Linearization (As Part of Core Syntax)

1. Introduction

Linear order is a precedence relation, namely a total strict order[1], defined between any terminal item (word/morpheme) in a pronounced sentence. No consensus either on the source of this constraint (Phonetic Form [PF] requirement or part of core syntax) or on its unicity (§31. Linearization - Idsardi & Raimi) is present yet.

The mainstream perspective considers linearization as a byproduct of the monodimensional reduction, imposed by the Sensory Motor (SM) system, of the bidimensional (dominance and sisterhood) phrase structure.

In these pages, I present what I believe to be the least common denominator of the minority perspective that considers a relevant part of linearization as strictly dependent on the asymmetric nature of the hierarchical structure. From this viewpoint, the induction of linear order is in fact a component of narrow syntax. Eventually, I discuss the hypothesis that the source of the relevant asymmetries is the incrementality of phrase structure building as it flows in time.

The chapter is structured as follows: a brief introduction to the nature of a "genuine explanation" in grammatical theorizing will set the stage (§2). The presentation of an influential hypothesis connecting asymmetric C-command with linear order, which is Kayne's Linear Correspondence Axiom (LCA; Kayne, 1994), will follow (§2). This section will show how the LCA intuition can be fruitfully maintained in minimalist (probe-goal) terms. Crucial empirical evidence will be offered in §3, suggesting that incrementality in the application of phrase structure building operation Merge justifies a genuine necessity of considering linearization as a core syntactic component. This evidence will include contrasts in constituency tests, such as coordination, fronting, and binding (Phillips, 1996, 2003; §4.1), the "leftward" bias for movement (Abels & Neeleman, 2012; Cinque, 2005; Kayne, 2020; §4.2), and subject islands, parasitic gaps constructions, and preposition-stranding vs pied-piping preferences (Bianchi & Chesi, 2014, 2015; §4.3). The relevance of the notion of phase (§13. Phase theory) and workspace in terms of linearization constraints will be discussed in §4.4. A theoretical consideration (the "telepathy paradox") will precede the conclusion of this chapter, showing in which sense the discussed asymmetries should be necessarily originated from narrow syntax.

2. A Genuine Explanation

Chomsky (2020) claims that a "genuine explanation" of fundamental properties of human language should face evolvability and learnability problems: on the one hand, a fundamental property should be simple enough to be

[1] I.e., it can be defined between any terminal item (total), it is irreflexive (if α precedes β, then β does not precede α) and transitive (if α precedes β and β precedes γ, then α precedes γ). Angled brackets will be used to indicate ordered sets; that is, α precedes β and β precedes γ will be represented as <α, β, γ>, or simply <α β γ>.

evolved (§54. Biology, genetics, and evolution), and on the other it must support language learnability solely on the basis of positive evidence available to children from primary linguistic data (§38. Learnability). In this vein, X-bar theory (Chomsky, 1970) must be rejected since it collapses together three distinct aspects, namely hierarchy, projection, and linearization. Chomsky argues convincingly that only the first aspect, hierarchy, is the fundamental property that must be expressed by the simplest possible mechanism, in compliance with the evolvability requirement. With Merge (§18. Merge) being the simplest possible set formation operation, namely, the operation taking two items α and β (either from the lexicon or the results of other Merge operations) thus forming the set consisting of the unordered pair {α, β}, the natural assumption is that language faculty simply evolved when Merge appeared. Merge, by itself, in its two flavors, external and internal (§22. Copy and Move), is sufficient to tackle the Galilean Challenge; namely, it succeeds to ensure an infinite generation of structured sets out of a finite set of lexical items. Merge(α, β) is then a binary function producing unordered sets that poses constraints neither on the category of α and β nor on their linear order (which "simply doesn't have anything to do with the core of language" Chomsky, 2020:16). From this perspective, linearization is part of externalization (§4. Externalization), namely a necessity imposed by the SM interface and, possibly, only by certain articulatory systems (§32. Linearization in sign languages). Consequently, if we aim at an adequate grammatical theory (Chomsky, 1957:13), we must filter the unordered sets produced by free Merge so as to get only grammatical phrase structures effectively corresponding to the descriptions of an infinite number of sentences that are part of the language we want to describe. These restrictions are often formulated as filters such as parameterizations at SM interface for inducing systematic linear ordering on the line of the head-directionality parameters (Travis, 1984). If certain parameters, like head directionality, are suggestively supported by evidence indicating that they are early settable (possibly based on prosodic cues to which children are very early sensitive, Christophe et al., 2003), no evidence of learnability is discussed for other filters such as Labeling (§14. Labeling theory), which is the ultimate algorithm in charge of the rejection of merging two incompatible items (e.g., *Merge(the, the))[2]. Notice that postponing the rejection of an unwanted phrase structure either to PF or Logical Form (LF) seriously calls into question what the role of "narrow syntax" is, if systematic, universal, structural constraints are expressed elsewhere. More importantly, this would lead to the conclusion that Labeling should be unnecessary (Collins, 2002) or necessary only at LF (§14. Labeling theory §26. Labels & Categories?). This second option would both pose evolvability and learnability problems.

Possibly in relation to this, Chomsky discussed recently the opportunity to restrict free recursion by limiting Merge to Workspaces (Ws) (MERGE, different from Merge in Chomsky, 2020). Technical details are unnecessary to understand that, in order to meet the basic descriptive adequacy requirement (Chomsky, 1965:24), the

[2] An important set of predictions, including the rejection of Merge(α, α), is derived from the Distinctness / Contiguity Theory discussed in Richards (2010, 2016). His interesting considerations are more in line with a view of linearization as a PF constraint. Though the consequences of this idea have an impact on Kayne's Asymmetric C-command discussed in §2, I point the reader to §31. Linearization by Idsardi & Raimi section of this handbook.

application of Merge must be restricted since very few of the theoretically possible combinations of lexical items and phrases lead to well-formed, explanatorily adequate, structures[3]. Two options are available: one possibility is to formulate independent constraints to limit free Merge, the other is to re-define Merge to limit its exuberance, hence reducing the set of legitimate results of this operation from the beginning. The first path has been explored by works on Labeling (§14. Labeling theory) and must face the learnability and evolvability issues previously discussed, the other is pursued here and it aims at the status of "genuine explanation".

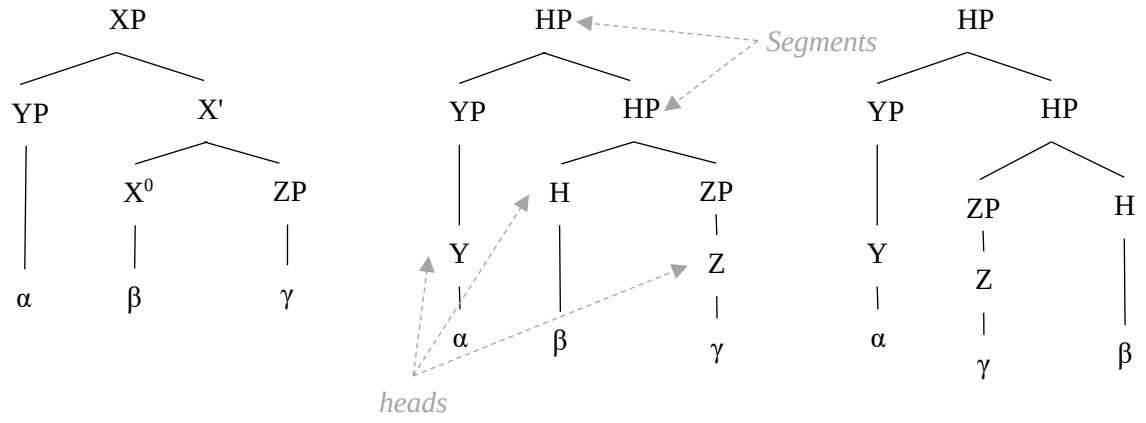
3. Asymmetric C-command and precedence: linking linearization to phrase structure building

Kayne argues that linear order is a byproduct of the asymmetric[4] C-command relation (Kayne, 1994). More precisely,

(1) if *X*, *Y* are non-terminals and *x*, *y* terminals such that *X* dominates *x* and *Y* dominates *y*, then if *X* asymmetrically c-commands *Y*, *x* precedes *y*. (Kayne, 1994, p. 33).

