

Syntax as Graph Theory

Diego Gabriel Krivochen

University of Reading, CINN

diegokrivochen@hotmail.com

Contents

1. Introduction.....	2
2. Definitions and generalities	14
3. Some notes on discontinuous dependencies	21
4. Comparing models.....	26
5. On English Complement Constructions.....	38
6. Some considerations about Clitic Climbing and Spanish auxiliary chains.....	48
7. On unexpected binding effects: a connectivity approach to Binding Theory	54
8. Complementation within the NP: on appositive relatives.....	60
9. Islands and lexical restriction: opening and closing scope	65
10. On MIGs and prizes (and embedding, and complements).....	69
11. Two kinds of ‘and’ and the Coordinate Structure Constraint	74
12. A small collection of ‘transformations’	79
13. Some open problems and questions	81
13.1 A brief note on Left and Right extractions.....	82
13.2 On deletion.....	84
14. (Some) conclusions.....	87
15. References.....	87
Appendix A: On some inadequacies of trees to represent Merge.....	98
Appendix B: What remains of the generative barrel.....	103

*Distincte de la logique, distincte de la psychologie,
la syntaxe n’a à chercher sa propre loi qu’en elle même.
Elle est autonome.*

Lucien Tesnière (1959) *Éléments de Syntaxe Structurale*. Chapitre 20, §21.

Abstract

In this work we explore the consequences of questioning and replacing long-held assumptions about the way in which syntactic structure is built. Rather than minimizing connectivity between nodes in order to get ‘unambiguous paths’, we will pursue the consequences of the idea that syntactic structure grows by maximally connecting sub-graphs in local domains. The result will be a model of grammar

in which aspects that have proven problematic for mainstream generative grammar (MGG), like discontinuity and crossing dependencies, have a natural place.

Keywords graph theory; discontinuity; multidominance; syntax

1. Introduction

Diagrams of structural descriptions for natural language strings may adopt many forms. A structural description for a string is an annotated analysis of said string, in which local hierarchical relations are indicated by various means: phrase structure grammars appeal to a combination of phrasal labels (NP, VP, and the like) and configurational information. These configurations are defined in mathematical objects which Postal (2010: 7) refers to as *L-trees*. An L-tree is a *set* of nodes and arcs (or ‘edges’, or ‘branches’), and should not be confused with *diagrams* of L-trees (which are merely pictures composed of lines and symbols; typographical objects rather than mathematical objects); as McCawley (1998: 47-48) warns. L-trees are *generated* (read: *recursively enumerated*) in the sense of Post (1944: 285-286) by a system of ordered rules operating over a set of symbols (an *alphabet*), which is customarily referred to as a *grammar*. The role of the grammar, in this view, is to recursively enumerate the grammatical strings of a language (Chomsky & Miller, 1963: 283). The full set of non-isomorphic L-trees generated by a grammar *Gr* (i.e., the full set of *structural descriptions*) defines the *strong generative capacity* of that *Gr* (Chomsky, 1965: 60). Frameworks that operate with rules of inference applying from axioms and yielding sets of strings are called ‘proof-theoretic’ or ‘generative-enumerative’ (Postal, 2010; Pullum & Scholtz, 2001 respectively). A *derivation* in *Gr* is the ordered sequence of strings obtained by applying the rules of *Gr* stepwise: these rules rewrite the symbols on the left-hand side of the rule as the symbols on the right-hand side sequentially (a symbol at a time). For example:

1) Alphabet: {S, A, B, P, Q, a, b, c}

Rules: S → A, B

A → a, P

P → c

B → Q

Q → b

Derivation: S

AB

aPB

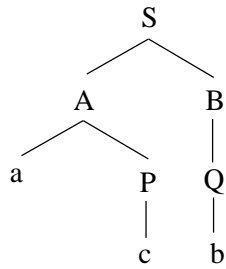
acB

acQ

acb

Now, given a derivation like the one above (which contains exclusively context-free *phrase structure rules* PSR), a *tree* can be constructed by relating nodes in any line to nodes in the line immediately above, such that the relation *rewrites as* becomes *is the mother of* (see McCawley, 1968: 245 for discussion and more details). The L-tree corresponding to the derivation above can be diagrammed as follows:

2)



A work as early as Bach (1964: 71) already formulates conditions on phrase markers (P-markers) in terms of their ‘topological’ [sic] properties:

A proper P marker (when represented in tree form) is a topological structure of lines and nodes conforming to the general requirement that a unique path be traceable from the termination of every branch to the point of origin of the whole tree (or for that matter from any node to any other node)

It is crucial to note that the requirement that there be unique, unambiguous paths (with the additional, but at this point covert, assumption that any node in such a path must be visited *only once*) is a staple of phrase structure in transformational generative grammar.

Formal grammars of the kind in (1) have been the object of studies since the very early days of generative grammar, and a big part of the early generative literature was devoted to proving the limitations and inadequacy of phrase structure rules to generate structural descriptions for all and only grammatical strings in natural languages (Chomsky, 1956, 1959; Postal, 1964, among others). To address the shortcomings of pure context-free PSRs, Harris (1952) and Chomsky (1955) introduced a further type of rule in the design of grammars for natural languages: *transformational rules*. These do not map symbols onto symbols, but rather *trees onto trees* (in what follows we will omit the *L*-qualification when talking about trees, presupposing it):

A transformational rule may be thought of as an ordered triplet $[T, B, E]$ consisting of an IC [Immediate Constituent] derivation tree T , a particular analysis of bracketing B of the last line of T , and an elementary transformation E indicating how the elements of B are to be converted so as to yield a new, derived tree T' (Lees 1976: 36)

More specifically, they map a *structural description* into a *structural change*, specifying the positions of terms before and after the transformation. Consider that each of the elements (or ‘terms’) of B is assigned a positive natural number (an ‘integer’), which corresponds to its place in the sequence, such that B is an ordered set $(X_1, X_2, X_3, \dots, X_n)$. Then,

*If the structural index of a transformation has n terms, a_1, a_2, a_n , it is a **reordering transformation** if its structural change has any a_i as its k^{th} term, or if a_i is adjoined to its k^{th} term, where $i \neq k$ (Ross, 1967: 427. Emphasis in the original)*

Transformations must always preserve the ‘topological’ properties of the P-marker: if branches cannot cross in a *kernel sequence*, they cannot be made to cross by means of a transformation; similarly, transformations cannot yield ‘ambiguous paths’, such that there is more than a unique path between any two nodes in the L-tree. In this sense, transformations are homomorphisms, as they preserve (topological) relations within selected structure. The generative power of transformations was greatly increased with the introduction and definition of *variables*, which can range over strings or arbitrary complexity (Ross, 1967). For example:

3) Interrogative:

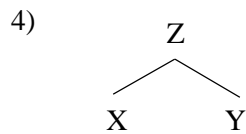
$$S[\text{structural}]D[\text{escription}]: \left\{ \begin{array}{l} \text{(a) } NP, C, VP \\ \text{(b) } NP, C + M, X \\ \text{(c) } NP, C + \textit{have}, X \\ \text{(d) } NP, C + \textit{be}, X \end{array} \right\}$$

$$S[\text{tructural}]C[\text{hange}]: X_1 - X_2 - X_3 \rightarrow X_2 - X_1 - X_3$$

[where C = Pres/Past and M = modal auxiliary]

Chomsky (1964: 227)

Derivations in transformational generative grammar instantiate a more general Immediate Constituent view on natural language structure (Wells, 1947; see Schmerling, 1983b for discussion) and the even more general view of language as a recursively-enumerable set (Pullum & Scholtz, 2001; Pullum, 2007). One of the roles of transformations is precisely to impose underlying continuity on superficial discontinuous dependencies (such as *verb-direct object* in a partial interrogative like *what did John buy?*, where the direct object *what* appears linearly away from its governor, the transitive verb *buy*), at the cost of multiplying the levels of representation in the grammar. The development of the theory of *displacement-as-movement* in the 70's and 80's gave rise to *traces* to mark the position *from where* a syntactic object had been displaced by a transformational rule (Fiengo, 1977), and later on, *copies* (Chomsky, 1995: 202)¹. Now, we need to note that transformational rules (as conceptualized in the Standard Theory, Government and Binding, and Minimalist frameworks) because they map (kernel) trees to (derived) trees, alter the configurational relations between nodes in these trees. In this sense, transformations are *not* homomorphic mappings: they do *not* preserve relations within selected structure; rather, they *create new relations*, deleting old ones. Let us illustrate. Imagine we have a phrase marker in which objects X and Y are in a local relation, as represented in (4):



