in the bottom-to-top derivation and creates the freezing effect) in the sense of McCloskey (2002). Following this idea, the derivation can be rephrased as follows:

(16) a. [$_F$C] do you think [$_f$C] Mary said ... [$_f$C] everybody admired [$_F$who]?

a'. [$_{\cancel{F}}$ [$_{\cancel{F}}$ who] C] do you think [$_{\cancel{f}}$ <$_F$ who> C] Mary said ...
[$_{\cancel{f}}$ <$_F$ who> C] everybody admired <$_F$ who> ?

This solution removes three important problems: first, it removes the inconvenient asymmetry between interpretable and uninterpretable features (as Rizzi noticed, we can hardly believe that question-related criterial positions and/or *wh*-elements involve morphologically overt uninterpretable features); second, it does not violate the finitary nature of the lexicon (since the *wh*-element only bears one relevant *F* feature); third, it is compatible with the freezing effect and can circumvent the expletive-strategy (when the *wh*-element is marked with *F*, if an expletive satisfies the complementizer *F* requirement by matching, then the *F* features on the *wh*-element would stay unmatched; if we assume that *F* requires a matching in a criterial position, sentences such as (15') are correctly predicted to be ungrammatical).

There is however one last standing problem: why are formal features *f* inserted phase-by-phase in the computation? "Nothing but a teleological device so as to reach a criterial position" can hardly be considered an explanatorily adequate answer and it diminishes the advantages of having a derivation by phase.

4.3 A Possible Solution: Move as a Left-to-right Dependency Using a Memory Buffer

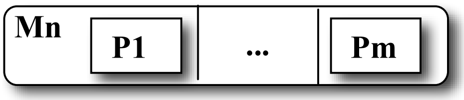
A solution for the problem of successive-cyclicity without using teleological devices, comes from parsing: memory buffered long distance dependencies.

Minimally speaking, to implement a memory-based dependency we need:

a. a definition of the memory buffer;
b. a deterministic procedure to move elements into the memory buffer;
c. a deterministic procedure to reintegrate moved elements in the structure;
d. a condition of success that allow us to determine, depending on the content of the memory buffer, whether or not the sentence is grammatical.

a. As for the structure of the memory buffer, let us assume a standard Last In First Out (LIFO) memory (used for instance in push-down automata[10]): the last element moved in the memory buffer is then the first one to be re-inserted in the structure. This would easily allow to capture nested dependencies[11]. Henceforth I will use a graphical representation (rounded boxes) to represent the memory buffers:

(17)



M stands for M(emory)-buffer and will be followed by an identification number *n* which will relate it, univocally, to a specific syntactic object (a phase in a top-down derivation). *P1... Pm* are elements stored in the memory buffer: *P1* is the first inserted element, *Pm* the last inserted one. Due to the LIFO structure of the memory buffer, *Pm* will be retrieved and reintegrated in the phrase structure before *P1*.

b. c. In order to regulate storage of an element in the memory buffer and the unloading procedure, we need to define a principle (strongly inspired to the *Linear Correspondence Axiom* proposed by Kayne 1994):

(18) **Linearization Principle (LP)**
if A *dominates* B, then either
a. <A, B> if B is a *complement* of A (that is, A selects B), or
b. <B, A> if B is a *functional projection*[12] of A

Following Grimshaw (1990), assume functional projections to be structured as extended projections (possibly with a fixed order as expected following cartographic inquiries, Belletti 2004, Cinque 2002 and Rizzi 2004b); consider then complement projections as larsonian VP-shells. This is what we would expect in terms of phrase structure:

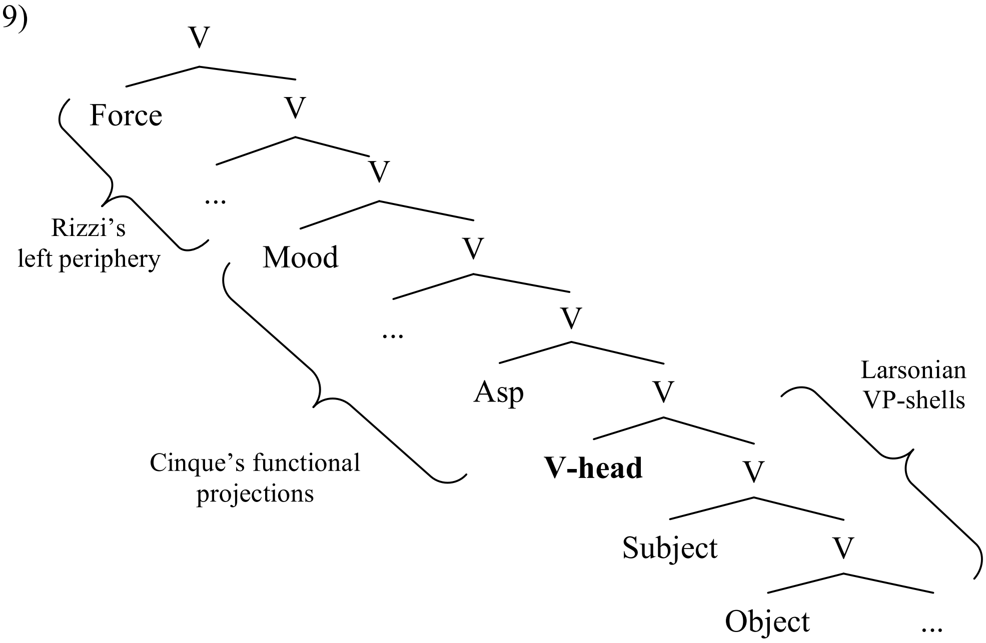
---

[10] This is not a dangerous assumption in computational terms, our grammar could in fact be no more powerful than context-free grammars. See Hopcroft and al. (1990) for a discussion of this point.

[11] Probably, this assumption will turn out to be too strong and it could be somehow weakened in order to capture some relevant empirical phenomena (cross-serial dependencies, apparent freedom in word order for topicalized elements). See Chesi 2004 for a discussion on how to capture these phenomena. Refer to Chesi 2004 also for a discussion on how this device is sufficient to capture relativized minimality effects (Rizzi 1990) using different slots in order to store different typologies of elements (e.g. A(rgumental), Mod(ifier), Q(uestion) etc..). The discussion of these aspects is irrelevant for the present discussion and will be omitted.

[12] Assume *functional projection* to subsume both the notion of *functional head* and of its *specifier*, following Starke 2004.

(19)



In this sense the verbal head projects upward and dominates (in a label-free grammar, Collins 2001) every functional projection that precedes it. On the other hand, Larsonian shells are projected in accordance with *selectional requirements* encoded on the verbal head in the lexicon[13].

According to the Linearization Principle in (18), an argument is inserted in the memory buffer if it is placed on the left of the V-head: *Wh*-elements in the left periphery, topicalized arguments and the subject in SVO languages, for instance, will be all inserted in the memory buffer and discharged in the argumental positions as soon as they can be selected[14].

d. Lastly, we need to specify a clear relation between the content of the memory buffer and the (un)grammaticality of the sentence. The most natural assumption is that the memory buffer must be empty at the end of the derivation: if not, we would have an unselected element within our structural description and that would violate the principle of *Full Interpretation* (Chomsky 1986). Then to guaranty a full interpretation we simply need to explicitly state a success condition:

(20) **Success condition**
(in order for the sentence to be grammatical) memory buffer(s) must be empty at the end of the derivation.

To summarize, a left-to-right movement operation stores in a memory buffer elements which

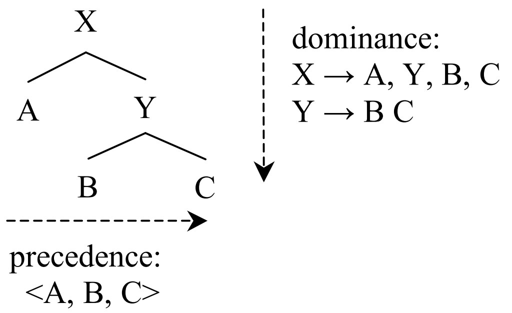
[13] Mutatis mutandis, we can assume that nominals project a parallel structure. For an attempt to derive some interesting parallelism among nominal and verbal functional domains refer to Ogawa 2001, Grohmann 2003 and Laenzlinger 2005 among others.

[14] This leads to a theory in which *move* preempts *merge*. This assumption seems to be empirically tenable as suggested by Richards (1999).

are not properly selected (according to LP) and seek to satisfy, before the end of the derivation, the required selectional dependencies according to the lexical heads that will be processed.