This intuition had a great impact on phrase structure description: X-bar theory, (2).a, must be abandoned in favor of a more primitive constraint explaining the asymmetry between specifiers and complements by relying on the difference between categories (consisting of heads and non-heads) and segments[5] (2).b-c.

(2) a. X-bar b. linearizable structure c. non-linearizable structure



[3] Same concerns, in terms of learnability and evolvability, apply to the notion of Numeration (Chomsky, 1995:225) which I will not consider here. Notice that Numeration is crucial in various derivational approaches that share the perspective here discussed (e.g., the one based on "split-Merge"; Zwart, 2009; see also §4.4).

[4] Asymmetric is an antisymmetric relation *R* such that if *X R Y* then *Y* ¬*R X*.

[5] Categories are non-terminal items. Heads are those categories immediately dominating a terminal item. A segment is a category split in multiple contiguous positions (adjunction) (Chomsky, 1986). LCA, in its 1994 formulation, has been seriously challenged (Abels & Neeleman, 2012; a.o.). Here I agree with those critics, and I assume implicitly (i) single headness, (ii) endocentricity, and (iii) singularity of specifier and complement positions to derive the expected asymmetries. Nonetheless, these (implicit) assumptions will be unnecessary given the minimalist perspective discussed later.

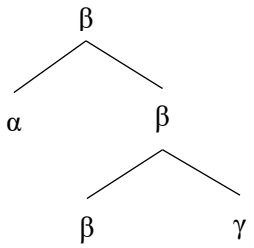
The definition of asymmetric C-commands becomes central for the rejection of (2).c, corresponding to the base-generated SOV word order, in favor of (2).b, the allegedly base-generated SVO word order:

(3) *X* c-commands *Y* iff *X* and *Y* are categories, *X* excludes *Y*, and every category that dominates *X* dominates *Y*.

The segment *HP* excludes *H* and *ZP* from C-commanding asymmetrically *Y*; therefore, since *YP* asymmetrically C-commands *H* (producing $<\alpha, \beta>$) and *H* asymmetrically C-commands *Z* (hence $<\beta, \gamma>$), (2).b is the linearizable (LCA-compliant[6]) representation of (2).b and not as (2).c (which is excluded by LCA, since $d(X)=<\beta, \gamma>$ and not $<\gamma, \beta>$). The theoretical import of the exclusion of (2).c is the conclusion that the theory is incompatible with a directionality parameter. Therefore, complement-head order must be derived through movement (§4.1)[7].

In Minimalist terms, (2).b is obtained as a result of two Merge operations as indicated in (4), adopting a label-free description (Collins, 2002; Stabler, 1997): Merge(β, γ) = $\{\beta, \gamma\}$, then Merge(α $\{\beta, \gamma\}$) = $\{\alpha$ $\{\beta, \gamma\}$.

(4)



The distinction between "terminal nodes"/"heads" and "nonterminal nodes"/"phrases" (including intermediate and maximal projections) is absent in (4) (in line with the Bare Phrase Structure approach and the Inclusiveness Condition Chomsky, 1995:249) but can be inferred in various ways: one simple (possibly simplistic[8]) way is to consider terminals/heads the elements at their first Merge and phrases as those re-merging multiple times (in (4), only β at its second merge qualifies as a phrase in this strict sense). From a purely descriptive perspective, this solution alone would be insufficient to qualify γ as a complement (C) of β (the head, H) and α as its specifier (S). Practically, the antisymmetric relation S-H-C (specifier-head-complement) (as well as the ban on SOV base-

[6] Given phrase marker P, with *T* being the set of terminals and *A*, containing all pairs of non-terminals such that the first asymmetrically c-commands the second, then the LCA applies: "*d(A)* is a linear ordering of *T"* (Kayne, 1994:6).

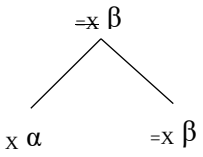
[7] I will keep defending here the idea that asymmetric C-command maps to linear order. For an alternative suggestive perspective, more in line with a PF parameterization of linear order, see (Abels & Neeleman, 2012).

[8] As discussed in Rizzi (2016), this might not be a sufficient strategy and a more articulated labeling procedure would be necessary to resolve critical (apparently) symmetrical cases (for instance for merging two heads or two XPs). Here I conclude that these configurations pose a problem only if we assume that the label must be "the closest" head as assumed in Rizzi (2016). Assuming a probe-goal asymmetry, as discussed later (a perspective which is more in line with that of Cecchetto and Donati, 2010 probing algorithm), the critical issues dissolve.

generated word-order) does not follow from the bare phrase structure perspective and must still be derived from independent assumptions.

Kayne (2018) explicitly proposes to adopt a Merge operation that generates ordered pairs <α, β> instead of unordered sets {α, β}. Kayne's argument lies in the intuition that probes precede goals since "probe-goal searches share the directionality of parsing and production" (Kayne, 2018, p. 9). If on the one hand, this solution might appear to be a pure stipulation, on the other, the intuition of considering parsing (Phillips, 1996, 2003) and generation (Momma & Phillips, 2018) directionality as the null hypothesis for phrase structure building has been proved to be empirically adequate for instance in capturing conflicting constituency test results (§4.1) and explaining structural priming effects in predicate structure production (Momma et al., 2018). Furthermore, from a purely configurational perspective, thinking in terms of probe-goal, a clear asymmetry can be exploited: several convergent hypotheses suggest that probes operate in head-complement relations (Probing algorithm, Cecchetto & Donati, 2010; select features, Chesi, 2004, 2012, Stabler, 1997; C-selection Sheehan, 2013; a.o.). If we accept the idea that Merge is feature-driven (exactly as Internal Merge) and that (5) is the result of a Merge operation[9], not only can we obtain a clear asymmetry between the two elements merged, but we can also conclude that there is an unambiguous ordering strategy for α and β. If looking at the tree from a Bottom-Up perspective there is no reason to process α and β in a specific order; building the tree Top-Down, from the root (which is, by definition, unique), the perspective leaves us no choice: the probe/select feature is processed "before" the goal, hence β should be processed before α in (5).

(5)



This minimal condition imposed to Merge to be successful has important consequences for the phrase structure building dynamics. The relevance of these consequences will be summarized in the following section. Similarly, the trigger for standard (A/A') movement operations and, more in general, the asymmetric nature of the specifier/adjunct (α in (4)) will follow from this perspective (§4.3).

4. Consequences of the hierarchical asymmetries and their effects on linearization

4.1. No directionality parameter needed

[9] Adopting Stabler's (1997) formalism, I will use =X to express the selection/probe features, where X are the relevant categorial features expected in the goal.

As anticipated, the first consequence of LCA was that no directionality parameter is possible and head-final order must be the result of the movement of the complement over the head (Kayne, 2011). Focusing on OV order, as a prototypical instance of head-final "parameterization", Kayne's argument consists of two parts: first, he observes that the object is able to move in a preverbal position as exemplified below:

(6) They are having [their car]$_i$ washed $\__i$ (Kayne, 2018:2)

Second, in allegedly head-final languages, the object does not remain in its thematic position, as testified by the particles (*neg* and *aux* in Korean (7) and Infinitivus-pro-participio (IPP) constructions in German (8) respectively) intervening between O and V as default word order:

(7) Mica-ka hakkyo ey *an* ka-ss-ta. Korean (S O Neg V) (Whitman, 2005)
Mica NOM school to NEG go-PAST-INDIC
Mica didn't go to school.

(8) Ich glaube, dass er das Buch *hätte* l esen wollen. German (S O Aux V) (Zwart, 2007)
I believe that he the book would-have to-read to-want
I believe that he would have wanted to read the book

These facts constitute a cue about the lack of necessity of a directionality parameter, but they are not the ultimate counter-evidence to send into retirement the head-directionality parametric approach. Moreover, the analyses deriving OV order include both V-head movement and consequent remnant movement (Nkemnji, 1995) and direct O movement (Kandybowicz & Baker, 2003) attracted by specific (possibly AgrO) functional heads. Only this second solution would qualify as a genuine counterexample to the assumption that the complement remains in the base thematic position and the necessary proof is the absence of any evidence suggesting that O remains in situ in OV, namely, there should always be some low adverbial Advx such that <O, Advx, V> in any OV language.