¹ Strictly speaking, *trace theory* and *copy theory* are not always too different. The version of *trace theory* in Fiengo (1977: 44-45) decomposes *movement* as follows:

...movement of NP_i to position NP_j (where A and B are the contents of these nodes) in (30) yields (31) as a derived constituent structure.

$$(30) \dots NP_j \dots NP_i \dots$$

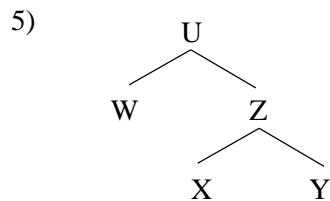
$$\begin{array}{cc} / & / \\ A & B \end{array}$$

$$(31) \dots NP_i \dots NP_i \dots$$

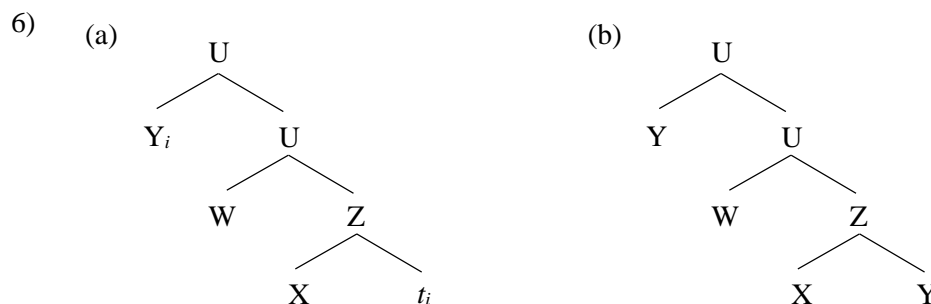
$$\begin{array}{cc} / & / \\ B & e \end{array}$$

On this view, NP_i and its contents are copied at position NP_j , deleting NP_j and A, and the identity element e is inserted as the contents of (in this case the righthand) NP_i , deleting B under identity.

Now suppose that there is some relation R between X and Y : for instance, say X theta-marks Y . That relation needs to be maintained throughout the derivation, or reconstructed at the level of semantic interpretation if disrupted by a reordering or deletion rule. We have seen some problems with the latter option, so we would like to give some general prospects to explore the former. Let us now introduce a further element in the derivation, W , which requires a local relation with Y in order to satisfy some requirement (which one in particular is not relevant for the present argument). W is external to $\{X, Y\}$, following a cumulative approach to derivational dynamics (which Chomsky, 1995: 190 encodes in the so-called ‘Extension Condition’):



But, what happens if a local configuration between W and Y is required (because, for instance, Y satisfies a criterial feature on W), and such relation cannot hold if X is in between? A *displacement-as-movement* approach can either (a) move Y to a higher position in the checking domain of W (extending U), outside the scope of X leaving a co-indexed trace behind (the so-called *trace theory*), or (b) *copy* Y and *re-introduce* Y in the derivation (the so-called *Copy Theory of Movement*, or *Copy+Re-Merge theory*; Chomsky, 2000; Uriagereka, 2002; Nunes, 2004 and much related work). Both options are diagrammed below:



In the most recent generative framework (the Minimalist Program), in which the *Copy Theory* emerged, derivations operate from the bottom-up, via a strictly binary, stepwise operation called *Merge*, which takes X, Y of arbitrary complexity, and forms an ordered set $\{Z, \{X, Y\}\}$, where Z is the *label* of $\{X, Y\}$ (more on this in Appendix A). Chomsky (1995: 225) presents the conception of the syntactic computation in the Minimalist Program as a procedure that maps an Array of lexical items to phonological and semantic representations. In this context, the operation *Merge* manipulates elements present in what is called the Numeration, a set of pairs (LI, i) , where LI is a Lexical Item and i is an index (specifically, an integer) indicating how many times LI is used in the computation. Chomsky (1995, 2000) correctly claims that the inclusion of *traces* in a derivation violates the Inclusion Condition (which bans the inclusion of elements that are not present in the Numeration during the derivation), and therefore, that they are to be replaced by Copies. However, as we observed in Krivochen (2015b), the operation *Copy* also introduces new elements in a derivation, provided that the copies are not present in the initial Lexical array or Numeration; therefore, copies also violate the Inclusion Condition if this condition is to be understood strictly: information cannot be deleted or lost, but it should also be impossible to add information (in the form of syntactic terminals, for

instance) that is not present in the Numeration, including distributional specifications for copies². If an element of arbitrary complexity is copied and then Merged, there is no reason to believe that element was present in the Numeration (unless we assume a massive amount of looking ahead that allows the Lexical Array/Numeration to see what is going to be the output of the derivation and thus have all needed elements, including copies, ready beforehand), it is therefore treated as a whole new element: after all, until we get to the interface levels, we have no possibility of establishing a referential connection with an object already introduced in the derivational space, except under special stipulations which depart from the simplest scenario and require justification. Conversely, if copies were indeed present in the NUM, the operation Copy would be superfluous, since there would be nothing to copy: all usable elements would be already predicted (somehow) in the NUM. It must be noted that copies are *not* identical elements: obviously they are distinct in their syntactic contexts (the structural relations they establish with their neighbours, what each copy is *mother of*, *daughter of*, *sister of*...), but also in their internal composition. LIs in Minimalism are assumed to be bundles of features (phonological, semantic, formal), which come in two variants: *interpretable* features can be read by the systems in charge of semantic and phonological interpretation, whereas *uninterpretable* features need to be discharged in the course of the derivation to prevent it from *crashing* (Chomsky, 1995: 232, ff.). This discharge requires a local relation between at least two categories sharing a feature, and if such a local relation does not hold in a tree, a term may be moved in order to create such a configuration (see Chomsky, 2000: 100-101). In the light of the Copy theory of movement, syntactic objects move up a tree by checking / valuating and discharging features that form lexical items and cannot be interpreted by the semantic or the morphophonological systems (so-called ‘uninterpretable’ or ‘formal’ features; see Chomsky, 1995: 276, ff.). This means that in any configuration, no two copies of a syntactic object are defined by the same feature specification if operations are indeed driven by the necessity to value/check features. Identity is not possible in a system that admits feature checking relations, since the feature matrix of an element varies as the derivation unfolds and copies are (Internally) merged in places in which they can value and erase uninterpretable features. The systems in charge of phonological and semantic interpretation (the ‘interfaces’) *cannot* establish a dependency between two objects, say, α and β , as in (7):

$$7) \quad \alpha = \{i-F_1, i-F_2, u-F_3, u-F_4\}$$

$$\beta = \{i-F_1, i-F_2, u-F_3, u-F_4\}$$

Where i = interpretable, u = uninterpretable.

because there is nothing inherent to α and/or β that suggests they are linked. Any link should be encoded as a diacritic of sorts (e.g., a referential index that is carried throughout the derivation, which would of course violate the *Inclusiveness Condition*).

² Stroik & Putnam (2013: 20) express a similar concern:

To “copy X” is not merely a single act of making a facsimile. It is actually a complex three-part act: it involves (i) making a facsimile of X, (ii) leaving X in its original domain D1, and (iii) placing the facsimile in a new domain D2. So, to make a copy of a painting, one must reproduce the painting somewhere (on a canvas, on film, etc.), and to make a copy of a computer file, one must reproduce the file somewhere in the computer (at least in temporary memory).

The nature and internal dynamics of such temporary memory is not addressed in MGG.

All in all, transformational grammar followed closely on the structuralist conception of clause structure in terms of immediate constituency, phrase structure, and dominance. But that conception was not unchallenged: as Perlmutter (1983: ix) points out, ‘*there are significant generalizations, both cross-linguistic and language-internal that can be captured in terms of grammatical relations but not in terms of phrase structure configurations*’. This work essentially agrees with Perlmutter’s insight, and explores an alternative to constituency-based frameworks (including all frameworks based on phrase-structure grammars), transformational and non-transformational. The shortcomings of assuming a unique, strictly binary and stepwise operation to generate structure via discrete recursive combinatorics have been analysed in past works (Krivochen, 2015a, b; 2016; 2017; Krivochen & Schmerling, 2016; Bravo et al., 2015, among others), and we will not repeat that case. In this work we will propose a radically different view of syntax, its role, and the kind of objects that we can use to assign strongly adequate structural descriptions to natural language strings while minimizing the primitives of the theory. We begin by considering the following quotation from McCawley (1982):

I will assume that the deepest relevant syntactic structures are ordered continuous trees and will investigate the possibility of discontinuity arising in the course of derivations through movement transformations that alter left-to-right ordering without altering constituency (McCawley, 1982: 94)

The possibility of having linear order divorced from constituency is essential, and we can go even further if we give up the requirement of structures being *trees* while keeping *order* as a basic admissibility condition for structural descriptions. In this sense, McCawley’s quotation is an excellent transition between traditional PSGs and the strongly representational, *derivation-less* approach that we will present.