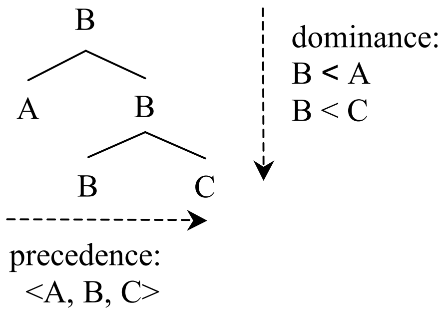
Before investigating the relevant domain where successfully seeking for a satisfaction of the selectional requirements, let me introduce a simple formalism to properly describe a *Structural Description* (henceforth SD). By default, usually, we assume that a sentence can be structurally described in terms of *dominance* and *precedence* relations as sketched below:

(21)



In a label free system, as the one I am assuming now, X and Y should be replaced by the projecting head as shown below:

(22)



We can now subsume the set of *dominance* relations in local terms by using *immediate dominance*: B *immediately dominates* A if B is the first (branching) node above A. When B will *immediately dominates* A we will write “B < A”.

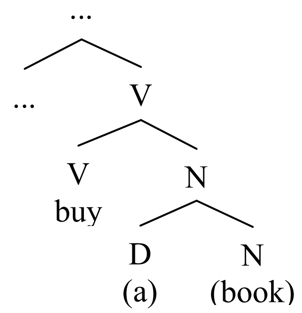
Another technical aspect to clarify is related to lexical selection: when a lexical element has to take an argument, this selectional requirement has to be encoded somehow in the lexicon. In order to avoid non-determinism[15] I assume that this selection is related to thematic selection. Functional specifications are then not properly selected, but stipulatively ordered/licensed according to cartographic inquiries. Following a fairly natural assumption, we refer to the feature that requires merge of an argument as a *select feature*. The crucial aspect of a top-down derivation hinges on these features that allow the derivation to add pieces of structure to the processed sentence. This intuition can be stated as follows:

[15] Refer to Chesi 2004 for the full discussion of this point.

(23) **Top-down projection**
a lexical head projects the minimal SD (that is the minimal set of necessary immediate dominance relations) based on its *select features*.

This implies, for instance, that as soon as we process a verbal head with a select feature requiring a determined nominal, we add to the SD the minimal set of dominance relations satisfying this requirement, i.e. V < N, N < D, or, more explicitly[16]:

(24)



The set of dominance relations introduced in the SD should be unified (in the sense of Shieber 1986) by merging (merge right, as in (11)) either an element already present in the memory buffer (first option) or another element newly introduced in the derivation (second option).

We are now in a position to introduce the last piece of theory to deal with movement from a left-to-right perspective: the notion of phase in a top-down derivation (Chesi 2004):

(25) **Notions of phase**

a. **phase** (definition)
a phase is a complete derivation involving:
- a *top-down projection* of a potential SD structure,
- a set $\boldsymbol{i_p}$ of lexical items such that any item is either the head of the phase or a functional specification of the head of the phase,
- a memory buffer M, to store unselected items and retrieve selected ones, initialized as empty, unless some items be inherited from the un-empty memory buffer of a selecting previous phase (this inherited element will be part of $\boldsymbol{i_p}$ and will be legitimated within the phase by being merged in the relevant position of the left-periphery, *edge* of the phase, as phonologically null element[17]);

b. **completeness** (definition)
a phase is complete if the head of the phase is saturated, namely if any mandatory

[16] Notice that we do not need to specify the set of precedence relation, since implied by LP. I will skip here the discussion of how to codify select feature which is now an irrelevant unwanted complication (cfr. Stabler 1997, Chesi 2004).

[17] Unless specific setup of some parameterized options.

selectional requirement (agent, patient etc.) is satisfied/projected;

c. **complementation** (definition)
a phase takes only complete phases as complements (a top-down projection can be considered a complete phase in order to close the phase generating this expectation);

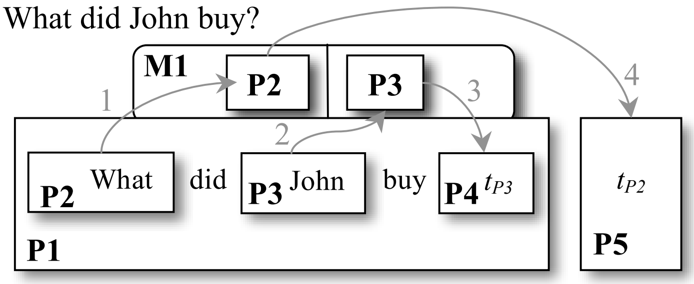
d. **phase selection/licensing requirement** (definition)
a phase has to be selected (exceptional default selection comes for the first projected phase); unselected phases can be introduced in the computation before their selection point to comply with a functional specification of the superordinate phase;

e. **memory buffer inheritance**
items in the memory buffer, at the end of the phase, can be transferred only to the memory buffer of the last selected phase.

Notice that this definition allow us to process phases either in parallel or sequentially: in the first case, a superordinate phase is still open when we process another phase (we will call this phase *nested* from now on); in the second case, we close the previous phase before processing the sequential one; this happens essentially when we project the last selected complement of a phase (hereafter this second phase will be the *selected*, or *sequential*, phase).

A minimal assumption for the system to work is to assume that phase heads are N and V[18]:

(26) What did John buy?



The derivation of the sentence in (26) proceeds as follows:

a. a default verbal phase is projected (P1)

b. since the verbal phase is interrogative, this functional feature has to be explicitly marked; in English this can be done by merging the relevant *wh*-element within the

[18] This will roughly corresponds, respectively, to DP and CP phases in a bottom-to-top derivation. As for *vP* phase, it is not completely clear to me what relevant empirical data could be captured with this phase level in a top-down grammar: notice that reconstruction effects could be predicted using VP-shells and intermediate traces on the line of Barker (2007).

specific left-peripheral position. That is how [$_N$ what], phase P2, is introduced in the derivation

c. since it is unselected (by LP) it is inserted (step 1) in the memory buffer (M1) of the matrix V-phase (P1)

d. *did* is compatible with a tense functional specification of the matrix V-phase, then licensed in this position[19]

e. [$_N$ John] (phase P3) is introduced to satisfy a subject-criterial requirement (functional specification of P1) and moved in the memory buffer since unselected (step 2)

f. then [$_V$ buy] is processed as the head of the matrix V-phase (P1). Since it has two selection requirements to be satisfied (an agent and a patient, both N-phases), these select features will project two phases, P4 e P5.

g. P4 is a nested phase and will be unified by merging P3 (the first accessible element in M1);

h. P5 is the last phase projection then is a selected/sequential phase which will inherit the content of the memory buffer of M1, eventually satisfying the last selectional requirement of the previous phase by merging P2.

I would like to stress that, despite the similarity with a parsing algorithm, this is not a parsing strategy: the elements entering the derivation can be thought of as selected using “standard” (sub)numeration(s) device. Full Interpretation and the non-ambiguity of the cartographic structural positions force left-to-right linearization in a deterministic way. From a chain-driven perspective, this solution simply shifts the focus from the standard tail-position (selected position, always lexically derivable), to the head-position of a chain (scope position, dynamically determined by the speaker in a multiple clausal sentential context). The shift is necessary in order to make the movement operation deterministic without any look-ahead device for intermediate steps.[20]

We are ready now to account for successive cyclicity: this definition of movement predicts that an element (e.g. a N(ominal) phase) can be discharged only in a selected/ sequential phase (i.e. post-head, e.g. after a V(erbal) head). Suppose then that the next projected phase is not a N-phase, as expected, but a V-phase (e.g. *believe* can project a V-phase): the N-element in the memory buffer will be however discharged in the new V-phase

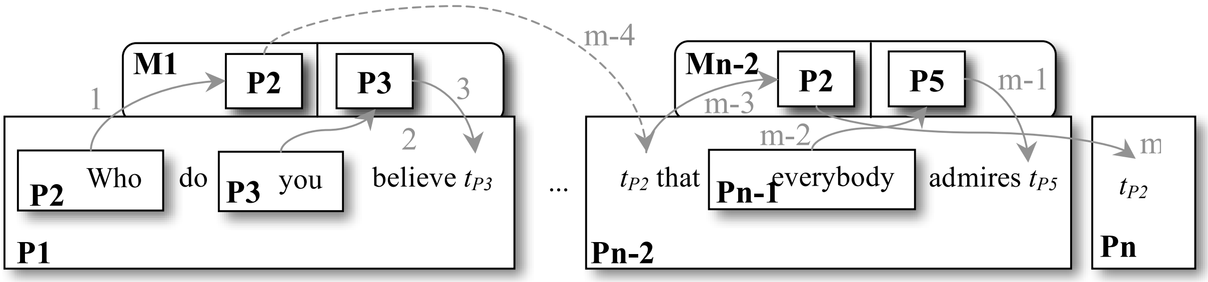
[19] We could easily implement an aux-to-C movement in interrogatives, but this would uselessly complicate the derivation.