A more perspicuous argument against head directionality can be built on the evidence indicating universal ordering of the functional projections: much of the data collected under the so-called Cartographic approach (Belletti, 2004; Cinque, 2002; Rizzi, 2004) indicates that functional projections are ordered in a cross-linguistically consistent way (Cinque, 1999) and this order is stable both in allegedly head-initial (Italian, French, English) and head-final (both agglutinating languages like Turkish and Korean and non-agglutinating languages like German and Hindi, with the first complying with the Mirror Principle, Baker, 1988, obtained through head-movement plus incorporation, or roll-up movement, Roberts, 2010).

Once the functional projections are considered "maximal projections" (that being the only viable way to constrain rigidly their order), rather than free adjuncts, the head directionality parameter would predict perfect mirror languages, but no perfect mirror image of English (or Japanese) exists (Kayne, 2011:1).

Cinque's work is especially enlightening from this perspective: counterexamples of Greenberg's universal 20[10] have been found, but only as long as the relevant functional elements appear post-nominally. In fact, assuming that the hierarchical functional structure is basically the one indicated in (9).a[11], the mirror order is obtained through roll-up movement as in Gungbe (9).b:

(9) a. [$_{Dem}$ These [$_{Num}$ three [$_{Adj}$ big [$_{N}$ crabs]]]

b. [[[[$_{N}$ àgásá [$_{Adj}$ ɖàxó _$_{N}$]][$_{Num}$ àtòn __$_{Adj}$]][$_{Dem}$ éhè _$_{Num}$]]]$_{i}$ [$_{D}$ 1ɔ́ _$_{i}$ [$_{NumP}$ 1ɛ́ _$_{i}$]]

crabs big three this

these specific three big crabs (Aboh, 2004:92–93)

Given four functional heads, Cinque's analysis explicitly predicts certain word orders such as (10).a-b and excludes others because neither NP movement nor pied-piping triggered by the NP (both of the [X [N]] and [N [X _$_{N}$]] kind) can derive them (10).c-d[12]:

(10) a. **(N)** [Dem **(N)** [Num **(N)** [Adj _$_{N}$]]] (NP movement)

b. **(Adj N)** [Dem **(Adj N)** [Num _$_{[Adj\ N]}$]] (NP pied-piping)

c. * (**N)** Num (**N)** Dem (**N)** A **(N)**

d. * A Dem **(N)** Num **(N)**

An alternative analysis is proposed by (Abels & Neeleman, 2012) and derives the very same restrictions without relying on specifier-head-complement asymmetric assumptions or LCA, but only assuming that movement is leftward. Their proposal is more in line with the minimalist perspective here adopted. Although no linearization (and projection, cf. Roberts, 2017) is assumed in their model, a base Merge order (resulting in asymmetric C-command relations) is assumed ({{{N A} Num } Dem}) as well as the stipulation that movement is necessarily toward a C-commanding position, which is assumed to be "leftward". Assuming that Movement, but not Merge, needs to be directional because of parsing (temporal) asymmetry (Ackema & Neeleman, 2002) poses a logical problem: if performance operates on one structure building operation to "optimize" it in processing terms, why shall this only apply to Move/Internal Merge? The fact that (External) Merge does not leave traces (in the technical sense) to be "interpreted" in parsing does not constitute, per se, a relevant asymmetry: bottom-up Merge requires parallel workspaces to build a complex subject before the predicate can be merged with it in head-initial, SV, languages like English (see §4.4). Different assumptions on how to deal with independent workspaces have an

[10] "When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or its exact opposite" (Greenberg, 1963:87). A relevant generalization that I do not have space to discuss here is the Final-Over-Final Condition (Sheehan et al., 2017).

[11] This word order is the most attested one according for instance to (Hawkins, 1983).

[12] According to Cinque, NP movement to specific AgrP position interspersed with the functional items is necessary to licence the functional positions as extended projections of the lexical NP, in the sense of Grimshaw (1991).

impact in terms for processing efficiency that could be approached from the parsing perspective as well. Identical considerations involve agreement and labeling.

Notice, moreover, that Cartography (in this sense the {{{N A} Num } Dem} basic Merge order) represents a puzzle yet to be solved: functional layers are not only strictly ordered, but also non-recursive; that means, unless a new lexical head (N or V, for instance) is merged, the sequence of Merge introducing functional layers on the top of a lexical category will always be limited[13].

4.2. Asymmetry of time, incrementality, and intermediate constituency

Kayne (1994:36–38) derives the superiority of the "precedence" relation over the "following" one based on a purely geometrical consideration by exploiting the asymmetry of time flow[14]. Its argument (orthogonal to the one discussed in §3) has been rephrased in minimalist terms by considering the asymmetry between probes and goals (Kayne, 2011:12):

(11) a. Probe-goal search shares the directionality of parsing and production.
b. Probe-goal search proceeds from left to right.

The consequence is that heads (probes) always precede their complements since the goal is either the complement (external merge) or within the complement (this would trigger internal merge).

The idea that parsing "influences" grammar is not new: Phillips persuasively argued that "Parser is the Grammar" should be the null hypothesis[15]. One of the main arguments he proposes to support this idea relies on conflicting constituency tests: coordination, (12).a, seems more liberal than fronting, (12).b, in targeting constituents (Phillips, 2003:39):

(12) a. Wallace gave [Gromit some biscuit] and [Shawn some cheese] for breakfast.
b. *[Gromit some biscuit]$_i$ Wallace gave _ $_i$ for breakfast.

---

[13] A possibly relevant exception to this, according to (Rizzi, 1997:297), is the Topic position which seems to be recursive:
i. [$_{Topic}$ Il tuo articolo]$_i$, [$_{Topic}$ ieri sera], [$_{Topic}$ in ufficio], io proprio non l'ho potuto leggere.
The your paper, yesterday night, in office, I really not it$_{CL}$ can read.
Your paper, yesterday night, in the office, I really didn't succeed to read (it)
This is the only functional position which seems to allow stacking freely.

[14] Assuming an abstract node *A* adjoined to the root, such that *A* asymmetrically C-commands every node in *T*, *a* being the first terminal node such that *A* dominates *a* and d(A) = <*a*, …>, *a* will enter any precedence relation, for any time slot consisting of a substring of the final sentence produced or parsed. This qualifies *A* as the node for which a total strict order in terms of asymmetric C-command relations can be defined and can be exploited, incrementally, to order any terminal. A similar consideration would support the Top-Down approach based on "split-merge" adopted in Zwart (2009).

[15] Many different approaches have reached the same conclusion. A notable one, integrating parsing and semantic compositionality is Dynamic Syntax (Kempson et al., 2000). Also many linear constraints have been related to parsing strategies (notably Abels & Neeleman, 2012; Hawkins, 1994; Sheehan, 2017 a.o.). From a compositional semantic perspective, coherent with this directionality shift, are works by (Barker et al., 2007; Barker & Shan, 2014).

Many explanations for the limited restrictions of coordination as compared to movement have been offered in the literature, but no general unified account of such contrasts, obtained from basic structure building properties, exist, with this notable exception: Phillips derives impressive consequences from Merge application dynamics, simply assuming an Incrementality Hypothesis:

(13) *Incrementality Hypothesis* (Phillips, 2003:37)

Sentence structures are built incrementally from left to right.