Let us draw a direct comparison with PSG-based MGG to start making things explicit. In the view that we explore in this paper, the whole point of ‘syntax’ is not to not ‘form set’ (as Chomsky 1995: 243 suggests; see also Collins, 2017 and much related work; in this view the ‘syntax’ is a discrete combinatoric system) or recursively enumerate expressions; rather, the ‘syntax’ establishes connections between nodes and make the most out of the smallest amount of nodes by maximizing connectivity rather than introducing new nodes. Similarly, the whole point of ‘doing grammar’ is to fully specify these connections and formulate necessary conditions on the admissibility of structural descriptions. This is what we are trying to provide a map of: the full set of connections between semantically relevant syntactic objects in the structural description of a natural language string. In this sense, the theory we will sketch here applies *model theory* rather than *proof theory*:

- *Model theory* is concerned with finding interpretations for well-formed formulae (WFF henceforth) which make such formulae true: if an interpretation *I* makes a WFF *S* true, we say that *I* is a *model* of *S* (or, alternatively, that *I* *satisfies* *S*). Model theory is concerned with expressions, not with sets of expressions: an expression, but not a set of expressions, can be a model. Grammars, in this view, consist of finite sets of admissibility conditions (Pullum, 2007: 1).
- *Proof theory* is concerned with the enumeration of WFF by means of recursive operations implemented through a Turing machine. More often than not, these operations are combinatoric, and inspired in the syntactic rather than the semantic side of logic: a grammar is a set of rules that recursively enumerates the set of sentences which constitutes a language *L*. Proof-theoretic models of syntax adopt a *procedural* view, which translates in the central role of *derivations*.

Our view incorporates elements from both meta-theories: we formulate rules as statements about expressions, rather than production functions (thus, aligning with model-theoretic frameworks). However, because our focus is set on *descriptive adequacy*, we will make use of terminology and concepts from proof-theoretic theories, like the idea of ‘syntax’ being computational (Standard Theory, Categorical Grammar, and LFG), as well as insights from model-theoretic syntactic theories (Arc-Pair Grammar, Relational Grammar, and Dependency Grammar), like the idea of ‘grammar’ being a set of admissibility conditions over graphs. The fact that a(n admittedly strongly representational) theory that was devised in the generative-transformational tradition, namely Government and Binding (Chomsky, 1981 and much related work) could indeed be ‘translated’ into model theory (Rogers, 1998) reveals that we can indeed take advantage of the best of each. The theory presented here clearly contrasts with the modern Chomskyan perspective of linguistic structure built step by step by means of bottom-up discrete recursive combinatorics over strictly binary sets. A consequence of adopting a strong proof-theoretic stance is that, because the grammar is a set of production rules that recursively enumerate structural descriptions, there arises the question of where in the Chomsky Hierarchy the grammar is located: are natural languages finite-state? Context-free? Context-sensitive? Turing-complete? We have argued in past works that, empirically, the question makes very little sense, for descriptively adequate structural descriptions for natural language strings, which represent semantic dependencies between elements in strings, need not be confined to a single level of the Hierarchy. At times, restricting the generative engine (where, again, ‘generative’ is understood in Post’s sense) to a specific level in the Chomsky Hierarchy assigns too much structure to substrings (as is the case with CFGs and simple iterative patterns, see Lasnik, 2011; Krivochen, 2015a, 2016c); at times, that restriction falls short (as is the case with CFGs and crossing dependencies, see Joshi, 1985 and much related work). That empirical inadequacy of proof-theoretic grammatical systems, which are based on the theory of computable functions developed, among many others, by Alan Turing (1936) can be either avoided or circumvented. It can be avoided by radically changing the perspective and adopting a model-theoretic approach. It can be circumvented by allowing the grammar, conceived of proof-theoretically, to oscillate between different levels in the Chomsky Hierarchy when assigning structural descriptions to natural language strings. This latter option, which entails rejecting the function-based nature of syntactic computation that is at the heart of generative grammar, has been explored in past works (particularly Krivochen, 2017). In this paper we explore aspects of a graph-theoretically informed grammatical theory that aligns with the former: the conditions and definitions that we will propose are not transition functions that take a SD input and generate a SC output. This forces us to reinterpret the role of ‘transformations’, which we will keep as merely descriptive devices, pretty much in the same way Postal (2010) uses ‘passivization’, ‘raising’ or such terms. The names of constructions will be taken from the *aetas aurea* of the generative tradition, roughly the period identified with the Standard Theory and its expansions. What matters is that rules and constraints are not formulated in derivational terms: there is thus no notion of *rule ordering* or indeed of time (either real or proof-theoretic). This does not mean that rules and constraints do not interact; they do, but in a different way:

Taking any rule R as an implication of the form ‘A materially implies B’, R applies to any structure S if and only if R’s antecedent A is satisfied by S. That determines that S is well-formed only if it satisfies B as well (Postal, 2010: 7)

In the present view, the grammar is a finite set of admissibility conditions over graphs, which are structural descriptions of natural language sentences. A graph³ is a set $G = (V, E)$, where V is a set of *vertices* and E is a set of *edges*. $v \in V$ is a vertex, and $e \in E$ is an edge. An edge e joining vertices a and b is notated $e = \langle a, b \rangle$, and a and b are said to be *adjacent vertices*. The *neighbour set* of v is the set of adjacent vertices to v , usually notated $N(v)$, and the *degree* of v is the number of *edges* connected to it. Here we will indifferently refer to vertices as ‘vertices’ or ‘nodes’. Locally, our graphs are *connected*: within local domains (a notion that will be defined shortly), there is a walk from any point to any other point (we will see that these walks, in our model, need to be *trails*, technically).

Let’s give some more definitions, which will be essential when considering relations between sub-graphs in structural descriptions. A graph H is a *subgraph* of G iff $V(H) \subset V(G)$ and $E(H) \subset E(G)$. We will also need a definition of *irreducible graph*:

A connected graph on three or more vertices is irreducible if it has no leaves, and if each vertex has a unique neighbor [sic] set. A connected graph on one or two vertices is also said to be irreducible, and a disconnected graph is irreducible if each of its connected components is irreducible (Koyama et al., 2007: 35)

This does *not* imply, though, that irreducible graphs are *complete* (or *strongly connected*): a graph G is *complete* iff every vertex is adjacent to every other vertex (that is, if there is an edge between any and every pair of vertices; see also Gould, 1998: 10, who uses the term *strongly connected*).

Irreducible graphs can, however, be cyclic without being complete. It is in this respect that we find a first major difference between tree-based syntax and the kind of formalism we advance here: we do not *require* for graphs to be acyclic (cf. Huck, 1984: 64), but acyclicity may emerge depending on the specific connectivity patterns of a specific piece of data.

Given this scenario, we have two possibilities when formalizing the theory of grammar in the form of a set of well-formed graphs:

- 8)
- a) Tend towards maximizing connectivity (at the cost of giving up ‘unambiguous paths’ in c-command relations)
- b) Tend towards maximizing unambiguous paths (at the cost of introducing extra nodes)

³ Accessible introductions to Graph Theory, in which some of the concepts we use here are defined in more detail, are van Steen (2010), Wilson (1996), and Gould (1988).