[20] This is because the scope in A'-movement is always marked solely by the surface position of the A'-moved element. The selected position is indeed deterministically identifiable given the definition of phase and the inheritance mechanism of the memory buffer in (25). Covert movements (e.g. Quantifier Raising) are a different matter, however compatible with the proposed model (Bianchi and Chesi, 2007).

and, since it will be not properly selected (the phase projection added the minimal set of dominance relation so as to expect a V-phase, not a N-phase), it will be reloaded within the memory buffer of the just projected V-phase. And so on, up to a compatible selected position, if any, in a successive cyclic way as sketched below:

(27) Who do you believe that Mary said [that ...] that everybody admires?
[$_{CP}$ [$_{DP}$Who]$_i$ do you believe ... [$_{CP}$ t_i [that everybody admires t_i]]?

(28)



The algorithm initializes a V-phase (the CP in (27), P1 in (28)). Then it computes the *wh*-phrase, which constitutes a separate N-phase (P2). Since this *wh*-phrase is not selected, it is stored in the local memory buffer (M1) of P1 by Move (step 1). Then, the computation of P1 proceeds integrating the auxiliary as a tense specification and it stores the nested N-phase P3 which will be retrieved as soon as the head of the matrix V-phase P1, *believe*, is processed. At this point (step 3) P3 is discharged in the subject position and P1 is closed since we reached the last selected complement. The *wh*-phrase (P2) in the memory buffer M1 is then discharged in the memory buffer of the selected-sequential phase (this is in order to comply with the success condition). This memory buffer transmission minimally employs *step m-4* (re-merge the content of the memory buffer of P1 in the left periphery of the complement CP, *Pn-2*) then, since this position is unselected, the wh-phrase is re-stored in the local memory buffer of Pn-2 (*step m-3*). As a result, this "inheritance" mechanism leaves an intermediate copy/trace in the edge of the complement CP phase.[21] This way the computation can proceeds cyclically, phase-by-phase through selected phases, without using any formal/edge/teleological feature. Eventually the *wh*-phrase P2 will be discharged in a object position of a verb (e.g. phase Pn, selected by the V head *admires*, phase Pn-2) where it will be discharged from the local memory buffer of P4 and re-merged (step m): the Success Condition is thus satisfied at the end of the computation.

5. Argument 3: Nested Phases are (Strong) Islands

In SVO languages, sentential subjects of (di)transitive verbs behave like (strong, in the

[21] As pointed out by Luigi Rizzi, p.c., although this assumption is not strictly necessary for the algorithm to work, it seems fairly natural and it allows us to capture various successive cyclicity effects, like e.g. Irish complementizer alternations (McCloskey 2002).

sense of Cinque 1990) islands (29); on the other hand, subjects of unacussatives and passives (30) seem to allow for smoother extractions (Chomsky 2005b):

(29) a. *I wonder who [$_{IP}$ [books of *t*] are on the table]

b. *I wonder what [$_{IP}$ [reading *t*] would be boring]

(30) a. [Of which car] was [the driver t_i] awarded a prize?

b. It was the CAR (and not the TRUCK) of which$_i$ [DP the driver t_i] was found.

On a par with the first class of subject, "true" adjuncts too, (31) Vs. (32), are considered to be strong islands for extraction:

(31) a. *Which concert did you sleep [during t]

b. *How did you leave [before fixing the car t]

(32) a. What did John arrive [whistling t]? (Borgonovo and Neeleman 2000:200)

b. What did John drive Mary crazy [trying to _x t]?

Looking at cross-linguistic variation, adjuncts (33.a) but not subjects (33.b)[22] behaves like islands in head-final languages such as Japanese (Saito & Fukui 1998):

(33) a. ??Nani-o$_i$ [John-ga [$_{PP}$ Mary-ga *ti* katta kara] okotteru] no.
what-ACC John-NOM Mary-NOM bought since angry Q

'What$_i$, John is angry [because Mary bought t_i].'

b. ?Nani-o$_i$ [John-ga [$_{NP}$ [$_{IP}$ Mary-ga t_i katta] koto]-o mondai-ni siteru] no.
what-ACC John-NOM Mary-NOM bought fact-ACC problem-into making Q

'What$_i$, John is making an issue out of [the fact that Mary bought t_i].'

c. ?Nani-o$_i$ [John-ga [$_{CP}$ [$_{NP}$ [$_{IP}$ Mary-ga t_i katta] koto]-ga mondai-da to] omotteru] no.
what-ACC John-NOM Mary-NOM bought fact-NOM problem-is that think Q

'What$_i$, John thinks that [the fact that Mary bought *ti*] is a problem.'

[22] The sentence is degraded since it involve an extraction from a complex NP, but crucially does not contrast with (33.c) (Saito & Fukui 1998).

5.1 Standard Solutions: Condition on Extraction Domains

The classical solution to account for the islandhood of both subjects and adjuncts has been proposed by Huang (1982):

(34) **Condition on Extraction Domains (CED)**
extraction is only possible out of phrases which are *properly governed*, where a node A is taken to properly govern a node B iff
i. A c-commands B and no major category boundary intervenes between A and B,
ii. B is contained within a maximal projection of A, and
iii. B is assigned its thematic role by A

This condition essentially predicts extractions from objects.

Another interesting solution, from a derivational minimalist perspective, is proposed by Uriagereka (1999): he suggests that cyclic spell-out could explain islandhood if we take into account economy conditions in terms of linearization requirement: in fact it could be simpler for the linearization system (i.e. some implementation of Kayne's LCA) to apply to smaller chunks rather than to the whole set of lexical elements in the structure. In this sense, subjects and adjuncts can be shipped to spell-out as independent workspace within which linearization can target only the relevant subset of elements. The matrix sentence and its object(s?) are indeed linearized/spelt-out at once since linearized within the same workspace. Assuming that elements from independent derivational workspaces cannot freely enter the linearization of other workspaces, this proposal elegantly accounts for islandhood effects.

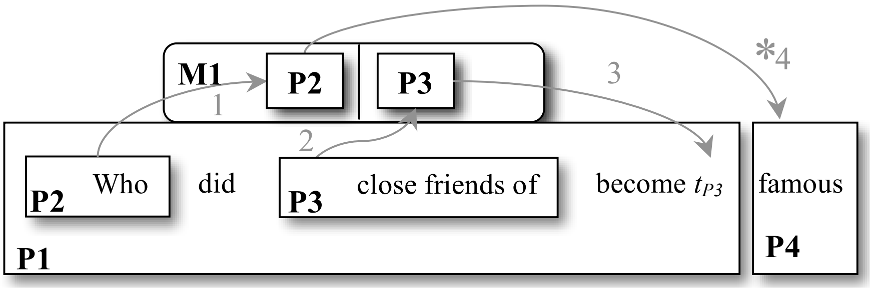
5.2 Inadequacies: Intermediate Status of Adjuncts and Cross-linguistic Variation

Needless to say, both solutions require non trivial modifications so as to accommodate the contrast (29) Vs. (30) and (31) Vs. (32). Also the cross-linguistic variation pointed out in (33) seems to be unexplained. More precisely, we could force the application of CED at specific derivational point in order to allow for extraction from passives/unaccusatives (then moving this condition from a representational to a derivational framework) so as to accommodate (29) Vs. (30); moreover we could attribute the transparent status of some "lower" adjuncts (such as *without*- or *after*-clauses) Vs. "higher" adjuncts (*because*- clauses) to the fact that these adjuncts have to be somehow thematically related to the predicated event structure (on the line of Borgonovo and Neeleman 2000 and Truswell 2006). Japanese data is however unaccounted for. The same is true for a multiple spell-out approach (where in fact it would be more problematic to explain why the subjects of unaccusatives/passives should be linearized within the matrix clause and not be independently shipped out to distinct derivational workspaces as other subjects).

5.3 A Possible Solution: Islands are Nested Phases

The top-down perspective on the ungrammaticality of strong islands is radically different from the standard view: as shown in (35), the ungrammaticality of this sentence, repeated below, does not come out from the impossibility of extracting an argument from the island, but from the impossibility of integrating the unselected element within the phase in which it has been introduced or within another phase, sequential with respect to the originating one:

(35) *Who did [close friends of _] become famous?