Let us take, for instance, a long-standing puzzle related to the contrasting results of binding vs fronting. Two possible structural analyses, (14).a and (14).b, can be associated to the same sentence:

(14) a. John [gives [the books] [$_V$ [to *them*$_i$] [$_V$ [in the garden] [$_V$ [on *each other*$_i$'s birthday]]]]]

b. John [[[[gives the books] to them] in the garden] on each other's birthday]

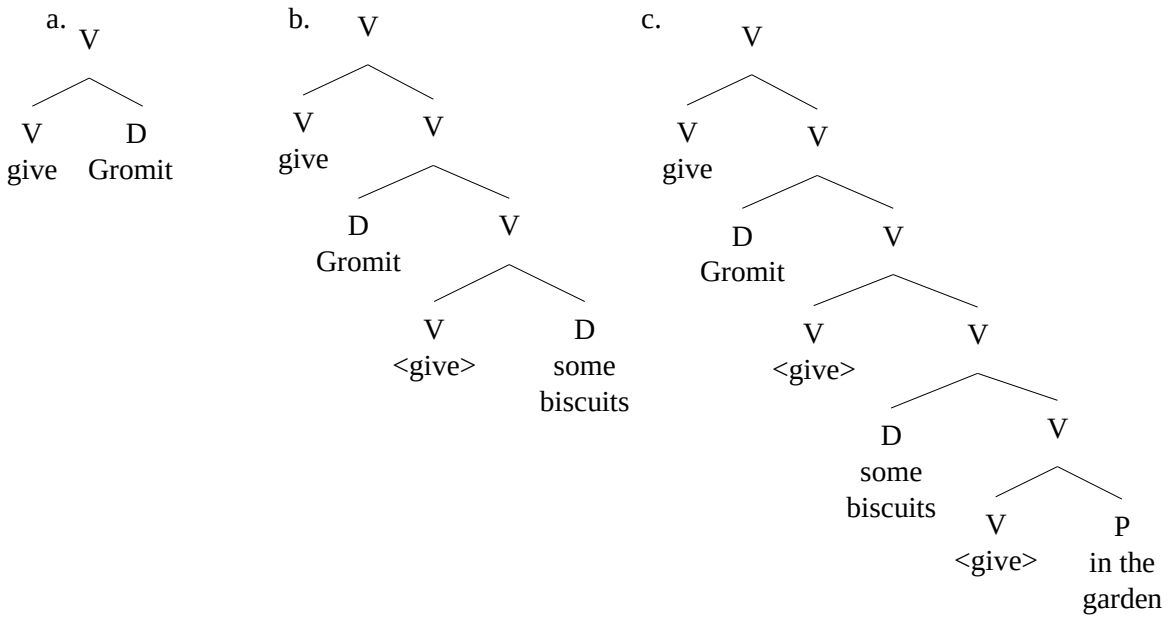
Both (14).a and (14).b descriptions are apparently necessary: for binding purposes, "them" should C-command[16] the reflexive anaphor "each other", and this indicates a right-branching structure based on VP-shells (Larson, 1988). On the other hand, from the VP-fronting perspective, a left-branching structure is expected, (14).b, as indicated by the possibility of fronting [give the books to them] while stranding [on each other's birthday] as in (15):

(15) John intended to give the books to *them*$_i$ in the garden and [give the books to *them*$_i$]$_j$ he did $\__j$ on *each other*$_i$'s birthday.

While (Pesetsky, 1995) considers both representations as legitimate (a "cascade" one, (14).a, for binding and a "layered" one for fronting, (14).b), (Phillips, 1996, 2003) minimally derives this and other contrasts by assuming that Merge operates incrementally from left to right and creates constituents that can be destroyed during the derivation, as schematized in (16) for the relevant VP shells in (12):

(16)

[16] For a resolutive discussion of the critics that binding does not necessarily require C-command, refer to (Phillips, 2003).



Simplifying a bit the original argument, [Gromit some biscuit] can be coordinated in (12).a since it appears as a constituent at step (16).b, where coordination applies, but it is not licensed yet as a constituent in (12).b. From this perspective, conflicts in constituency tests are explained by relying on the time at which a specific operation applies: coordination can target intermediate constituents like (16).b that can then be later destroyed, since it applies locally, while fronting cannot access the same constituents either because they are not constituents yet or they will be destroyed/modified by the time the gap will be licensed. Notice that the interpretation of the fronting case in (15) is that the "PP stranded" is in fact "late adjunct" after the VP is "reconstructed" in the gap position[17]. Interestingly, C-commands will always proceed incrementally in a monotonic way, without destroying any asymmetric relation previously built. This solution, based on a sort of Larsonian VP-shells, explains the anaphor binding observed (14).a by deriving the structure proposed in (16)[18].

4.3. A new perspective on movement, islandhood, and parasitic gaps

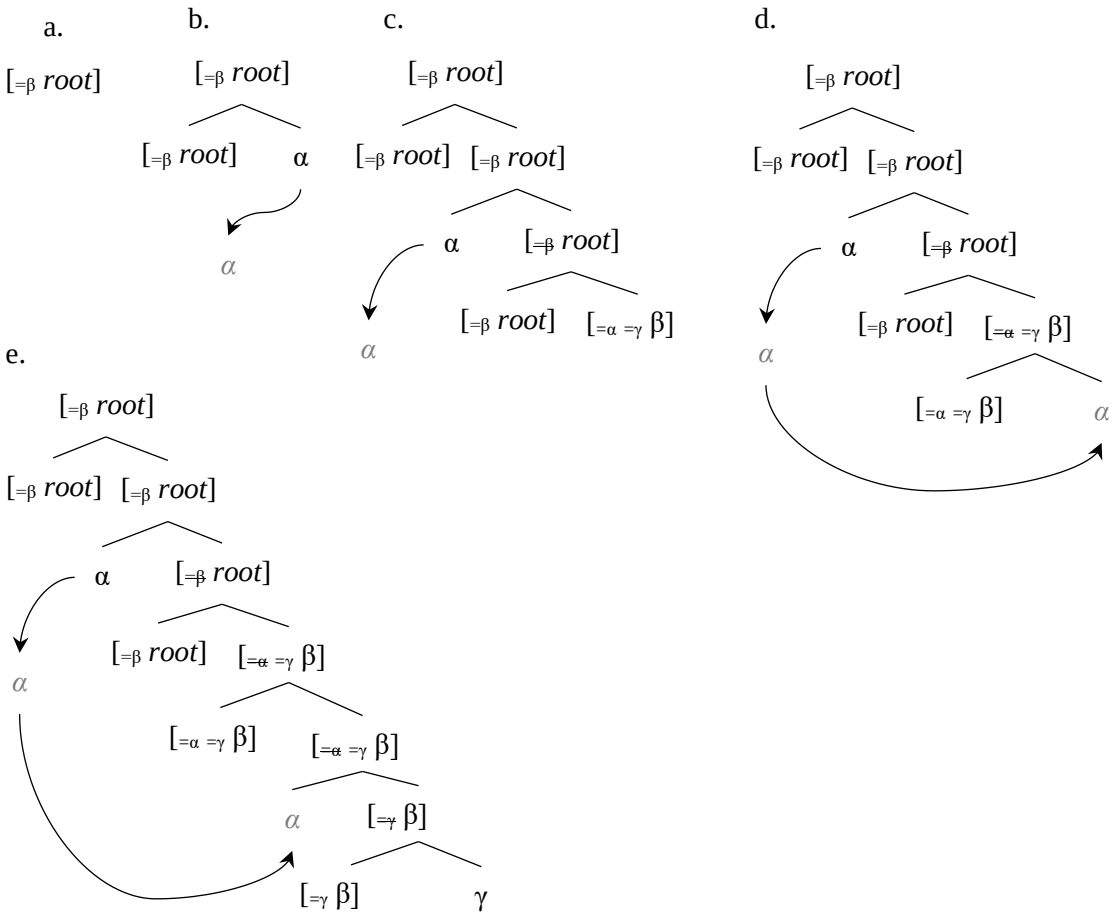
Back to (4), [$_{\beta}$ α [$_{\beta}$ β γ]], if β and γ, respectively the head and the complement, are systematically ordered as <β, γ> in virtue of a probe-goal relation, the very same relation can not be used to order specifiers (if needed at

[17] See (Baltin, 2006) for discussion of more complex, possibly problematic, cases.