(8 b) is the option chosen by most works on generative grammar since GB⁴, as well as LFG, HPSG, and (locally), APG. Since the base component of a generative grammar⁵ is a context-free grammar of the type $\Sigma \rightarrow F$ (rewrite a possibly unary sequence of non-terminal or ‘intermediate’ symbols as a - possibly null- sequence of terminal and/or non-terminal symbols; see e.g., McCawley, 1968), further restricted by axioms imposing binarity as a condition over the format of rules (Kayne, 1984; Chomsky, 1986 et seq.) such that F is *always* a string of *two* symbols, the following conditions ensue:

- a) Every node has only one mother
- b) Every branching node has at most two daughters

That means that any terminal (i.e., nonbranching) node t has a neighbour set $N(t) = 1$, and any nonterminal (i.e., branching) node n has a neighbour set $N(n) = 3$ (two daughters and a mother), apart from the root node S , whose neighbour set is $N(S) = 2$ (two daughters, but no mother). Given these *a priori* conditions (we call them *a priori* since they would be given by Universal Grammar, see Kayne, 1994; Chomsky, 1995), phrase markers grow introducing extra nodes in the form of nonterminals or empty categories (phonologically null terminals). The graphs generated by a generative context-free grammar are connected, rooted, and labelled binary branching trees. This view of structure building, while common to transformational and non-transformational models alike is perhaps best exemplified by Kayne (1984), in which explicit conditions pertaining to *unambiguous paths* in c-command relations between terminals and nonterminals are introduced into the grammar. Kayne’s relevant definitions are the following:

Let a path P (in a phrase structure tree T) be a sequence of nodes $(A_0 \dots A_i, A_{i+1}, \dots A_n)$ such that:

- a. $\forall i, j \ 0 \leq i, j \leq n \ A_i = A_j \rightarrow i = j$
- b. $\forall i \ 0 \leq i < n \ A_i$ immediately dominates A_{i+1} or A_{i+1} immediately dominates A_i (Kayne, 1984: 131-132)

Condition (a) amounts to saying that a path is a sequence of *distinct* nodes; there are no loops and a path does not double back on itself (which is consistent with the graph theoretic notion of *path*, which requires that no vertex be visited more than once, as well as with that of *trail*, which requires that no edge be walked on more than once). Condition (b) means that a path is a sequence of adjacent nodes. Let us now proceed to the definition of ‘unambiguous path’

An unambiguous path in T is a path $P = (A_0, \dots A_i, A_{i+1}, \dots A_n)$ such that:

⁴ Although we do have to note that Chomsky & Miller (1963) and Katz & Postal (1964) assume that Generalized Transformations apply to a *pair* of strings:

*The basic recursive devices in the grammar are the generalized transformations that produce a string from a **pair** of underlying strings* (Chomsky and Miller, 1963: 304. Our highlighting)

*The recursive power [of a generative grammar] resides in Generalized Transformations, i.e., those which operate on a set of **P**-markers [phrase markers] (**probably always two**) to produce a single new derived **P**-marker (...)* (Katz & Postal, 1964: 12. Our highlighting)

The roots of binarity, which is at the heart of unambiguous c-command paths and more recently also labelling, would then be much older than X-bar theory.

⁵ Or, more specifically, the *constituent structure* subcomponent of the base, since the Lexicon (i.e., the alphabet of allowed terminal and non-terminal symbols) was also part of the base in the *Aspects* model and later incarnations of the generative-transformational theory.

$\forall i \ 0 \leq i < n$

- a. If A_i immediately dominates A_{i+1} , then A_i immediately dominates no node in T other than A_{i+1} , with the permissible exception of A_{i-1}
- b. If A_i is immediately dominated by A_{i+1} , then A_i is immediately dominated by no node in T other than A_i , with the permissible exception of A_{i-1} (Kayne, 1984: 132)

Kayne's conditions forbid upward branching (the so-called 'Single Mother Condition' SMC), non-binary branching, and –when labels are introduced-, discontinuity. In this context, the 'permissible exception of A_{i-1} ' in (b) is only added 'for symmetry' [sic], because it *never* comes into play (as the SMC is respected throughout).

In Kayne's vision, representative of the GB/MP tradition, terminals need to be nodes with degree 1, and nonterminals are nodes with degree 3, except for the root which has degree 2 (as seen above). Moreover, in the more recent (1994) version of the theory, the most deeply embedded node in a tree T must be an empty category (either a trace or a base-generated e) so as to comply with the so-called Linear Correspondence Axiom: $d(A)$ is a linear ordering of T . [for A a set of non-terminals, T a set of terminals, d a terminal-to-nonterminal relation] (Kayne, 1994: 6). Here, d is a syntactic relation known as asymmetric *c-command*, applying to the set of nonterminals. In simple terms, the LCA states, on the weak interpretation, that 'when x [asymmetrically] *c-commands* y , x precedes y ' (Uriagereka, 2012: 56), and on the strong interpretations, that x precedes y *iff* x [asymmetrically] *c-commands* y , *c-command* being both *necessary* and *sufficient*. The resulting model of phrase structure is one in which *each and every object is of one of these types*⁶:

- a. A terminal node
- b. A nonterminal node dominating two terminals (one of which must then be a *trace* in the sense of Fiengo, 1977⁷)
- c. A nonterminal node dominating a terminal and a nonterminal

We have provided empirical arguments against that rigid conception of phrase structure, based on mixed computational dependencies being established within local derivational units in natural languages in previous works (Krivochen, 2015a, 2016; Bravo et al., 2015; Krivochen & Schmerling, 2016, among other works). 'Mixed computation' means that

phrase markers, as structural descriptions of strings, are not computationally uniform but mixed. By saying that a system is 'computationally mixed', we mean that the structural descriptions assigned to substrings in $L(G)$, for L an enumerable set of strings and G a grammar [...], need not all be formally identical. [...] a computationally mixed system assigns

⁶ As David Medeiros (p.c.) pointed out to us, there is nothing in the LCA that prevents a fourth scenario (*d*) *A nonterminal node dominating a single terminal*. However, if the LCA applies to structures generated via Merge (Chomsky, 1995), unary branching is banned on independent grounds: Merge is by definition binary (Chomsky, 1995: 226; 2000: 81, and much subsequent work, even going back to Chomsky, 1955; see also Collins & Stabler, 2016, *Definition 13*).

⁷ In a theory that models *displacement* via literal *movement* of constituents, the notion of 'trace' is defined as follows:

Let us call the position from which movement occurs [...] the trace of the node that moves, and let us define proper binding as a relation that holds between a node and its trace only if the node precedes its trace (Fiengo, 1977: 45)

a substring the simplest structural description that captures and represents the formal and semantic relations between syntactic objects. (Krivochen & Schmerling, 2016: 36)

This proposal comes as an answer to a computational conundrum: structural descriptions must be *minimally appropriate*, that is, they should assign no more structure than strictly needed for substrings in L. But phrase structure grammars are in a very specific sense *procrustean*, only allowing for a single kind of computational dependency in structural descriptions for natural language strings: those allowable in Context-Free languages, nothing more and –crucially- nothing less. Because of this, it has long been recognized that they can be too powerful:

a constituent-structure grammar necessarily imposes too rich an analysis on sentences because of features inherent in the way P-markers are defined for such sentences. (Chomsky, 1963: 298. Highlighting ours)

But they can also fall short when assigning structural descriptions to natural language strings (Joshi, 1985). So, what can be done about this? A possibility that has just begun to be explored is that linguistic computation is an *oscillatory process*, which moves up and down the Chomsky Hierarchy in local domains. Some antecedents of this view can be seen in works by Howard Lasnik and Juan Uriagereka (see also Saddy, 2017; Saddy & Krivochen, 2017):

[...] what we need should be, as it were, ‘dynamic flatness’. But this is the sort of concept that sounds incomprehensible in a classical computational view, while making sense to those for whom syntactic computations are psychologically real. (Lasnik & Uriagereka, 2012: 21)

*In a manner of speaking, what we really want to do is **move down** the [Chomsky] hierarchy. Finite-state Markov processes give flat objects, as they impose no structure. But **that is not quite the answer either**. While it would work fine for coordination of terminal symbols, phrases can also be coordinated, and, again, with no upper bound.* (Lasnik, 2011: 361)

In this work we pursue a line of thought that is an even more radical departure from Immediate Constituency tenets than the works referred to above. We eliminate empty nodes (including traces), restrictions on the degree of vertices, and the requirement of graphs *minimizing* connectivity (*as per* the SMC and strict binarity requirements), thus proposing a theory aligned with tendency (1 a).

We have formulated (8 a) and (8 b) as ‘tendencies’, because derivations as global processes can dynamically oscillate between those extremes; this is another way of capturing the computationally mixed properties of natural language structures. This view of grammar as a dynamical oscillatory process yields graphs that can be evaluated as maximizing or minimizing connectivity *only within local domains*. That means that a model that embraces *n*-ary branching, discontinuity and multiple motherhood does *not* necessarily need to reject binarity and monotonicity altogether: these are properties of phrase markers *at a local level*, whereas computational mixture is a global property of structural descriptions. More on this below.