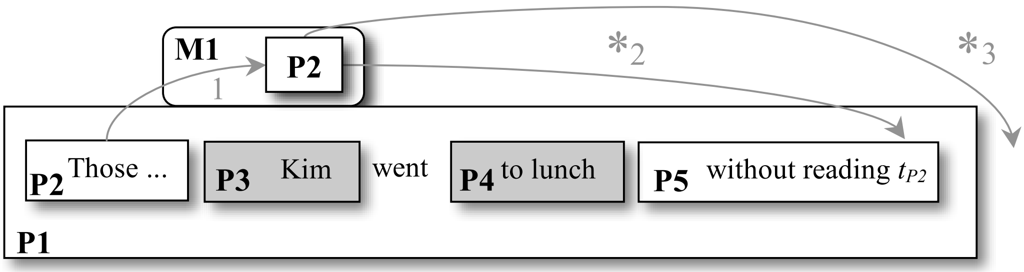
While the islandhood of subjects of (di)transitive verb is directly derivable, subject of passives/unaccusatives permeability needs a very strict application of the memory buffer inheritance mechanism; the definition is repeated below:

(25) e. **memory buffer inheritance**
items in the memory buffer, at the end of the phase, can be transferred only to the memory buffer of the last selected phase.

In this sense we need, first, to discharge the preverbal subject in the very thematic position projected by the unaccusative/passive phase-head, then, since last selected argument, discharge within this fulfilled projection the other element inserted in the memory buffer. Unless the nominal phase have no thematic requirement fulfilled by "external merge", the sentence will be predicted to be grammatical.

The contrast (31) Vs. (32) is indeed more problematic: first observe that the LP in (18) does not allow an adjunct to be to the right to the related lexical head for which it is a functional specification[23]. Despite the fact that we need to weaken the LP so as to capture the possibility to have adjuncts to the right, we should observe that adjuncts, since attached to a "functional specification", in the sense of (18)-(19), have to be considered genuine nested phases. This intuition is depicted below:

(36) ??[Those boring old reports]$_i$, Kim went to lunch [without reading e_i].



[23] This is a standard prediction in any theory that assume some version of Kayne's LCA.

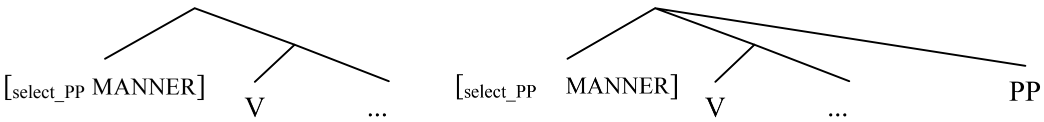
Ignoring the behaviour of phases P3 and P4, the relevant part of the derivation is that P2 is stored in the memory buffer since unselected (*step 1*) but then it can be neither discharged in P5 (*step 2*) since P5 is nested/unselected, nor in a sequential phase, since *went* does not project any other phase.

In order to explain why a functional-related constituent is linearized, against LP, to the right of the relevant modifying head, I would like to sketch a notion of heaviness:

(37) **Heaviness**
when a constituent licensed in a functional position bears more features than the functional licensed ones (then it is a (nested) complex phase), it would rather be linearized in a phase-peripheral position

Obviously the intuitive motivation of this definition (reducing complexity, by marginalizing nesting) has to be systematically explored; especially whether this is a preference or a constraint has to be empirically determined; however notice that some important constraints (e.g. right-roof constraint) would be readily predictable by (37). Tentatively, as in Bianchi and Chesi (2005), I will assume that the "complex" structure of the functional phase-head includes some sort of select feature (parallel to the one on the phase head, e.g. a *manner* phase-head as in (39.a) selects a PP as an argument) and, coherently with top-down projection definition (23), this feature introduces the needed SD at the end of the matrix phase (keeping however its nested relationship with this superordinate phase as in (39.b)[24]:

(39) a. b.



Since [manner] is a functional specification of the verbal phase, it has to be computed while this superordinate phase is still open (then it is a genuinely nested phase as expected). Only the special (complex) structure of this functional projection should be responsible for the right-hand position of the selected PP.

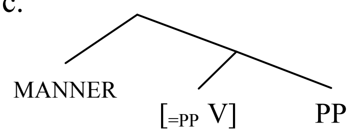
Note that the right-hand modifier is *not* selected by the lexical head (unlike Larson 1988, 1990)[25] hence, it is not a sequential phase but a nested one. This idea would allow us to capture the unexpected behaviour of some right-hand modifiers which seem to be transparent for extraction (e.g. (32)) by assuming a minimal difference between (39.b), where the island

[24] This is necessary in order to prevents the matrix phase from being interrupted. Then we need to assume that the top-down projection of this functional specification will follow the top-down projection of the head of the matrix clause.

[25] In this way the modifier is structurally superior to the VP-internal constituents; this avoids a number of problems with a generalized Larsonian "adjunct as complement" analysis (see Bianchi 2001 for discussion).

PP is "selected" by the functional specification, and (39.c), where the PP is a sequential phase because the select feature is specified on the verbal head:[26]

(39) c.



In the end we need to account for cross-linguistic variation (33): an interesting unconventional solution is suggested by Choi & Yoon, 2006 (as mentioned in section 3.2) and it challenges directly the head directionality parameter. This proposal can be simply stated as follows: while in English predicates selects their argument (a *P(redicate)-centered* language), in Japanese arguments selects their predicates (*A(rgument)-centered* language). The idea of inverse selection adopted is very powerful and potentially very restrictive and I do not have enough space here to explore many potential implications. Let me simply notice that to accommodate this parameterization within our model we simply need to allow nominal phases, by means of their case-marking, to project intersectively their top-down requirements: that is, a nominative and an accusative nominal phase would conspire so as to project a relevant verbal phase (or better, the minimal set of verbal shells constituting a sequential-selected verbal phase). In this sense, case-marked nominals are not nested phases then can be landing site for unselected phases. On the other hand adjuncts do not select any verbal phase, but simply license a functional specification of a verbal-phase (in accordance with LP, they lay on the left of the verbal-head). Then we expect adjuncts but not case-marked arguments to behave like nested phases, or if you prefer, strong islands.

6. Argument 4: Growing Complexity in Nested Phases

Before exploring another advantage of thinking to strong islands in terms of nested phases from a top-down, left-to-right derivational perspective, it is worth pointing out that the distinction nested/selected phase is interesting also from a computational point of view. Phases are introduced in the linguistic theory to reduce complexity (Chomsky 2001). More precisely, chunking the derivation by phases reduces the space of most problems by reducing the number of states to be explored. This can be shown if we take a simple combinatorial problem such as lexical ambiguity: there are lexical items, in any natural language, which are ambiguous among various syntactic classes (e.g. "run" can be both a verb and a noun); assume we need to process n ambiguous items, all of them ambiguous among a certain number c of categories. A brute force algorithm would evaluate c^n possibilities. Now assume

[26] The idea that a selectional specification for a manner PP can be associated directly to a lexical head is made plausible by the existence of "selected adjuncts" (cf. Rizzi 1990): e.g., a verb like *behave* requires a manner specification; a verb like *weight* requires a measure specification of a certain kind; a verb like *be born* requires a locative or temporal specification, etc.

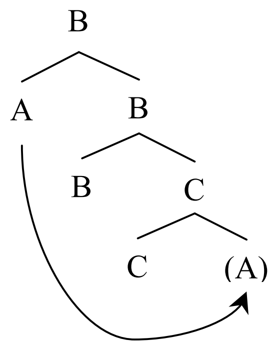
that phases are the natural domain within which ambiguities resolution takes place[27]; if we would have 8 elements distributed in 2 phases (4 per phase), to be disambiguated among 3 target categories, using or not using phases make a big difference as shown in the following table (162 possibilities, using phases, versus 6561 possibilities without phases):

(38)

	without using phases	using phases
space of the problem	c^n	$c^{n\_ph1} + c^{n\_ph2}$
e.g. 8 ambiguous elements, two phases (4 elements per phase), 3 categories	3^8	$3^4 + 3^4$ (=81 + 81)
number of possibilities to be explore to disambiguate the example sentence	**6561**	**162**

The same is true if we consider the ambiguity generated by movement: minimally speaking any movement operation introduces at least an extra immediate dominance relation (i.e. "C < A" in the example below):

(39)



Combinatorially speaking, exploring which dominance relations have to be associated to a given set of precedence relations has, at least, the complexity order of $\mathbf{O(2^{n-1})}$, where n is the length of the totally ordered sequence (namely the number of words in the processed sentence): this is because among n items we should define n-1 relations, at best (the minimum number of relations that would make a tree of n leafs, fully connected) and any of these relations could be ambiguous about the projecting head (e.g. A<B or B<A). Complexity increases, if we consider simple movements (even without considering successive cyclic movements the complexity grows significantly): at worse any item could have been potentially moved from/to any lower (licensed/selected) position, the complexity order of the problem increases up to $\mathbf{O(2^{(n^2-n)/2})}$: this is because, potentially, any element could establish a dominance relation with any other element that follows it: e.g. with 4 elements {A, B, C, D}