[18] Den Dikken (2018, p. 81) noticed that this analysis is problematic if applied to ellipsis (included in the original discussion, see also Phillips & Lewis, 2013), where the assumed VP left-branching structure, necessary to avoid impossible interpretations, conflicts with anaphoric binding. In his work, Den Dikken convincingly argues that fronting and ellipsis differ in their "reconstruction" nature, with the first, maintaining a parallel structure with respect to the fronted VP shells, and the second, consisting of an elliptical VP with an implicit (since just mentioned) complement unable to bind properly the anaphor. Disregarding technical details, this analysis of ellipsis, on the one hand, solves the binding puzzle without relying on different cascades/layers structural descriptions, whereas, on the other hand, it suggests that the ellipsis motivation, a sort of "avoid redundancy" (Bianchi, 2009; Schlenker, 2005), is directional: only what has been already merged can be "reconstructed" (in a pre-theoretical sense).

all) and adjuncts[19] to the left of (i.e. before) the head. A practical problem here is that α, in (4), if selected/probed by β, should have been processed/linearized later according to (5) and (11).b. A licensing solution for α before its probe/licensor recurs to Movement: in a probe-first (Top-Down) derivation, resulting in a left-right structure building as assumed by (Phillips, 1996, 2003), the absence of a proper probe can be interpreted exactly as the trigger for movement (Chesi, 2004, 2012, 2015). If α bears unselected/unprobed (categorical) features, it remains as a pending constituent in the current workspace as long as these unprobed features get properly probed/selected. In the case of the Subject-Verb-Complement canonical word order, represented as α-β-γ by the tree in (4), the derivation would then proceed as schematized in (17):

(17) Lexical items: [=β *root*], α, [=α =γ β], γ



Root is whatever node we need to start a derivation (e.g., a possibly phonetically empty C selecting the highest functional head of an extended VP projection). α, once merged in (17).b, being unprobed, will be readily "moved"

[19] While the necessity of adjunction is hardly deniable (Chomsky, 2020:49) and explicitly assumes a Pair-Merge operation, distinct from MERGE to deal with adjunction), specifiers have often been questioned (Abels & Neeleman, 2012; Starke, 2004).

in the pending area of the workspace waiting to be properly selected by a following head[20]. Then β will be merged, (17).c, satisfying the root selection requirement. Therefore, the pending features of α will be re-merged satisfying the first selection requirement of β (17).d (thematic role assignment). In the end, γ will be merged, (17).e, satisfying the last selection requirement of β. γ then qualifies as the internal argument. A derivation will be successful if, and only if, no pending item remains in the workspace and no more selection request must be satisfied. Notice that selection requirements must be lexically ordered to derive the relevant predicate argumental shells as discussed in (16) and (17) (à la Larson, 1988).

Bianchi & Chesi (2006, 2014, 2015) use the "memory buffer" metaphor to deal with movement "from the specifier" position: a near-sighted memory that makes available only the last compatible unselected/unprobed item[21]. Applying a version of the Earliness Principle, "moved" items are the first ones to be remerged[22], but their probe/selector must be found in the current workspace W, where the W is the extended projection ending with a lexical category (N, V or A) and continuing through the last selected complement (right recursive branch of the tree, namely the sequential workspace/phase in Bianchi & Chesi, 2006). Any expansion of a functional position as a complex phrase, that is a phrase containing one lexical item, constitutes a nested workspace (a nested phase in Bianchi and Chesi's terminology). It seems that nested workspaces are unable to satisfy the pending dependency of a superordinate workspace, but they can make use (parasitically) of the elements pending in the "memory buffer" of the superordinate workspace[23]. This perspective clarifies in which sense linearization must be

[20] Various proposals have been put forward to license α in this position. Building on Stabler's formalism, Chesi & Brattico (2018) rely on selection (=x) for head-complement linearization, Merge([$_{=X}$ H], [$_{X}$ C]) = < $_{H}$ H, C> (or [H [C]]), and licensing (+y) for specifier/adjunct linearization Merge([$_{Y}$ S], [$_{+Y}$ H]) = <$_{H}$ S, H> (or [[S] H]). The same result is obtained here relying on superordinate selection and free Merge of α. This might be too unrestrictive and a more complex C selection would be necessary, such us [$_{=AgrS\ =T}$ C], where =AgrS (the licensor of the thetic subject position in Bianchi & Chesi, 2014) will license the agreement morpheme of α, but not the argumental one (e.g. D) that would force α to remain in search for an argument selector (e.g. [$_{=D}$ V]). In the meanwhile, =T will be satisfied by the relevant predicate-attached morphemes (possibly triggering agreement/feature unification). I avoid discussing the pros and cons of these proposals here as well as the necessity of considering heaviness as another trigger for procrastinating selection satisfaction (Chesi, 2013; Chesi & Brattico, 2018).

[21] Technically speaking, a Last-In-First-Out memory buffer. This plainly implements locality conditions, shortest move and the Minimal Ling Condition (Chomsky, 1995, p. 311). As results, counting dependencies can be easily described, but not cross-serial dependencies (Stabler, 1997) and the grammar would be as powerful as a Context-Free-Grammar (Chesi, 2004). The careful reader should have noticed that SOV word order could pose a problem from this perspective. This problem dissolves both assuming that canonical base SOV ordering always requires case features on S and O that can be selected by the predicate, or that either S and O are licensed above the predicative nucleus as discussed in (Bianchi & Chesi, 2014). Successive cyclic dependencies and superiority effects can also be captured, but as noticed by den Dikken (2018:91–94), the original re-merge option into the left periphery of a selected phase/workspace (Chesi, 2004) would produce inconsistent predictions with multiple *wh*- movement. Partial spell-out of intermediate copies (Rugna, 2020) suggests that re-merge is necessary and den Dikken (2018) analysis is superior in capturing multiple, long- distance, wh- dependencies.

[22] As first noticed by (Richards, 1999), Move preempts Merge in a Top-Down derivation.

[23] Certain right-hand "adjuncts" can pose a problem for this account: (Chesi, 2007) tries to extend (Bianchi & Chesi, 2006) approach to true adjuncts that seems to involve a delayed expansion of a select feature (due to the heaviness of a PP), while others qualify as true arguments of the predicate (Truswell, 2007) and, in fact, license matrix wh-related gaps as in (i).a. As noticed by (den Dikken, 2018:97), under this analysis, their ability to license another legitimate gap, as in (i).b is unexpected (unless we assume that the temporal argument is not the last selected one, but a "delayed" nested argument, Chesi, 2013).

(i) a. who did John drive Mary crazy [talking to _]
b. who did John drive _ crazy [talking to _]

considered as a purely structural constraint: parasitic gap (P-gap) constructions seem able to recover a strong/configurational island violation (Engdahl, 1983). The contrast in (18) indicates that an offending gap in the subject island can be rescued by an extra licensed gap in the canonical object position:

(18) a. *[$_{matrix\_W}$ [$_{nested\_W}$ Which famous playwright]$_i$ did [$_{nested\_W}$ close friends of _$_i$] become famous]?

b. ?[$_{matrix\_W}$ [$_{nested\_W}$ Which famous playwright]$_i$ did [$_{nested\_W}$ close friends of _$_i$] admire [$_{sequential\_W}$ _$_i$]]?

This fact poses a serious challenge for theories that only rely on the status of the island phrase to explain the ungrammaticality of (18).a. On the one hand, it is attested that the subject can be transparent for extraction in specific contexts (Abeillé et al., 2020) and neither its base vs derived position (base, as in post-copular subjects vs derived as the subject of a transitive/unergative predicate) nor its role (internal, as is unaccusative vs external, as in unergative subjects) is sufficient to explain the relevant contrasts. To explain the transparency of certain pre-verbal subject positions without weakening the assumption that configurational islands are always nested workspaces, Bianchi & Chesi (2014) propose that transparent subjects are those that reconstruct (get interpreted into the predicate), while non-transparent ones are those "presupposed", namely selected, by an higher category that forces them to be interpreted directly in the highest (criterial, in Rizzi's sense) surface position[24].

This approach to reconstruction as "delayed interpretation" is coherent with another subtle linearization contrast:

(19) a. *the man who [pictures of _$_{DP}$] are on the table (Chomsky, 1986, p. 32)

b. he is the person of whom [pictures _$_{PP}$] are on the table

Preposition stranding, which is the default option in English, is unavailable when the subject "reconstructs" in its base position, where it becomes transparent for extraction. Bianchi & Chesi (2015) assume that the presence of the preposition indicates that the probe selecting the argument of "pictures" is expanded/merged in the surface position, leaving the subject no room either for discharging the pending wh-phrase as its argument (in its surface position the subject is a nested workspace), or waiting to be interpreted within the predicative nucleus, since the probe of the nominal head has been already processed, as testified by the presence of the preposition. This hypothesis shed some light on the long-standing puzzle on preposition linearization.