Formalisms that position themselves at the (8 b) end of the spectrum, which include transformational generative grammar as well as LFG, HPSG and CG, assume -explicitly or implicitly- what Sampson (1975) calls the ‘Single Mother Condition’:

D is a set of NODES, α is a function from D into a vocabulary of symbols (if $\alpha(d) = s$ we say that s is the LABEL of d), and δ is a partial function from D into strings over D (...)

(iii) for any d, d' in the domain of δ , if $\delta(d) = e_1, e_2, \dots, e_n$, $\delta(d') = e'_1, e'_2, \dots, e'_n$, and $e_i = e'_i$, for some $1 \leq i \leq n$, $1 \leq i' \leq n'$, then $d = d'$ and $i = i'$. That is, nodes may not branch upwards. We shall call property (iii) the *SINGLE MOTHER CONDITION* (SMC).

For LFG, Bresnan et al. (2016: 46) claim that c-structures (which are generated by means of a CFG) also obey the SMC. As for CGs, Partee (1974), among others, argues that traditional CGs have the strong generative power of CFGs: it is possible to use a ‘pure’ [sic] CG (i.e., Ajdukiewicz / Bar-Hillel style) as a model for the base component of a transformational grammar (Partee, 1974: 7). The general CF format of CG rules is translated into Chomsky Normal Form as follows:

$$9) \quad c \rightarrow (c/c_1 \dots c_n) + c_1 + \dots + c_n \text{ (Lewis, 1972)}$$

Which is a Context-Free rule (things may not be all that simple, though: not all variants of CG have the same generative power. See Van Benthem, 1988 and Buszkowski, 1988 for detailed discussion about the generative power of different variants of CG, including Ajdukiewicz-style and Lambek calculus). In CGs the SMC is respected, although diagrams have a much less significant role (if any at all) than in PSGs.

A similar condition to the SMC was initially assumed by McCawley (1968) in his formalization of the base component of a transformational generative grammar. Let x and y be nodes, ρ be a dominance relation between nodes, and λ be a precedence relation. Then:

if $x\rho y$ and $x'\rho y$, then $x = x'$

*for any two nodes x and y , if $x \neq y$, then either $x\rho^*y$ or $y\rho^*x$ or $x\lambda y$ or $y\lambda x$ (see also Wall, 1972: 149 for an equivalent condition)*

Both Sampson (1975) and McCawley (1982) argue against SMC-based theories (see also Jacobson, 1977), yet they do not go as far as proposing a theory of the (8 a) type. The expansion of phrase structure grammars in Peters and Ritchie (1981), called ‘Phrase Linking Grammars’ also allows for multidominance *locally*, and is thus closer to (8 a) than MGG and LFG. Arc-Pair Grammar (APG henceforth; Johnson & Postal, 1980) and Metagraph Grammar (Postal, 2010: 20) also reject the SMC. We share some of their arguments and conclusions, but not all; importantly, our proposal is not equivalent to theirs.

Since LSLT (Chomsky, 1955), a generative grammar –in principle, any fully explicit grammar- has incorporated a *transformational component* which mediates between kernel sequences (which are derived only by means of phrase structure rules) and morpho-phonological interpretation. Here, as will become obvious below, we will take *transformations* to have an exclusively *descriptive* value. In this sense, we would like to propose that *transformations*, i.e., mappings from *structural descriptions* to *structural changes* using traditional terms, can be divided in two types:

- 10) a. Transformations that introduce new vertices and new edges
- b. Transformations that introduce new edges between already existing vertices

And that this division corresponds to McCawley’s, in the sense that transformations of the type (10 a) also change constituency (‘relation-changing transformations’ for McCawley, RCT henceforth); whereas transformations of type (10 b) do not (‘relation-preserving transformations’, RPT henceforth, an obviously derivative term). In this work, we are *not* concerned with the linearization of graphs, our goal is to provide exhaustive descriptions of connectivity patterns in representations (thus, strictly speaking, it would be a bit misleading to use McCawley’s term *order-changing transformations* for transformations of the type (10 b)). By ‘constituency’ –above- we simply mean ‘grammatical

relations’ (but see below for some discussion): if relations are indeed defined in a pre-transformational structure, as in Chomsky (1965: 71), ‘be in construction with’ is equivalent to ‘be F of’, where F is a grammatical function or relation (‘subject of’, ‘object of’, etc.). But there is no need to resort to base structures (which need to be reconstructed when linearity is disrupted) if grammatical relations are in fact primitives: MGG has the X-bar format (or some version of that, more or less representationally oriented) as a primitive given by UG, and grammatical relations are defined over structural templates generated by X-bar. On the contrary, Relational Grammar (and its descendants) takes grammatical relations as primitives, and eliminates derivations (and the need for triggers for derivational operations with them; that means no features, no Agree, no Merge, etc.). There is still the question of how to link roles with lexical items, and that is a question that we will return to below.

2. Definitions and generalities

A grammar, we said, is a finite set of conditions over what is an expression of the language: well-formed expressions are models of the language. Thus, the conditions that the structure of expressions must satisfy need to be made explicit. In this section we present some (semi-formal) definitions and conditions over allowable structures, to be complemented by conditions over possible dependencies within graphs, to be explored in detail in the following sections:

1. Let X, Y, \dots, n be variables over sub-graphs, where each sub-graph is understood as a connected set of vertices v by means of edges e . We distinguish vertices by using numerical subscripts: $X = (v_1, v_2, \dots, v_n)$. The subscripts v_x, v_y, v_z, \dots will be used to denote variables over vertices. We will use ‘vertices’ and ‘nodes’ indistinctly.
2. Edges specify the *directionality* of the connection: $e \langle v_1, v_2 \rangle$. Thus, $e \langle v_1, v_2 \rangle \neq e \langle v_2, v_1 \rangle$. If a graph contains a function i from $E \rightarrow V \times V$, i is called an *incidence function*, and the graph is called a *digraph* (short for *directed graph*).

In this sense, our edges are not unlike *arcs* in Arc Pair Grammar (Johnson & Postal, 1980):

An arc represents the existence of a relation at a nonnull [sic] sequence of levels (Johnson & Postal, 1980: 37)

We assume throughout that there is *only one* level at which relations are established, because essentially what we are doing is creating a full map of the state of the derivational engine generating natural language structures conceived as a dynamical system at a given point in time (more on this below). Our graphs might be thought of as *exhaustive descriptions of ‘snapshots’ of a dynamical system*, once it has reached a stable state.

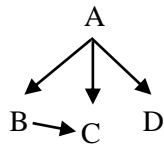
The reader will note that whereas all our graphs are *connected* and *directed* (thus, strictly speaking, we should call them *digraphs*, but we will stick to just ‘graphs’), not all of them are *strongly* connected (only *irreducible graphs* are), where a graph G is *strongly connected* iff there exists a directed path between any pair of distinct vertices in G (Gould, 1988: 10; Van Steen, 2010: 61).

3. Let ρ be a binary relation ‘immediately dominates’. Let ρ^* be a binary relation ‘transitively dominates’, where v_1 dominates v_2 iff there is some $e \langle v_1, v_2 \rangle$ (i.e., there is a directed edge from v_1 to v_2)
4. The ρ -domain of a vertex v is the set $\{v_1, v_2, \dots, v_n\}$ that v dominates either directly or transitively. The ρ -set of a graph is the full set of dominance relations between vertices in that graph.
5. Let τ be an n -ary relation ‘sister of’, as an abbreviation for:

$$11) \tau(v_1, v_2, \dots, v_n) \text{ iff } \exists (v_x) \in X \mid v_x \rho(v_1, v_2, \dots, v_n)$$

Here we do not require that sisters do *not* dominate each other, which seems counter-intuitive from the perspective of tree-based syntax. Thus, in the following treelet (where directed edges have been marked with arrows)

12)

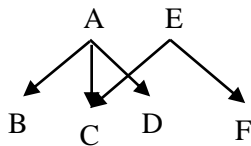


We have that

$$\rho = \{(A, B), (A, C), (A, D), (B, C)\}$$

And, since $\text{Ap}(B, C, D)$ holds, we say that B, C, and D are sisters because they share the mother A, regardless of the fact that $(B, C) \in \rho$. In a slightly different case:

13)