[27] Given the definition of phase as extended projection, this is a reasonable assumption: functional elements (such as determiners or adverbials) often are unambiguous, then they can help disambiguating other elements within the same phase. However, even if the domain of disambiguation would turn out to be bigger (2-3 phases, for instance) the structure of the argument presented in this section would be still valid.

we could have 6 possible dominance relations {A-B, A-C, A-D, B-C, B-D, C-D}; with 5 elements {A, B, C, D, E} we could have 10 possible dominance relations {A-B, A-C, A-D, A-E, B-C, B-D, B-E, C-D, C-E, D-E} and so on. Then (((n-1)+1)*(n-1))/2. This is clearly not a satisfactory result, since the growing rate of any of these functions would make the problem quickly intractable.[28]

Interestingly enough, intractability can be tackled by adopting the phase idea. We have assumed, following the definition I gave in section 4.3, that movement is triggered by the presence of an element that is not (licensed/)selected; in this case, this element will be stored in "memory" up to the point where another element selects it. Following Chomsky (2001), let us assume that if movement happens, it has to happen within the phase: then given a limited context corresponding to the phase boundaries, either we find the selecting head within it, so we can connect the moved object to its base position (projecting a set of dominance relations, according to Phase Projection, (25), that minimally satisfy this selection necessity), or we do not find any selecting element, then the expectations to find a selecting head for the unselected moved element will be projected onto the lower (selected) phase[29]. And so on.

Properties of intermediate positions[30] strengthen the idea that traces on the *edge* of the phase serve as "memory refresh" (*phase balance*, Felser 2001), that is, the undischarged elements within the previous memory buffer are allowed to enter the "phase numeration" (then being part of the next selected phase) in order to make (certain kind of) Long Distance Dependencies possible. This mechanism produces a remarkable reduction of complexity: for any moved element, we would have only two possible "landing site" positions within a phase to be evaluated (one *left-edge* position, used as a memory refresh or one *selected* position) [31]. Then for any phase the number of dominance relations required would be, at worst, $\mathbf{2^{2k}}$ (in case anything would have been moved to the left periphery of the phase[32]). The complexity

---

[28] We do not need to argue that complexity, in fact, is even worse than that since unpronounced elements can enter dominance relations (possibly without being linearized) and there is no way to determine how many empty elements could be present, in principle, in a sentence of length *n*.

[29] Leaving an intermediate trace on the "left-periphery" of this lower phase, as discussed in section 4.3.

[30] They are not *selected* positions; they are available for binding, reconstructions effects; they are not triggered by any apparent satisfaction of semantic requirements, they are in fact accounted for, in a bottom-to-top perspective, by purely formal features (cf. section 4).

[31] It has been often assumed that moved constituents stop by intermediate positions which are different from the "base position" and from the "edge" of each phase (see Sportiche 1988 for a discussion on floating quantifiers). Notice that despite the fact that in order to account for these positions we need to consider a weakening of some stronger assumptions about movement as expressed in 4.3, these considerations do not affect the argument in this section.

[32] **k** is the number of functional features, plus **1** lexical head, minus **1**, since for linking *n* elements we would need *n-1* relations.

order of the problem, considering any dominance as ambiguous, would be $\mathbf{O(2^{2k})}$. Opening a phase before having closed the previous one, again, leads to a duplication of the number of possible dominance relations and so on as long as new phases are opened. The real order of the problem is then $\mathbf{O(2^{p2k})}$ with *p* representing the number of open phases at the same time. The relation between the length of the input (*n*) and this function is expressed in terms of phases, since any phase represents a chunk of this input; in particular, the number of lexical items in *n* would determine the number of phases (which could be *n*, at worst; *k* can be arbitrarily fixed).

Functions:	*N° of relations to be evaluated* *n=6* *(p=2, k=2)*	*n=9* *(p=3, k=2)*
Brute force $2^{\frac{(n^2-n)}{2}}$	≃ 32K	≃ 68.000M
Nested Phases 2^{p2k}	256	4096
Selected Phases $p \cdot 2^{2k}$	32	64

Note that long distance dependencies within a phase would not produce any remarkable effect on the complexity of the problem, while discontinuous constituency relations among phases would increase the complexity of the problem in a linear way if any phase is closed before the next one starts (selected phases). On the other hand, there is an exponential increase of complexity any time we open a phase before having closed the previous one (nested phases). This expected increase of the complexity, related to the movement of elements across open phases could be linked, potentially, to the ban on memory discharge within strong islands (as shown in the previous section). Then "extraction from a strong island" or, from this perspective, movement within nested phases, could also receive a principled explanation in terms of economy conditions.

7. Argument 5: Parasitic Gaps

The last piece of evidence supporting a top-down, from-left-to-right derivational view of the structures building operations comes from parasitic gaps constructions:

(40) a. * [Which famous playwright]$_i$ did [close friends of e_i] become famous?

b. $^?$ [Which famous playwright]$_i$ did [close friends of e_i] admire e_i ?

(41) a. * Who did [my talking to e_i] bother Hilary?

b. $^{\surd}$ Who did [my talking to e_i] bother e_i?

(42) a. * Who$_i$ did you consider [friends of e_i] angry at Sandy?

b. $^{\surd}$ Who$_i$ did you consider[friends of e_i] angry at e_i ?

Violations of extraction from preverbal subject islands (40.a) gerundive subject islands (41.a), and small clause subject islands (42.a) can be rescued by a legitimate gap on a right branch (40.b), (41.b), and (42.b). This data shows that extraction from the left-branch subject is possible under certain configurations. Parallel examples come with adjuncts, as shown by the contrast below:

(43) a. $^{??}$Those boring old reports, Kim went to lunch [without reading e_i].

b. $^{\surd}$Those boring old reports, Kim filed e_i [without reading e_i].

Interestingly enough, not all adjuncts behaves as islands as noticed in section 5 (following examples are from Pollard and Sag, 1994):

(44) a. Who did you go to Girona [in order to meet *e*]?

b. This is the blanket that Rebecca refuses to sleep [without *e*].

c. How many of the book reports did the teacher smile [after reading *e*]?

Moreover, bleeding complex NP island constraint seems to be possible, but only when the complex NP is within the subject:

(45) a. *A person who [people [$_{CP}$ that talk to e_i]] usually have money in mind

b. $^{?}$A person who [people [$_{CP}$ that talk to e_i]] usually end up fascinated with e_i

(46) a. ?*[Which book]$_i$ did John meet [a child [$_{CP}$ who read e_i]]

b. * [Which book]$_i$ did John gave e_i [to a child [$_{CP}$ who read e_i]]]

7.1 Standard Solutions: Connectedness and Sideward Movement

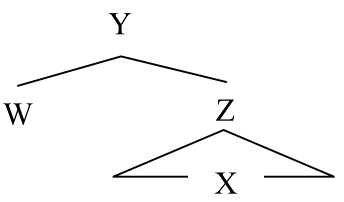
A representational solution for parasitic gaps constructions is due to Kayne (1983) and it is know as the *Connectedness Condition*. Roughly speaking, two nodes (e.g. the moved object and its trace) are connected when there is a continuous path between them formed by a set of nodes that respect a specific configuration. More precisely, this set of nodes forming the path is called a *g-projection*. The g-projection is defined recursively so as to be able to include canonically governed nodes upward in the tree as described in (47)-(48):

(47) W and Z (Z a maximal projection, and W and Z immediately dominated by some Y) are in a *canonical government configuration* iff
a. V governs NP to its right in the grammar of the language and W precedes Z
b. V governs NP to its left in the grammar of the language and Z precedes W

(48) Y is a *g-projection* of X iff

i. Y is an (X') projection of X or of a g-projection of X, or

ii. X is a structural governor and Y immediately dominates W and Z, where Z is a maximal projection of a g-projection of X, and W and Z are in a canonical government configuration



Then in a VO language like English, every right branch is in a canonical government configuration, by (48).