To summarize, in this probe-first (Top-Down) derivation, the movement trigger becomes the absence of a preceding probe (Chesi, 2004, 2012, 2015; Sobin, 2020). Coherently with (11).b, without need to bring up parsing and generation, the absence of the relevant licensing environment (a selecting head in Chesi's 2004; Stabler's 1997

[24] In a nutshell, this is the implementation of D-linking (Pesetsky 1987) in terms of presuppositionality (Bianchi & Chesi, 2014:534; Diesing, 1992): In i. the subject is presupposed, then it is fully interpreted in its surface position leaving only a coindexed variable (x) for filling the thematic role of the predicate "perfect". In ii. the subject is not presupposed, then it is reconstructed/interpreted in the predicative nucleus where it qualifies as a sequential workspace (then becoming transparent).

i. ?*[Of which masterpiece]$_i$ is [one reproduction _ $_i$]$_j$ absolutely [perfect x_j]?

ii. [Of which masterpiece]$_i$ is [one reproduction _ $_i$]$_j$ already [available _$_j$]?

terms, a probe in Cecchetto & Donati's 2010; Kayne's 2011 sense) triggers movement of the whole workspace headed by a lexical item (the wh-phrase or the DP subject, a nested phase in Bianchi and Chesi's terms) that will remain active in the matrix workspace as long a compatible probe is found[25]. Notice that this approach is not simply "parsing-oriented". From this perspective, licensing a presuppositional subjects requires a specific probe in the higher IP functional domain[26]. Picking up a different head, with different probes, would generate completely different derivations in a way that is more in line with the generation perspective than with the parsing one (the parser needs to figure out the relevant phonetically empty probes used by the generator).

Assuming that only lexical elements (N, V, and certain Adjectival/Adverbial heads) can introduce "left-peripheral" probes (then instantiating new workspaces/phases), we derive some important facts:

(20) functional projections are strictly ordered since each functional head probes the immediately lower one, necessarily non-recursively, unless a lexical item (N, V, or A) is probed, which can probe, in its turn, a higher functional category (explaining iteration, namely the simplest form of recursion – tail recursion, to be precise, Abelson & Sussman, 1996, p. 457).

(21) unselected features[27] force the whole phrase bearing them to remain available in the workspace until a probe for them is merged; in the meanwhile, the very same whole phrase qualifies as an independent nested workspace.

While (20) is a necessary step toward direct compositionality, to be explained in line with Ramchand & Svenonius (2014), (21) constitutes the only source of "true" recursion (vs. tail recursion) in phrase structure building from this perspective and plainly derives islandhood and p-gaps by reversing the classic perspective: the problem with configurational islands is not the non-extractability of a complement from the island phrase, but the absence of the legitimate gap in the matrix working space where the wh-item first merged (licensed in virtue of a probed *wh* criterial feature, but unlicensed as a D argument). The approach is an explicit derivational rephrase of the representational connectedness idea (Bianchi & Chesi, 2006; Kayne, 1984)[28]. Moreover, it can be interpreted as a Top-Down analysis of the Bottom-Up derivation discussed by Nunes & Uriagereka (2000) and presented in the next section.

---

[25] The Memory-buffer metaphor (Bianchi & Chesi, 2006, 2014; Chesi, 2004) might be inappropriate because of its similarity with Push Down automata and with much of the psycholinguistic literature in the processing literature. The intent there was purely theoretical and probably the notion of pending elements in the current workspace is the most accurate interpretation of this metaphor.

[26] SubjP in Cardinaletti (2004) terms. See also Shlonsky (2000).

[27] Possibly clustered with selected ones, forcing "pied-piping" in Chesi & Brattico's 2018 terms.

[28] Only the immediate superordinate workspace (W) is accessible to parasitic copy. This correctly excluded unwelcome generalizations forcing the intermediate W2 to request parasitic copies of the pending items if lower W3 needs them:

(i) *A person [$_{W1}$ who$_i$ you admire _$_i$ [$_{W2}$ because [$_{W3}$ close friends of _$_i$] became famous]

(ii) ?A person [$_{W1}$ who$_i$ you admire _$_i$ [$_{W2}$ because [$_{W3}$ close friends of _$_i$] criticize _$_i$]

4.4. Cyclic linearization: the notion of phase and workspace

Various intuitions adopted in the previous sections have been framed also in a more canonical Bottom-Up perspective. Two crucial concepts for understanding structural linearization constraints in mainstream minimalism are those of *phase* and *workspace*. The notion of phase is originally motivated, by complexity considerations (Chomsky, 2001): the "active memory" burden should be reduced, during the computation, by spelling-out (§34. Spell-out), cyclically (33. Cyclicity), the phrase structure built so far (Nissenbaum, 2000; Uriagereka, 1999). The domain of a cyclic spell-out, in the current theorizing, is then often associated to the notion of phase: a constituent with a certain degree of independence in terms of interface conditions is released from active memory after it has been delivered to PF and LF. In this chapter, I remained agnostic on the exact definition of the phrasal nodes that qualify as phases (see §13. Phase theory), and I have adopted the less controversial definition of "derivational workspace". This notion fits perfectly with the necessity of chunking the derivation into pieces with a certain degree of independence. Nunes & Uriagereka (2000) exploit the idea of the independency between two parallel workspaces to propose an elegant redefinition of the Condition on Extraction Domains (CED) effects (Huang, 1982) in derivational terms (see also Stepanov, 2007): the contrast in (18) can be captured, also from a Bottom-Up perspective, assuming that a "nested workspace" ("close friends of which famous playwright") is spelt-out before merging with the matrix workspace (i.e. "become famous"). Since spell-out literally send the phrase structure to PF for linearization, when the nested workspace merges with the matrix one, no possible re-ordering is available of the elements contained and these are frozen in that specific order. This approach has the advantage of dealing in exactly the same way (as predicted by CED) both with left-branching constituents and with right-hand adjuncts ("*Which paper did you read Don Quixote [before filing _]?"; see note 23): They are both frozen in a specific precedence order when they merge with the matrix workspace. This is not necessary for the direct object, in SVO languages, that lies on the matrix workspace in which the predicate is merged[29].

To explain parasitic gaps, as discussed in §4.3, the authors rely on a "defective Numeration" and assume that if in the matrix workspace a missing argument would lead the derivation to crash, then the system can re-use, via sideward movement (Nunes, 2001), the relevant wh- phrase merged within the adjunct/subject constituent that, in the meanwhile, was made prominent in virtue of "whatever feature is involved in successive cyclic A'-movement" (Nunes & Uriagereka, 2000, p. 30).

Alternatively, a more direct reference to the notion of phase to recast linearization constraints in derivational terms is proposed by Fox & Pesetsky (2005). Their approach to cyclic linearization crucially depends on the assumption than linear order is preserved during the derivation and "frozen" on a phase-basis.

[29] This implies that the matrix workspace is not spelt-out until the relevant reordering of the wh- phrase is realized. In successive cyclic movement, intermediate triggers must be explicitly postulated as in any bottom-up approach (e.g. "edge" features Chomsky, 2005). See Chesi (2007) for a discussion of the problems related to feature-driven successive cyclicity.

Interesting consequences of this idea are, for instance, the derivations of Holmberg's Generalization on Object Shift in Scandinavian languages and successive cyclicity: since ordering must be preserved, if a wh- item is not moved into the left periphery of each intermediate phase (e.g., <saw close friends of who>) a precedence order conflict would occur if the element reaches the left periphery (e.g. <who$_i$ do you think that John <saw close friends of who$_i$>>, since the lower copy of *who* follows <saw close friends of> while the higher copy precedes these words)[30]. Similarly, the well-known fact that Object movement is blocked in Swedish, (22).b, unless the verb moves with it (22).a, leftward, preserving the VO order (also in case V2 is satisfied by an auxiliary as in (22).c) is derived: since no trigger for O movement in the VP edge is present, the <V O> order in the VP phase must be preserved in the later phases, hence either both <V O> move or O remains in situ[31].

(22) a. Jag [$_v$ kysste] [$_o$ henne] inte [$_{VP}$ t_v t_o] (Holmberg, 1999)
I kissed her not

b. * … att Jag [$_o$ henne] inte [$_{VP}$ [$_v$ kysste] t_o]
… that I her kissed not

c. * Jag har [$_o$ henne] inte [$_{VP}$ [$_v$ kysste] t_o]
I have her not kissed

From the linearization perspective, both approaches (Nunes & Uriagereka and Fox & Pesetsky) are compatible (necessary) solutions for what we can call "the telepathy paradox", which I will introduce in the next section.

5. The telepathy paradox

Chomsky (1995:221) argues that "if humans could communicate by telepathy, there would be no need for a phonological component […] to the use of language generally […] inducing [requirements that are] departures from 'perfection'".