B, C, and D are sisters by virtue of sharing the mother A, and C and F are sisters by virtue of sharing the mother E. We might notate this using a subscript on τ , such that $\tau_A(B, C, D)$ is the set of nodes that are sisters because they share the mother node A.

$$6. \quad X \subset Y \text{ iff } \exists (v) \in Y \mid y\rho^*X$$

$$7. \quad \exists (v_0) \in X \mid \forall (v_x) \in X \ v_0\rho^*v_x \text{ (i.e., sub-graphs are rooted; adapted from McCawley, 1982: 93)}$$

This definition does not imply that there is always a *unique* root in a graph, as a matter of fact, the second sample graph above has two roots (A and E). This diagram is an example of what Morin & O'Miley (1969: 182) call a *vine*: a directed, acyclic, labelled, rooted graph.⁸ However, in the definition of linguistic *cycles* the restriction of the existential quantifier in *definition 7* from *at least one* to *one and only one* is relevant: a *cycle* is, in purely configurational terms, a *single-rooted graph*. Multi-rooted graphs, thus, are obtained by relating cycles: concretely, by conjoining transformations (in the sense of Fillmore, 1963) and identification of common vertices. In this work, we will use *cycle* in the sense that it has in generative linguistics since the 1960s; in graph theory, a *cycle* is a closed path (Gould, 1988: 10), but we will not use this definition. The condition that a *single-rooted graph* be a *cycle* becomes important in a strongly cyclic model in which sub-graphs connect but do define opaque domains, if connections occur *at the root*. We will come back to this below.

$$8. \quad \text{A relation } \mathbb{R} \text{ between sub-graphs } X, Y, \dots, n \text{ is total and transitive, but not antisymmetric (see also Ojeda, 1987: 258).}$$

$$9. \quad X \text{ is 'in construction with' } Y \text{ if either (a) } \exists (Z) \mid Z\rho X \wedge Z\rho Y \text{ or (b) } X\tau Y.$$

⁸ We need to note that Morin and O'Miley's work was primarily focused on the inadequacies of single-rooted trees headed by a single performative verb in an early Generative Semantics framework (that is, they argue against representations of the kind assumed by Ross, 1970); their concern was more *semantic* than *syntactic* (in the contemporary use of the terms). The consequences that their formal framework have for syntax are as far-reaching as they are usually ignored (a point also made in Postal, 2010: 394).

10. Let π be an ordering of X , such that $\pi(X)$ is the (possibly unary) set of *trails* within X starting at the root and following the direction of edges.

The notion of *trail* is weaker than that of *path*: a *path* is a walk in which no *vertices* are repeated, whereas a *trail* is a walk in which no *edges* are repeated (Gould, 1988: 9; Wilson, 1996: 26; Van Steen, 2010: 37). Because we will make use of repeated visits to certain vertices, as a crucial aspect of our model, we settle here for the weaker notion of *trail*.

11. A vertex v_1 is *ordered* with respect to v_2 iff either $\rho(v_1, v_2)$ or $\rho(v_2, v_1)$ or $\rho^*(v_1, v_2)$ or $\rho^*(v_2, v_1)$. If v_1 is ordered with respect to v_2 , either v_1 is in the ρ -domain of v_2 or v_2 is in the ρ -domain of v_1 .

Note that the notions we have introduced here use only very basic set-theoretic notation. For the phenomena analyzed here, we will see that it is enough. Unlike some previous graph-theoretic or geometrical analyses of syntax (e.g., Kracht, 2001; Beim Graben & Gerth, 2012) we do *not* attempt to encode GB/MP-type derivations in graphs, because –as we have seen– we depart from most basic assumptions in that framework (monotonicity, the SMC, movement). We will thus not define notions that are crucial for GB/MP approaches to syntax (e.g., *c-command*, or *chain*), but which play no role in a representational model (on the nature of the representations assumed here, see below). The framework presented here is orthogonal to MGG in more ways than one. Generative Grammar is an Item-and-Arrangement (IA) framework, in which endocentric constituent structure is monotonically built by means of discrete recursive combinatorics. Moreover, MP-style formalisms follow two additional conditions: the *No-Tampering Condition* NTC (operations over X and Y , for X and Y syntactic objects, leave X and Y unchanged; Chomsky, 2007: 8) and the Extension Condition EC (prevent counter-cyclic operations; Chomsky, 1995: 190). IA frameworks must be distinguished from Item and Process (IP) grammars, in which

A language L is a system consisting of the following:

- 1) *an algebra consisting in a non-empty set A of expressions and a (possibly empty) indexed set of operations defined over A ; A is the smallest set containing as members all the basic expressions of L and closed under the operations,*
- 2) *a Lexicon, or set of basic expressions indexed by a set of category indices, and*
- 3) *a set of syntactic rules that recursively assign any derived expressions of L to indexed syntactic categories. An n -place rule is a triple whose first member is the index of an n -place operation, whose second member is an n -place sequence of category indices (those of the inputs to the rule), and whose third member is the index of the output category of the rule. (2) and (3) constitute a recursive definition of the subset of A that is syntactically well formed (Schmerling, 1983a: 395)*

The proposal we put forth here is somewhat anomalous in the syntactic landscape. In this work we remain agnostic with respect to the IA vs. IP debate (because graphs can be used to represent diagrams corresponding to structural descriptions in either kind of theory), make no use of endocentricity (against a generalization of which we have argued in past works, particularly Krivochen 2015a) or structure building by recursive discrete combinatorics. Furthermore, because we allow for multidominance, discontinuity, *and* indirect loops, both the NTC and the EC are violated; the theory advanced here contrasts heavily with GB/MP-inspired MGG. We will see that there are some simplifications on representations assigned to natural language strings that follow directly from these violations.

At this point, we need to stress that *discontinuity* and *multidominance* have to be carefully distinguished. Discontinuity can be captured without multidominance (an example would be transformational accounts of Right Node Raising, as in Postal, 1997), and in the same way, multidominance structures do not necessarily allow for discontinuity: it is possible to have multidominance theories of Immediate Constituency grammars (e.g., TAGs with links, as in Joshi, 1985; see also the multidominance expansion of GB grammars in Kracht, 2008). In this work, we allow for *both*; this separates the proposal made here from MGG-inspired graph-theoretic works⁹, and puts us closer to APG and Metagraph Grammar (Johnson & Postal, 1980; Postal, 2010). Furthermore, multidominance does not necessarily generate *cyclic graphs*, as can be seen in the sample multidominated trees after (3): a multidominated node can have two mothers in different treelets (each mother can be the root of said treelet, as in our example), in which case the graph is directed and *acyclic*. A *directed cyclic* graph can be *rooted* if we have a designated node which is not dominated by any other node, in the sense that there is no edge *towards* that designated node (as we defined *root* above). If there was no encoding of *directionality* in edges, we would have to give up either roots or cyclic graphs. Thus, *discontinuity*, *multidominance*, and *cyclicity* (in the graph-theoretic sense) must be carefully kept apart.

Nodes in our graphs, relevantly, do *not* correspond to terminal or nonterminal nodes (or lexical items): rather, they correspond to the *translation of terms into intensional logic*, following Montagovian practice (Montague, 1973: 25-26, ff.; also Schmerling, 1988). Thus, definitions 1 and 2 above should incorporate the following corollary:

Corollary:

Variables over sub-graphs and nodes belonging to sub-graphs stand for their translation into *intensional logic*. Thus,

$d_x \in X$ is semantically interpreted as $\lambda PP \{^{\wedge}d_x\} \in \lambda PP \{^{\wedge}X\}$ or, abridged, $d_x' \in X'$

$\lambda PP \{^{\wedge}d_x\}$ is in turn an abbreviation (omitting parentheses) of $\lambda P[[^{\vee}P](^{\wedge}x)]$. The lambda tells us that this denotes a function from things of type P, which is the type of properties (which are the intensions of (the propositional functions of) sets) to propositions, which in turn consist of a set variable ($^{\vee}P$ denotes the *extension* of P -of the type of the characteristic function of a set- applied to the intension of x . This means that we have a function that is looking for a set (something of type $^{\vee}P$) as its argument.