The *Connectedness Condition* (henceforth, CC) requires that the set of the g-projections of (the governor(s) of) the empty category(ies) bound by a given binder and the binder itself form a connected subtree:

(49) The g-projection set G_β of a category β is defined as follows (where γ governs β):

a. $\forall\pi$, π = a g-projection of $\gamma \rightarrow \pi \in G_\beta$

b. $\beta \in G_\beta$ and

b'. δ dominates β and δ does not dominate $\gamma \rightarrow \delta \in G_\beta$[33]

(50) Connectedness Condition (CC)

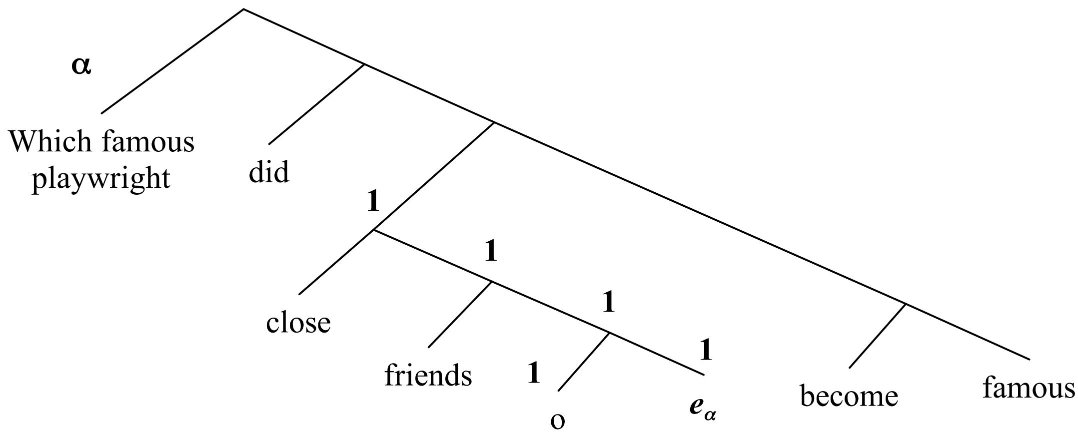
Let $\beta_1 \ldots \beta_j, \beta_{j+1} \ldots \beta_n$ be a maximal set of empty categories in a tree T such that $\exists\alpha\forall j$, β_j is locally bound by α. Then $\{\alpha\} \cup (\bigcup_{1=j=n} G_{\beta j})$ must constitute a subtree of T.

In case of a single gap, the CC requires that all the maximal projections in the path between the gap and its binder be on a right branch. Consider for instance the ungrammatical example in (1a): as the tree graph below makes clear, the g-projections of the gap stop at the level of the preverbal subject, which is a left branch and hence not in a canonical government configuration. Therefore, the g-projections cannot extend upward to reach the binder, and the CC is violated:[34]

[33] Clause b' takes care of government across a clausal boundary δ.

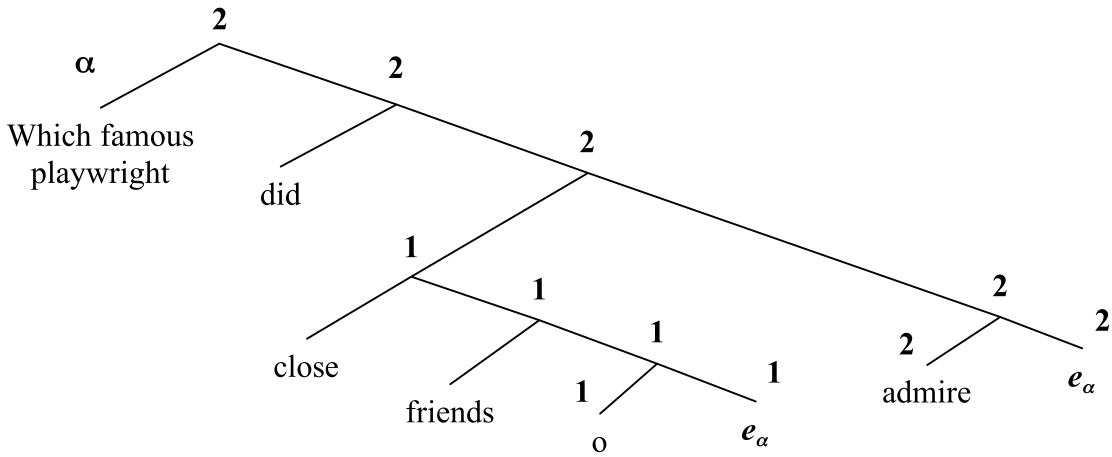
[34] As in Kayne (1983), the numerical indices are introduced for expository purposes to mark the g-projection paths of the empty categories, and have no theoretical significance.

(40.a) *



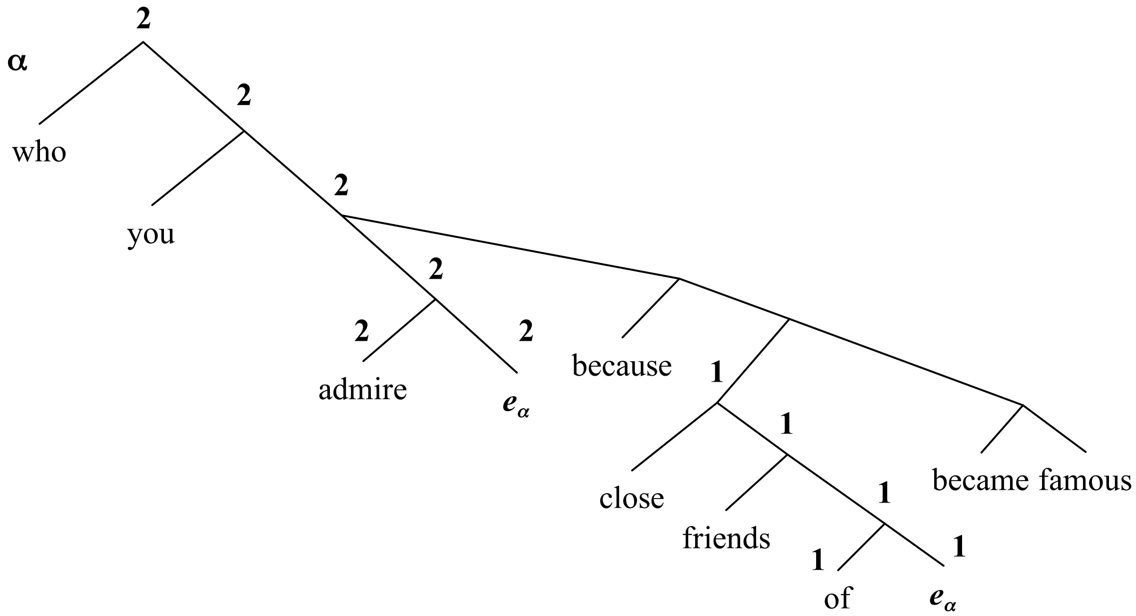
The rescuing effect in (40.b) is due to the fact that the g-projections of the lower gap in the object position extend upward and connect to the g-projections of the illegitimate gap within the subject, as shown in the tree below. As a result, the two g-projections sets form a connected subtree including the binder, and the CC is satisfied:

(40.b)



On the contrary, no rescuing effect arises if the legitimate gap is too high in the tree for its g-projection set to connect to that of the illegitimate gap, as in the following example:

(51) * a person who you admire *e* because [close friends of *e*] became famous



Nunes and Uriagereka (2000) recast the CC in derivational terms: following the argument briefly introduced in section 5.1 (Uriagereka, 1999), the subject/adjunct islandhood is described in terms of phase cyclicity with relative multiple spell-out plus the notion of *sideward movement*: this notion allows an element α, within a certain derivational space, to be extracted and merged within the matrix derivational space under certain circumstances, namely when the (sub)numeration of the matrix phase would not be sufficient to satisfy all thematic requirements of the phase head. To recover from otherwise ungrammatical results, the peripheral *wh*-element/argument of the independent derivational workspace can be extracted and merged in the requested argumental position (sideward movement) then behave as a common *wh*-element newly introduced in the computation by external merge. This proposal is sufficient to subsume both subject-related and adjunct-related connectedness effects (40)-(43).

7.2 Inadequacies: Complex NP-constraint

Observe that there are at least two empirical problems with the standard CC: adjuncts (44) and Complex NP-islands selectivity (45), (46).

Since adjuncts are on a right branch, they should be properly governed by the definition in (48); this should give rise to CC but in fact this is not the case since sentence in (43.a), reported here, is ungrammatical unless an extra gap is inserted in the matrix clause (43.b):

(43) a. ??Those boring old reports, Kim went to lunch [without reading e_i].

b. √ Those boring old reports, Kim filed e_i [without reading e_i].

Longobardi (1985) proposes a modification of the notion of g-projection in order to block it when it "starts" from a right-hand modifier. The minimal change he needs to introduce is that maximal g-projection be properly governed (where properly governed means to be an argument or a predicate of the governing category). In this sense adjuncts are not properly governed and they block g-projection and CC. Then, in order to explain why certain adjuncts are more transparent to extraction than other, we need to incorporate a more subtle proper government condition sensitive to the event-predicated semantic structure on the line of Borgonovo, Neeleman (2000) and Truswell (2006).

On the other hand, the mysterious behaviour of complex NP islands depending on their position does not follow either from Kayne's / Longobardi's revisited CC or from Nunes and Uriagereka's theory of multiple spell-out plus sideward movement: both theories have to implement ad hoc solutions for this problem.