Taking this statement seriously, if we would not need to cope with the SM interface, one could imagine being able to transfer instantly one sentence structure into each other's mind. Nevertheless, instantaneous transmission would be useless: since sentences can be as long as we want (possibly infinite) either we assume that our brain has an infinite number of processing units (but this would obviate the necessity of a recursive mechanism or, at least,

[30] The precise definition of phase is crucial to derive the relevant restrictions here: if DP is considered a phase, then an intermediate representation of <saw close friends of who> should be, obligatorily, <who$_i$ <close friends of t_i>>.

[31] Again, the ban on V movement alone should be derived assuming precise phasal boundaries and selective edge features: to exclude "Jag kysste inte henne" (I kissed no her), assuming that the negation is outside the VP phase, no edge feature should be able to attract V alone to check V2 (this would produce no linear order conflicts). According to Holmberg, the judgments on this construction are subject to dialectal variations. This complicates the analysis possibly requiring a non-trivial parameterization of the VP edge features.

the notion of phase/workspace) or, in case of a finite set of processing units (the only viable option), the same units should be reused to process incrementally the sentence received. This is what we can dub the "telepathy paradox": if we could use telepathy, this would be useless in terms of instantaneous message transmission, since the processing bottleneck (simply expressed in terms of finitary processing units) will persevere in forcing an incremental, piecemeal coding-decoding procedure. We can conclude that "incrementality" is a virtual conceptual necessity and telepathy should necessarily cope with this exactly as SM does. Temporal flow seems then an independent constraint imposing more relevant asymmetries than the SM system. This is somehow implicit in the intuition behind the notion of phase, that should induce a "reduction of computational burden" (Chomsky, 2001, p. 11), but it has never been stated as a "virtual conceptual necessity". In the probe-first (Top-Down) perspective presented in §4.2-§4.3, the various approaches to linearization discussed, all share the same intuition that a minimal asymmetry in phrase structure building can be exploited to induce relevant local linearization constraints only relying on the time of application of each Merge operation. This strongly complies with the incrementality requirement in the strictest possible sense. This is by no means the only possible alternative. As summarized in §4.4, other proposals approach the linearization problem by relying on the phase/spell-out dynamics and they still comply with the incrementality requirement. These last perspectives are in line with the mainstream (Bottom-Up) definition of Merge, but they need to deal with technical details, both from the evolvability and learnability perspective (e.g., Numeration, sideward movement, the definition of Phase, the countercyclic nature of reconstruction) that seem less severely affecting the probe-first, Top-Down, perspective.

6. Conclusion

In these pages, I summarized some of the arguments suggesting a clear connection between asymmetries in phrase structure (asymmetric C-command: Kayne, 1994, 2018, 2020) and linear order, possibly reducible to the incrementality assumption (Phillips, 1996, 2003). Taking Kayne's (1994:36–38) argument (see note 14) to its logical conclusion, in a derivational perspective in which asymmetric C-command is rephrased as "being merged before in the same workspace", the result is a left-right derivation tout court. This shift in the derivational perspective has important consequences ranging from the explanation of conflicts in constituency tests to islandhood and p-gaps reinterpretation. This also impacted other empirical domains that, for the sake of space, cannot be properly presented here. These include crucial limitations of QR and rightward-movement (deriving for instance the "right-roof constraint" along the lines of Bianchi and Chesi, 2010 and Chesi, 2013), expletives (Sobin, 2020), and the role of Agreement in relation to "domain opacity" (den Dikken, 2018). Despite the fact that some of these phenomena can be accounted for also from the classic bottom-up perspective, grossly chunking the derivation by phase/workspace and assuming a cyclic linearization that must be preserved, it should be acknowledged that, compared to the Bottom-Up perspective, the incremental, probe-first, Top-Down one is the only tenable hypothesis for recasting crucial linearization constraints in core syntactic terms, while permitting

incremental generation of a sentence for which the end is not yet planned[32]. In this respect, memory fading is not just a matter of performance (either parsing or generation), but the direct effect of "an infinite use of finite means". Considering that the proposed expectation-driven reformulation of Merge and Move is totally transparent both for parsing and generation, the legitimate conclusion is that some core constraints on the precedence relation dubbed "linearization" are just the side effects of the incremental phrase structure building. In this sense, the probe-first (Top-Down) solution is also a viable candidate for a "genuine explanation".

References

Abeillé, A., Hemforth, B., Winckel, E., & Gibson, E. (2020). Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition*, *204*, 104293.

Abels, K., & Neeleman, A. (2012). Linear Asymmetries and the LCA: Linear Asymmetries and the LCA. *Syntax, 15*(1), 25–74. https://doi.org/10.1111/j.1467-9612.2011.00163.x

Abelson, H., & Sussman, G. J. (1996). *Structure and interpretation of computer programs*. The MIT Press.

Aboh, E. O. (2004). *The morphosyntax of complement-head sequences: Clause structure and word order patterns in Kwa*. Oxford University Press.

Ackema, P., & Neeleman, A. (2002). Effects of short-term storage in processing rightward movement. In S. Nooteboom, F. Weerman, & F. Wijnen (Eds.), *Storage and computation in the language faculty* (pp. 219–256). Springer.

Baker, M. C. (1988). *Incorporation: A theory of grammatical function changing*. University of Chicago Press.

Baltin, M. R. (2006). The Nonunity of VP-Preposing. *Language*, *82*(4), 734–766. https://doi.org/10.1353/lan.2006.0181

Barker, C., Jacobson, P., Jacobson, P. I., & others. (2007). *Direct compositionality*. Oxford University Press.

Barker, C., & Shan, C.-C. (2014). *Continuations and natural language* (First edition). Oxford University Press.

Belletti, A. (2004). *Structures and Beyond: The Cartography of Syntactic Structures, Volume 3*. Oxford University Press.

Bianchi, V., & Chesi, C. (2006). Phases, left-branch islands, and computational nesting. *Proceedings of the 29th Annual Penn Linguistics Colloquium*, *12.1*, 15–28.

[32] E.g., "And water came and extinguished the fire that burned the stick that hit the dog that bit the cat that ate the mouse that my father bought at the market." A. Branduardi, translation of the lyrics "Alla fiera dell'est" (At the Eastern Fair).

Bianchi, V., & Chesi, C. (2010). Reversing the perspective on Quantifier Raising. *RGG. RIVISTA DI GRAMMATICA GENERATIVA*, 3–38.

Bianchi, V., & Chesi, C. (2014). Subject islands, reconstruction, and the flow of the computation. *Linquistic Inquiry*, 525–569. https://doi.org/10.1162/LING_a_00166

Bianchi, V., & Chesi, C. (2015). On a PP/DP asymmetry in extraction. In *Structures, Strategies and Beyond* (Vol. 223, pp. 47–66). John Benjamins. https://doi.org/10.1075/la.223.03bia

Cardinaletti, A. (2004). Toward a cartography of subject positions. In L. Rizzi (Ed.), *The structure of CP and IP. The cartography of syntactic structures* (Vol. 2, pp. 115–165). Oxford University Press.

Cecchetto, C., & Donati, C. (2010). On Labeling: Principle C and Head Movement: On Labeling: Principle C and Head Movement. *Syntax*, *13*(3), 241–278. https://doi.org/10.1111/j.1467-9612.2010.00140.x

Chesi, C. (2004). *Phases and cartography in linguistic computation: Toward a cognitively motivated computational model of linguistic competence* [PhD Thesis]. University of Siena.

Chesi, C. (2007). An introduction to Phase-based Minimalist Grammars: Why move is Top-Down from Left-to-Right. In *STIL - Studies in Linguistics—Vol. 1* (Vol. 1, pp. 38–75). CISCL Press.

Chesi, C. (2012). *Competence and Computation Toward a Processing Friendly Minimalist Grammar*. Unipress.

Chesi, C. (2013). Do the 'right' move. In *Special Issue on Directionality of Phrase Structure Building* (pp. 107–138). CISCL Press.

Chesi, C. (2015). On directionality of phrase structure building. *Journal of Psycholinguistic Research*, 65–89. https://doi.org/10.1007/s10936-014-9330-6

Chesi, C., & Brattico, P. J. (2018). Larger than expected: Constraints on pied-piping across languages. *RGG. RIVISTA DI GRAMMATICA GENERATIVA*, *2008.4*, 1–38.

Chomsky. (1970). Remarks on Nominalization. In R. A. Jacobs & P. S. Rosenbaum (Eds.), *Readings in English Transformational Grammar* (pp. 184–221). Ginn.