Now, because we are not concerned with issues of ‘reference’ in any sense that is external to the algebra itself, there is no need to differentiate between *de re* and *de dicto* readings when it comes to the translations of the nodes: as far as the present model is concerned, there is only one graph (or, equivalently, two *isomorphic* graphs¹⁰) corresponding to (14):

⁹ The present paper also differs from the more mathematically oriented works in dealing more with concrete linguistic examples rather than with proving theorems about the structures generated by the system. For the linguist, that might be a relief. For the mathematician, a nuisance. The reader is encouraged to compare the presentation here with the rigorous axiomatization of Copy-Chain and Multidominance Structures in Kracht (2001).

¹⁰ Where *isomorphism* is defined as follows:

14) John seeks a unicorn (Montague, 1973: 22)

The two corresponding interpretations, namely:

- 14') a. There is a specific unicorn such that John seeks him
b. John seeks any entity belonging to the set of unicorns

Do not seem to require us to change the dominance relations in the graph, such that in both cases we have

15) $\rho = \{(\text{seek}, \text{John}); (\text{seek}, \text{unicorn})\}$

Where *John* and *unicorn* stand for $\lambda\text{PP}\{\wedge\text{John}\}$ and $\lambda\text{PP}\{\wedge\text{unicorn}\}$ respectively. This is so because we are modelling *syntax*, not *logical form*. Our syntax is not multi-layered, which does not mean that there is only one kind of representation corresponding to each sentence: the reader may think of the present work as a specification of the *c*- and *f*-structures of the structural description corresponding to a given string (in LFG terms); or a full specification of syntactic structure and grammatical function tiers in *Simpler Syntax* (Culicover & Jackendoff, 2005); or a map which provides all relevant information about theta structure + surface structure + predicate structure in terms of configurational properties in graphs (three of the levels in Williams' 2003 *Representation Theory*), etc.. The theory of the grammar *may* (but *need not*) be multi-layered; what is crucial here is that *the syntactic component of the theory of the grammar is not*.

Similarly, we are not concerned with the phenomenological existence of unicorns. We only care, for purposes of the formalization in this work, about the connectivity patterns that we can find and model in natural language sentences. What those sentences are used for fall outside the scope of this work.

The fact that vertices do *not* correspond to 'lexical items' or 'phrases' (or 'constituents', in Item-and-Arrangement grammars of the kind described in Schmerling, 1983a, b), however that turns out to be encoded, will be essential in the formulation of transformations and the analysis of crossing dependencies, below. In this model, vertices correspond to *basic expressions*, which, as we will see, are *not* equivalent to *words* or *phrases*, being defined in a different kind of grammatical system (Item-and-Process grammars; see Hockett, 1954, Schmerling, 1983a, b for discussion). This aspect sets our proposal apart from other graph-based theories, including Relational Grammar (see Perlmutter, 1983) and its heirs, Arc Pair Grammar (Johnson & Postal, 1980) and Metagraph grammar (Postal, 2010): in these, nodes represent 'substantive linguistic elements' defined in terms of their lexical or phrasal status (although this is often not explicitly said). The choice of what counts as a suitable element to be a *vertex* or *node* in an L-graph (be it a minimally connected tree or a network) impacts on the generative power and descriptive adequacy of the theory, as we will see in some detail below.

The core idea that we want to put forth is that most 'transformations', understood here as *purely descriptive devices* (see fn. 10), actually *do not change relations*: at most, they *create* new relations, but *without disrupting those already existing*. In order to describe these relations, locally and globally, we make use of the tools that graph theory puts at our disposition, plus the empirical insights obtained through careful analysis in both transformational and non-transformational theories.

Consider two graphs $G = (V, E)$ and $G^* = (V^*, E^*)$. G and G^* are **isomorphic** if there exists a one-to-one mapping $\phi: V \rightarrow V^*$ such that for every edge $e \in E$ with $e = \langle u, v \rangle$, there is a unique edge $e^* \in E^*$ with $e^* = \langle \phi(u), \phi(v) \rangle$ (van Steen, 2010: 33)

The grammar, in the present conception, aims at maximizing connectivity and minimizing the number of *distinct* nodes within local domains, and not create any more walks than necessary to represent semantic dependencies between nodes. Only when adding a new edge connecting nodes already present in the graph is not enough because we are indeed *changing* grammatical function and not *adding* relations to those already existing do we resort to introducing a new node in the graph, following the usual MGG ‘unambiguous path’ requirements (rather, *unambiguous trail*, under the present assumptions). But this kind of operation seems to be the exception rather than the rule, and a quite rare exception also.

This approach to the analysis of dependencies in natural language strings yields an unexpected result for the transformationalist: *most transformations actually do not change grammatical relations which were created by lexical insertion*. A caveat on the status of transformations in this framework is necessary: transformations are taken here simply as descriptive devices, without implying that syntactic objects *actually* move or anything remotely similar. This view on what transformations are and what they can do was actually rather widespread in the early days of Transformational Generative Grammar¹¹, but the line between description and purported explanation was blurred in more recent times, and in orthodox circles (but see, e.g., Postal, 1997: ix-x, 1 for a take on ‘extraction’ as a purely descriptive term).

Conceiving of phrase markers as connected *cycles*, where a *cycle* is an irreducible *single-rooted directed graph*, has the advantages already noted by Grindrod et al. (2014), Saddy & Grindrod (2016) in their formalization of neural networks. Connections between nodes can be expressed by means of a matrix (see also Van Steen, 2010: 60). Consider a graph G with n vertices: we need a square $n \times n$ adjacency matrix $A(G)$:

16)

$$G = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nn} \end{bmatrix}$$

In $A(G)$, $v_{ij} = 0$ if there is no edge between v_i and v_j . In other words:

17) $v_{ij} = 0$ iff $(i, j) \notin \rho$

The adjacency matrix corresponding to a phrase marker is *not symmetric* (see definition 4): for any two vertices v_i and v_j , $A[v_i, v_j] \neq A[v_j, v_i]$. The main diagonal of $A(G)$ is composed just of zeros, because there are no direct loops: no v directly dominates itself:

18) $(v_n, v_n) \notin \rho$

This is rather conventional, see, e.g., APG’s *No Loop Condition* (also Postal, 2010: 12); Dependency Grammar trees by definition also comply with this condition, because dependency is a ‘strict mother-daughter relation’ (Osborne, 2008: 1122; see also Osborne et al., 2011; Kahane & Mazzotta, 2015), with *mother of* being a dyadic relation between distinct elements. However, we introduce the following condition, pertaining to *transitive domination*:

¹¹ For instance, Rogers (1974: 556) lucidly says that:

Transderivational constraints, global and interpretive rules, and transformations, it seems to me, don’t explain anything: they describe (underline in the original).

19) $(v_n, v_n) \in \rho^*$

This condition allows for a trail to visit a single node more than once, provided that said trail visits other nodes in between. This is a crucial aspect of our proposal, and intimately related to *multidominance* (i.e., the rejection of the SMC): we will see that this condition allows for a simplification of the *chain* mechanism that is present in transformational generative grammar (see also Kracht, 2001).

The reader familiar with APG will note that the *Bicircuit* relation (decomposable into $\text{Branch}(A, B) \wedge \text{Branch}(B, A)$ for any A, B) does hold in the theory exposed here (or *Circuit* in Postal's more recent Metagraph grammar; Postal, 2010: 12-13), however, our theory does *not* allow for the *Parallel* relation (where two arcs share both 'heads' and 'tails') to be reflexive (*contra* Johnson & Postal, 1980: 41). This should help us give a proper definition of anti-locality conditions. We allow for less primitive relations than APG, but those we allow are n -ary and defined over a *single* level of 'representation'. Our modest graphs are syntactic connections defined over semantic vertices, without even a spare thought for linearity or the morpho-phonological exponents of vertices, at least in the initial version of the theory. In this respect, the reader might find it useful to check out work on linearization of Dependency trees, like Kahane & Lareau (2016); and the axiomatization of precedence relations in Postal (2010: 26).