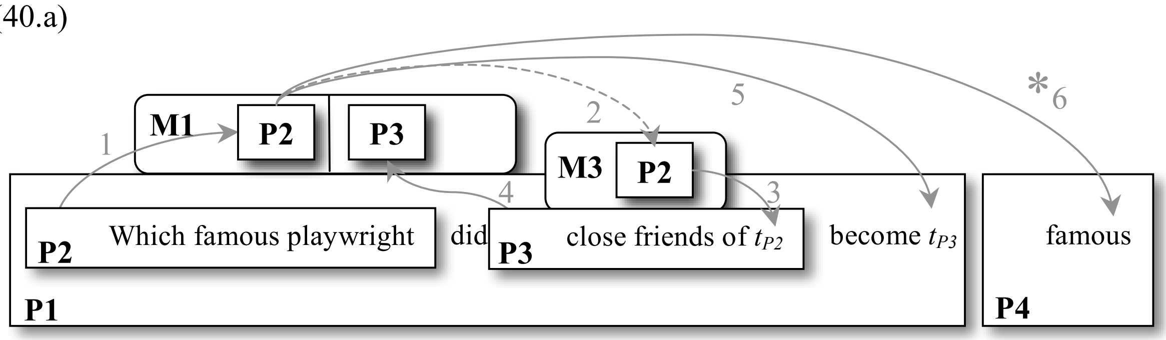
7.3 A possible Solution: Only Selected Phases are Recursive Branches

The theory of island-hood sketched in 5.3 is sufficient to subsume the CC if we add an extra possibility to our memory buffer devices:

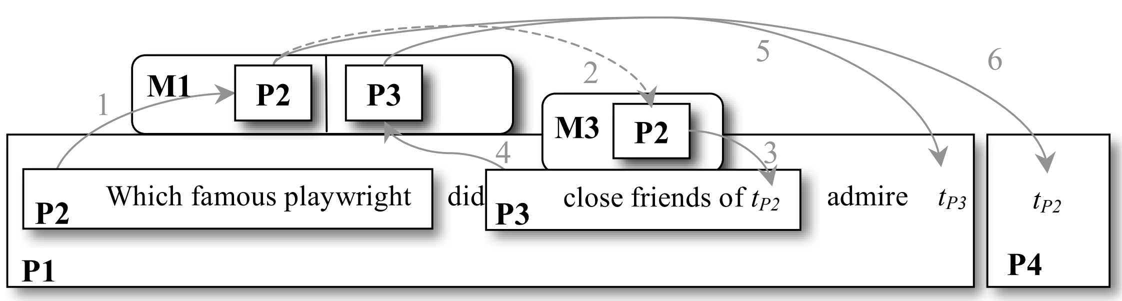
(52) **Parasitic use of the content of the memory buffer**
a nested phase can use elements in the memory buffer of the superordinate phases but only parasitically, that is, without the removing the used elements from the originating memory buffer.

This does not violate any phase-condition in (25) and allows us to capture every relevant example in a natural way (dotted lines mark parasitic use of the content of the memory buffer):

(40.a)

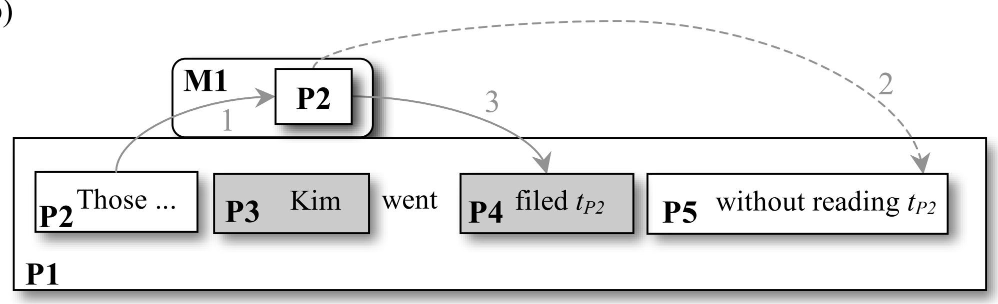


(40.b)



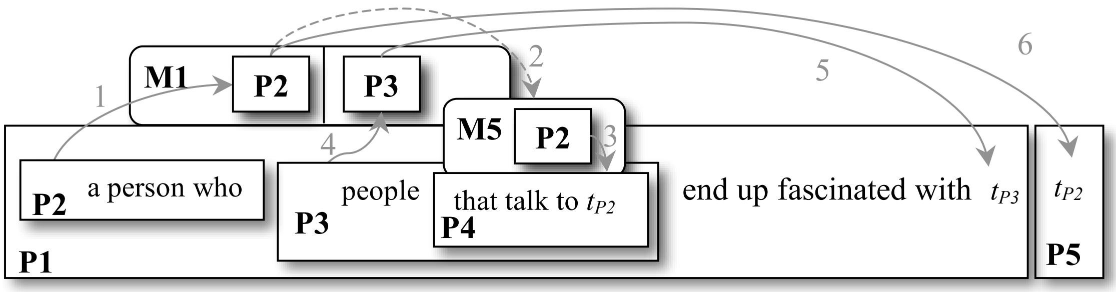
The top-down derivation predicts that (40.a) is ungrammatical not because P2 cannot be interpreted within P3, but because it could be discharged only within a phase which is selected by P1 which in fact is not there.[35] On the other hand, in (40.b) the verb *admire* projects the required phase and the memory buffer (M1) can be discharged, complying with the Success Condition at the end of the derivation. As for adjuncts the relevant derivation, is reported below (cfr. with (36)):

(43.b)

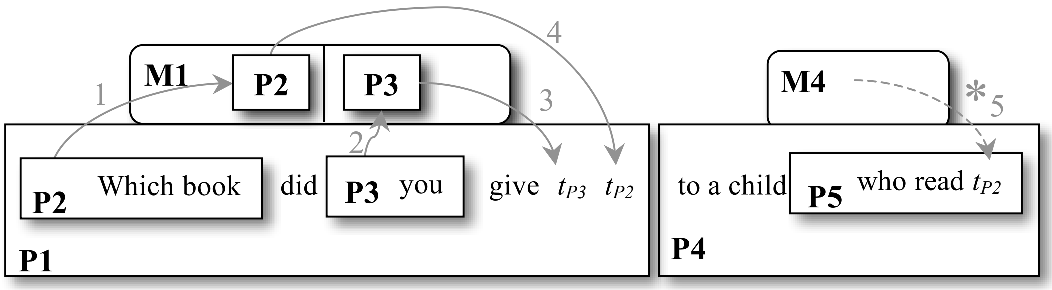


In the end, the ambiguous status of the complex NP constraint is easily accounted for:

(45.b)



(46.b)



In (46.b) (but not in (45.a)) the memory buffer of the relevant phase (M4 for (46.b) Vs. M5 for (45.a)) is empty then no element can satisfy the thematic requirement expressed in P5.[36]

---

[35] This solution is very similar to Nunes and Uriagereka's intuition which would predict (40.a) to be ungrammatical because of the absence of a selected position within the matrix phase which would have triggered sideward movement of the *wh*-element from the subject.

[36] In this case, depending on the structure we would assume for double object constructions, we could predict the following sentence to be significatively better than (46.b):
[Which book]$_i$ did John give [a child [$_{CP}$ who read e_i]]] e_i

8. Conclusion: An Overall Picture of the Proposed Grammar and Further Challenges

In this paper I summarized a few arguments showing that a derivational grammar is superior in capturing many crucial phenomena if we consider a directionality shift with respect to the standard minimalist approach. Adopting a left-to-right, top-down formulation of the structure building operations we can capture in a relatively simple way:

1. asymmetric results of constituency-based tests within the same language and across languages;

2. successive cyclicity in A'-movement without unmotivated intermediate steps;

3. (strong) islands conditions including tricking border-line phenomena with event-predicated arguments and with subject depending on verbal subcategorization ((di)transitives Vs. unaccusatives/passives); moreover we can explore productively some degree of cross-linguistic variation in terms of islandhood (Japanese object-subject non asymmetric behaviour Vs. adjuncts islandhood);

4. a well behaved characterization of strong islands in terms of computational complexity

5. a reformulation of the connectedness conditions in derivational terms, accounting for complex NP-islandhood variability depending on the structural position of the complex NP.