Chomsky, N. (1957). *Syntactic structures*. Walter de Gruyter.

Chomsky, N. (1965). *Aspects of the Theory of Syntax* (Vol. 11). MIT press.

Chomsky, N. (1986). *Barriers* (Vol. 13). MIT press.

Chomsky, N. (1995). *The minimalist program*. MIT press.

Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1–52). MIT Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, *36*(1), 1–22.

Chomsky, N. (2020). *UCLA Lectures*. lingbuzz/005485

Christophe, A., Nespor, M., Teresa Guasti, M., & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: The case of the head-direction parameter. *Developmental Science*, *6*(2), 211–220.

Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press.

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

Cinque, G. (2005). Deriving Greenberg's Universal 20 and Its Exceptions. *Linguistic Inquiry*, *36*(3), 315–332. https://doi.org/10.1162/0024389054396917

Collins, C. (2002). Eliminating Labels. In S. D. Epstein & T. D. Seely (Eds.), *Derivation and Explanation in the Minimalist Program* (pp. 42–64). Blackwell Publishing. https://doi.org/10.1002/9780470755662.ch3

den Dikken, M. (2018). *Dependency and directionality*. Cambridge University Press. https://doi.org/10.1017/9781316822821

Diesing, M. (1992). *Indefinites*. MIT press.

Engdahl, E. (1983). Parasitic gaps. *Linguistics and Philosophy*, *6*(1), 5–34.

Fox, D., & Pesetsky, D. (2005). Cyclic Linearization of Syntactic Structure. *Theoretical Linguistics*, *31*(1–2), 1–45. https://doi.org/10.1515/thli.2005.31.1-2.1

Greenberg, J. H. (1963). *Universals of language*.

Grimshaw, J. (1991). Extended projection. *Ms., Brandeis University*.

Hawkins, J. A. (1983). *Word order universals*. Academic Press.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge University Press.

Holmberg, A. (1999). Remarks on Holmberg's Generalization. *Studia Linguistica*, *53*(1), 1–39. https://doi.org/10.1111/1467-9582.00038

Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar*. MIT.

Kandybowicz, J., & Baker, M. C. (2003). On directionality and the structure of the verb phrase: Evidence from Nupe. *Syntax*, *6*(2), 115–155.

Kayne, R. S. (1984). *Connectedness and binary branching*. De Gruyter Mouton. https://doi.org/10.1515/9783111682228

Kayne, R. S. (1994). *The antisymmetry of syntax*. MIT Press.

Kayne, R. S. (2018). The place of linear order in the language faculty. *Ms. New York University*. https://doi.org/lingbuzz/003820

Kayne, R. S. (2020). Antisymmetry and Externalization. *Ms. New York University*. https://doi.org/lingbuzz/005554

Kayne, R. S. (2011). Why are there no directionality parameters. *Proceedings of WCCFL*, *28*, 1–23.

Kempson, R., Meyer-Viol, W., & Gabbay, D. M. (2000). *Dynamic syntax: The flow of language understanding*. Wiley-Blackwell.

Larson, R. K. (1988). On the double object construction. *Linguistic Inquiry*, *19*(3), 335–391.

Momma, S., & Phillips, C. (2018). The Relationship Between Parsing and Generation. *Annual Review of Linguistics*, *4*(1), 233–254. https://doi.org/10.1146/annurev-linguistics-011817-045719

Momma, S., Slevc, L. R., & Phillips, C. (2018). Unaccusativity in Sentence Production. *Linguistic Inquiry*, *49*(1), 181–194. https://doi.org/10.1162/LING_a_00271

Nissenbaum, J. W. (2000). *Investigations of covert phrase movement* [PhD Thesis]. Massachusetts Institute of Technology.

Nkemnji, M. A. (1995). *Heavy pied-piping in Nweh*. UCLA.

Nunes, J. (2001). Sideward movement. *Linguistic Inquiry*, *32*(2), 303–344.

Nunes, J., & Uriagereka, J. (2000). Cyclicity and Extraction Domains. *Syntax*, *3*(1), 20–43. https://doi.org/10.1111/1467-9612.00023

Pesetsky, D. (1987). Wh-in-situ: Movement and unselective binding. In E. Reuland & A. G. B. ter Meulen (Eds.), *The representation of (in) definiteness* (Vol. 98, pp. 98–129). MIT Press.

Pesetsky, D. (1995). *Zero syntax: Experiencers and cascades*. MIT Press.

Phillips, C. (1996). *Order and structure* [PhD Thesis]. Massachusetts Institute of Technology.

Phillips, C. (2003). Linear Order and Constituency. *Linguistic Inquiry*, *34*(1), 37–90.

Phillips, C., & Lewis, S. (2013). Derivational order in syntax: Evidence and architectural consequences. *Studies in Linguistics*, *6*, 11–47.

Ramchand, G., & Svenonius, P. (2014). Deriving the functional hierarchy. *Language Sciences*, *46*, 152–174.

Richards, N. (2010). *Uttering trees*. MIT Press.

Richards, N. (2016). *Contiguity theory*. The MIT Press.

Rizzi, L. (1997). The fine structure of the left periphery. In *Elements of grammar* (pp. 281–337). Springer.

Rizzi, L. (Ed.). (2004). *The structure of CP and IP*. Oxford University Press.

Roberts, I. (2010). *Agreement and head movement: Clitics, incorporation, and defective goals*. MIT Press.

Roberts, I. (2017). The final-over-final condition in DP: Universal 20 and the nature of demonstratives. In M. Sheehan, T. Biberauer, I. Roberts, & A. Holmberg (Eds.), *The Final-over-Final Condition: A Syntactic Universal* (Vol. 76, p. 151). MIT Press.

Rugna, G. (2020). German wh-copying: A top-down analysis. *Quaderni Di Linguistica e Studi Orientali*, 187–219. https://doi.org/10.13128/QULSO-2421-7220-9700

Sheehan, M. (2013). Some Implications of a Copy Theory of Labeling. *Syntax*, *16*(4), 362–396. https://doi.org/10.1111/synt.12010

Sheehan, M. (2017). The Final-over-Final Condition and Processing. In M. Sheehan, T. Biberauer, I. Roberts, & A. Holmberg (Eds.), *The Final-Over-Final Condition: A Syntactic Universal* (Vol. 76, pp. 79–98). MIT Press.

Sheehan, M., Biberauer, T., Roberts, I. G., & Holmberg, A. (Eds.). (2017). *The final-over-final condition: A syntactic universal*. The MIT Press.

Shlonsky, U. (2000). Subject positions and copular constructions. In H. Bennis, M. Everaert, & E. Reuland (Eds.), *Interface strategies* (pp. 325–347). Royal Netherlands Academy of Arts and Sciences.

Sobin, N. (2020). Directing Syntactic Traffic. *Syntax*, *23*(3), 241–274. https://doi.org/10.1111/synt.12194

Stabler, E. (1997). Derivational minimalism. In C. Retoré (Ed.), *Logical Aspects of Computational Linguistics* (pp. 68–95). Springer Berlin Heidelberg.

Starke, M. (2004). On the inexistence of specifiers and the nature of heads. In A. Belletti (Ed.), *Structures and Beyond: The Cartography of Syntactic Structures* (Vol. 3, pp. 252–268). Oxford University Press.

Stepanov, A. (2007). The end of CED? Minimalism and extraction domains. *Syntax*, *10*(1), 80–126.

Travis, L. deMena. (1984). *Parameters and effects of word order variation* [PhD Thesis]. Massachusetts Institute of Technology.

Truswell, R. (2007). Extraction from adjuncts and the structure of events. *Lingua*, *117*(8), 1355–1377.

Uriagereka, J. (1999). Multiple spell-out. In S. D. Epstein & N. Hornstein (Eds.), *Working Minimalism* (pp. 251–282). MIT Press.

Whitman, J. (2005). Preverbal elements in Korean and Japanese. In G. Cinque & R. S. Kayne (Eds.), *The Oxford Handbook of Comparative Syntax*.

Zwart, J.-W. (2007). Some notes on the origin and distribution of the IPP-effect. *GAGL: Groninger Arbeiten Zur Germanistischen Linguistik*, *45*, 77–99.

Zwart, J.-W. (2009). Prospects for top-down derivation. *Catalan Journal of Linguistics*, *8*(1), 161. https://doi.org/10.5565/rev/catjl.146