In sum, to capture these results we need to re-orient the following structure building components:

a. merge (Chomsky 1995) became **merge right** (Phillips 1996); it postulates a strict parallelism between linearization and right-branching structure: it predicts that every time an element is introduced in the computation (strictly from left-to-right) this element enters a constituency relation, potentially destroyed during the following derivation.

b. **phases** are derivational spaces (in the sense of Uriagereka 1999) which have to be properly selected and which roughly correspond to extended projections originating from the phase-head (minimally, nouns and verbs). The structure of the functional projections within the phase (following cartographic inquiries) is cross-linguistically fixed; thematic requirements of the phase head project (top-down, on the right edge of the structure) hierarchically ordered ph(r)asal shells (much as in Larson 1988). Phases can appear both in a functional position (then they are licensed but not selected) and in thematic, post-head, position. In the first case we consider these phases as *nested*; in the second case the phases are *selected*. Nesting and selection, a crucial distinction in order to account for movement-based dependencies, is regulated by the *Linearization Principle*, inspired by Kayne's LCA.

c. **movement** became a long-distance memory-buffered left-to-right dependency seeking for selected/licensed positions according to Linearization Principle. Any unselected element/phase is moved into the memory buffer (a LIFO memory) of the relative phase and discharged in the first compatible selected position. Every phase has its own memory buffer; memory buffers must be empty at the end of any phase in order for the sentence to be grammatical (Success Condition). The content of the memory buffer, if any, can only be discharged onto the last phase selected by the current phase head.

It is fair to acknowledge that there are also important challenges yet to be tackle: first of all we need to explain how case-marked phases (e.g. "to whom" like elements) can be nested phases; it is a matter of fact that as soon as we introduce in the numeration a case-marked element, we should have already processed the relevant selecting head (potentially arbitrarily far in the structure). A possible discussion of this problem could start from the solution sketched above to account for the non-asymmetric behaviour of subjects and objects in Japanese in terms of islandhood: assume a "case-marked" elements simply is allowed to project a specific top-down projection, "to whom"-like elements simply project a relevant VP-shell that, unless satisfied by further elements processed in some phase-tailored (selected) derivational workings space, would lead to ungrammaticality.

Another challenge is to determine how the relevant C-command relations can be captured in top-down derivational terms, especially the binding conditions: LP and phase-nesting could be sufficient to recast major empirical phenomena without making any reference to the (representational) notion of C-command. This seems to be possible along the lines of Schlenker (2005) and Bianchi (2007) (see also Shan and Barker 2006 for a left-to-right approach to quantificational binding which dispenses with C-command).

Many of these problems, so far, seem to have at least a(n interesting) conceivable explanation in this radically different, though very restrictive, directionally non-standard minimalist approach.

References

Barker, C. (2007) "Reconstruction as Delayed Evaluation." Ms., Los Angeles.

Belletti, A. (2004) *Structures and Beyond,* Oxford University Press.

Bianchi, V. (2001) "Antisymmetry and the Leftness Condition: Leftness as Anti-c-command." *Studia Linguistica* 55, 1-38.

Bianchi, V. (2007) "Non-redundancy and Backward Anaphora." XXX GLOW Colloquium, Tromso.

Bianchi, V. and Chesi, C. (2005) "Phases, Strong Islands, and Computational Nesting." Ms., University of Siena [presented at GLOW 28, 2005, Geneva].

Bianchi, V. and Chesi C. (2007) "Quantifier Raising in a Top-down Grammar." XVII Colloquium on Generative Grammar, Girona.

Borgonovo, C. and Neeleman, A. (2000) "Transparent Adjuncts." *The Canadian journal of linguistics* 45**,** 199-224.

Cann, R., Kempson, R. and Marten, L. (2005) *The Dynamics of Language: An Introduction,* Oxford: Elsevier.

Chesi, C. (2004a) *Phases and Cartography in Linguistic Computation: toward a Cognitively Motivated Computational Model of Linguistic Competence*, unpublished Ph.D. thesis, University of Siena.

Chesi, C. (2004b) "Phases and Complexity in Phrase Structure Building." *The 15th Meeting of Computational Linguistics in the Netherlands***,** 59-75.

Choi, Y., and Yoon, J. (2006) "Argument Cluster Coordination and Constituency Test (Non)-Conflicts." NELS 37**,** University of Illinois at Urbana-Champaign.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press.

Chomsky, N. (1986) *Knowledge of Language: Its Nature, Origin, and Use*, Praeger.

Chomsky, N. (1995) *The Minimalist Program*, Cambridge, MA: MIT Press.

Chomsky, N. (2000) "Minimalist Inquiries: The Framework." In Martin, R., Michaels, D. and Uriagereka, J., eds., *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, Cambridge, MA: MIT Press, 89-155.

Chomsky, N. (2005a) "Three Factors in Language Design." *Linguistic Inquiry* 36**,** 1-22.

Chomsky, N. (2005b) "On Phases." Ms., MIT.

Cinque, G. (1990) *Types of A-dependencies,* Cambridge, MA: MIT Press.

Cinque, G. (2002) *The Cartography of Syntactic Structures. Vol. 1, Functional Structure in DP and IP,* Oxford University Press.

Collins, C. (2001) "Eliminating Labels." *MIT Occasional Papers in Linguistics* 20**,** 1–25.

Epstein, S., Groat, E. M., Kitahara, H., and Kawashima, R. (1998) *A Derivational Approach to Syntactic Relations,* Oxford University Press.

Grimshaw, J. B. (1990) *Argument Structure*, Cambridge, MA: MIT Press.

Grohmann, K. K. (2003) *Prolific Domains: On the Anti-Locality of Movement Dependencies*, John Benjamins Publishing Company.

Hopcroft, J. E. and Ullman, J. D. (1990) *Introduction to Automata Theory, Languages, and Computation*, Boston, MA: Addison-Wesley Longman Publishing Co., Inc.

Huang, C.-T. J. (1982) *Logical Relations in Chinese and the Theory of Grammar*, Ph.D. thesis, MIT.

Kayne, R. S. (1983) *Connectedness and Binary Branching*, Foris.

Koizumi, M. (2000) "String Vacuous Overt Verb Raising." *Journal of East Asian Linguistics* 9**,** 227-285.

Laenzlinger, C. (2005) "Some Notes on DP-internal movement," *GG@ G* 4**,** 227–260.

Larson, R. (1988) "On the Double Object Construction in English." *Linguistic Inquiry* 19**,** 589–637.

McCloskey, J. (2002) "Resumption, Successive Cyclicity, and the Locality of Operations." In Epstein, S. and Seely, D., eds, *Derivation and Explanation in the Minimalist Program***,** 184–226.

Nunes, J. and Uriagereka, J. (2000) "Cyclicity and Extraction Domains." *Syntax* 3**,** 20-43.

Ogawa, Y. (2001) *A Unified Theory of Verbal and Nominal Projections*, Oxford University Press.

Phillips, C. (1996) *Order and Structure*, Ph.D. thesis, MIT.

Phillips, C. (2003) "Linear Order and Constituency." *Linguistic Inquiry* 34**,** 37-90.

Richards, N. (1999) "Dependency Formation and Directionality of Tree Construction." *MIT Working Papers in Linguistics* 34**,** 67-105.

Rizzi, L. (1990) *Relativized Minimality*, Cambridge, MA: MIT Press.

Rizzi, L. (2004a) "On the Form of Chains: Criterial Positions and ECP Effects." Ms.. University of Siena.

Rizzi, L. (2004b) *The Cartography of Syntactic Structures,* Oxford University Press.

Ross, J. R. (1967) *Constraints on Variables in Syntax,* Ph.D. thesis, MIT.

Saito, M. and Fukui, N. (1998) "Order in Phrase Structure and Movement." *Linguistic Inquiry* 29**,** 439-474.

Schlenker, P. (2005) "Non-Redundancy: Towards a Semantic Reinterpretation of Binding Theory." *Natural Language Semantics* 13**,** 1-92.

Shan, C. and Barker, C. (2006) "Explaining Crossover and Superiority as Left-to-right Evaluation." *Linguistics and Philosophy* 29(1), 91–134.

Shieber, S. M. (1986) *An Introduction to Unification-based Approaches to Grammar,* Center for the Study of Language and Information, Stanford, CA: Stanford University

Sportiche, D. (1988) "A theory of Floating Quantifiers and its Corollaries for Constituent Structure." *Linguistic Inquiry* 19**,** 425-449.

Stabler, E. (1997) "Derivational Minimalism." Logical Aspects of Computational Linguistics: First International Conference, LACL'96.**,** Nancy, France**,** 68.

Starke, M. (2004) "On the Inexistance of Specifiers and the Nature of Heads." In Belletti, A., ed., *Structures and Beyond*, Oxford University Press, 252-268.

Steedman, M. (2000) *The Syntactic Process*, Cambridge, MA: MIT Press.

Truswell, R. (2006) "Extraction from Adjuncts and the Structure of Events." Ms., London.

Uriagereka, J. (1999) "Multiple Spell-out." In Epstein, S., and Hornstein, N., eds., *Working Minimalism*, Cambridge, MA: MIT Press.