The representations whose connectivity patterns we try to describe exhaustively are subjected to the usual constraints over variables (*a la* Ross, 1967), which also apply to TAGs (Kroch & Joshi, 1987) and other, non-transformational models; this is desirable, given the remarkable empirical robustness of some of Ross' constraints (particularly insofar as English is concerned), however, these constraints have to be understood strictly as constraints over *walks* and more specifically *trails* rather than as filter over *allowable trees*. Essentially, we have a set of nodes, and we want to know if a walk between any two is legitimate (that is, we want to answer the question 'can I go from A to B?', where A and B are vertices): walking that *trail* is interpreting the structure, and if a *trail* is illegitimate, so is the interpretation of the relevant sub-graph. Consider Uriagereka's (2012: 56) observation that any exhaustively binary branched tree is expressible in a finite state fashion (the so-called 'Finite-State limit on phrase structure'), a point previously noted by Greibach (1965):

a given finite-state language L can be generated either by a psg [Phrase Structure Grammar] containing only left-linear rules: $Z \rightarrow aY, Z \rightarrow a$, or by a psg containing only right-linear rules: $Z \rightarrow Ya, Z \rightarrow a$, and a psg containing either only left-linear rules or only right-linear rules will generate a finite state language (Greibach, 1965: 44)

Consider now that a Markov chain M is *irreducible* if it is possible to get from any state $i \in M$ to any other state $j \in M$ ($i \neq j$) without 'leaving the chain'. The mutual accessibility of states i and j is expressed in terms of there being a non-zero probability associated to a transition function between i and j . The bottom line of the proposal here is that 'form graph' applies so as to maximize irreducible Markov chains (IMC) in derivations. The connection between these IMCs *can* be of a higher order (e.g., a PSG).

In this conception, the grammar does *not* generate a set of surface structures (as in ST-EST-REST), a set of derivations (Generative Semantics; *vanilla* Minimalist Program), a set of form-meaning pairings (Categorial Grammar), a set of constituent structure-functional structure pairings (LFG), or a set of strings (*Syntactic Structures*-model and pre-ST). *The grammar as understood in the present work* (strongly) *generates a set of graphs*.

The following sections will be devoted to the analysis of linguistic phenomena using the notions introduced thus far. We will focus on data and paradigms that have proven difficult to analyze in Immediate Constituent-based, SMC-respecting frameworks. The first phenomenon that we will turn our attention to is a particularly challenging one for PSG: discontinuity.

3. Some notes on discontinuous dependencies

In general, ‘discontinuous constituency’ refers to a semantic or syntactic dependency holding between nodes that are not adjacent in linear structure. The notion has been part of IC analyses since pre-generative times; that can be seen clearly in Wells’ (1947) definition of *discontinuous sequence*:

A DISCONTINUOUS SEQUENCE IS A CONSTITUENT IF IN SOME ENVIRONMENT THE CORRESPONDING CONTINUOUS SEQUENCE OCCURS AS A CONSTITUENT IN A CONSTRUCTION SEMANTICALLY HARMONIOUS WITH THE CONSTRUCTIONS IN WHICH THE GIVEN DISCONTINUOUS SEQUENCE OCCURS (Wells, 1947: 104. Capitals in the original)

Wells’ proposal for the revision of the IC system includes multiple immediate constituents depending on a single node, which translates into n -ary branching for tree representations. Quite more recently, Ojeda (2005) presents the issue in the following terms:

Let A, B, C, D be four constituents of some word or phrase. A is said to be discontinuous if and only if (i) B is a constituent of A, (ii) C is a constituent of A, (iii) D is not a constituent of A, and (iv) D is linearly ordered between B and C.

Note that ‘discontinuity’ pertains to linear order primarily: syntax and semantics are affected only insofar as they determine linear order: this is a prominent property of Minimalist syntax (see Kayne, 1994; Moro, 2000), but it is not unheard of in other frameworks. For example, Kahane & Lareau (2016) propose a Dependency Grammar approach to the relation between linear order and syntactic structure, such that governor-dependant relations are subject to a linearization rule that converts linear precedence links in dependency trees. Linear dependency links determine the position of dependants with respect to their governors in terms of ‘same’ and ‘opposite’ directions in a left-to-right parsing model. We will see how our proposal differs from these, and how it can accommodate for different patterns of discontinuous dependencies.

In the framework we pursue here, Ojeda’s definition would be represented as follows:

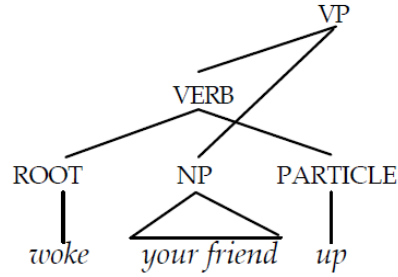
$$\rho^{(*)} = \{(A, B); (A, C)\}$$

$$\tau = \{(B, C); (B, D); (D, C)\}$$

$$e \langle A, B \rangle; e \langle B, C \rangle; e \langle B, D \rangle; e \langle D, C \rangle$$

Consider now the following phrase marker, which Ojeda offers as an example of discontinuity (see also Jacobson, 1987 and Bach, 1979 for an analysis of the operation Right Wrap which yields the V+NP+Prt pattern from V+Prt+NP):

20)



Here, we have the following relations:

$$\begin{aligned}
 21) \rho &= \{(VP, VERB); (VP, NP); (VERB, ROOT); (VERB, PARTICLE)\} \\
 \tau &= \{(VERB, NP); (ROOT, PARTICLE)\} \\
 e &= \langle VP, VERB \rangle; \langle VP, NP \rangle; \langle VERB, ROOT \rangle; \langle VERB, PARTICLE \rangle
 \end{aligned}$$

However, it is not clear that we need all that structure: we should be able to clean it up. The intermediate node [VERB], particularly, seems to be otiose. Moreover, Ojeda’s representation follows the Single Mother Condition, which we also challenge. An alternative structure would be:

$$\begin{aligned}
 20') \rho &= \{(VP, NP); (VP, ROOT); (VP, PARTICLE); (ROOT, NP)\} \\
 \tau &= \{(ROOT, NP); (ROOT, PARTICLE); (NP, PARTICLE)\} \\
 e &= \langle VP, ROOT \rangle; \langle VP, NP \rangle; \langle VERB, NP \rangle; \langle VERB, PARTICLE \rangle
 \end{aligned}$$

Note that we have eliminated VERB, and that the NP has two mothers: VP and ROOT. But, do we really need VP as a node in the graph? VP is the ‘name’ we give a subgraph (that is: the graph ‘VP’ is the set of nodes v_{ROOT} , v_{NP} , and $v_{PARTICLE}$), not a node in that subgraph. Thus, it seems we can dispense with VP as a node: VP is a set of connected nodes, *in and of itself a connected subgraph*; not a node within a subgraph. Thus, what we have is¹²:

$$\begin{aligned}
 20'') \rho &= \{(ROOT, NP); (ROOT, PARTICLE)\} \\
 \tau &= \{(NP, PARTICLE)\}
 \end{aligned}$$

¹² The rationale behind the ‘tree pruning’ undertaken in the paragraphs above, and which we will assume for the rest of the paper, is not too different from what underlies LFG’s Economy of Expression in *c*-structures:

All syntactic phrase structure nodes are optional and are not used unless required by independent principles (completeness, coherence, semantic expressivity). (Bresnan, 2001: 91)

As in LFG, the restriction assumed here applies to terminals as well as nonterminals: here, a terminal is a node whose ρ -domain is empty (note, incidentally, that this ‘theorematic’ conception of terminals differs from that of other graph-theory based frameworks like Arc Pair Grammar, in which terminals and nonterminals are primitives, see Johnson and Postal, 1980: 29, ff.). The whole point of this work is have as few terminals as possible, but also as richly connected as possible. It is important to note here that, strictly speaking, *economy of expression* is a condition over all expressions in the language (Pullum, 2007: 2), which makes it incompatible with model-theoretic theories. If *economy of expression* is understood in a meta-theoretical sense, however, as an admissibility condition over which structural descriptions can be proposed within the theory, it might survive (or, for instance, as a rule pertaining to *sponsor* and *erase* arcs in APG: something like ‘only S-graphs are well-formed graphs’; Postal, 2010: 27; Johnson & Postal, 1980: 78, ff.). We will not attempt to reformulate the principle, however, because for all descriptive intents and purposes it is clear as it is, and because it is not clear that there would be any significant difference beyond purely theoretical concerns.

$e = \langle \text{ROOT}, \text{NP} \rangle; \langle \text{NP}, \text{ROOT} \rangle; \langle \text{ROOT}, \text{PARTICLE} \rangle; \langle \text{PARTICLE}, \text{ROOT} \rangle$

The root selects the particle, and the particle delimits the event denoted by the root (wake up ~ heat up), thus the two edges from ROOT to PARTICLE and from PARTICLE to ROOT. The root subcategorizes for the NP, and the NP defines the Aktionsart of the root. In this example in particular that is not very clear because ‘wake’ is an achievement, just like ‘wake up’, so $e \langle \text{NP}, \text{ROOT} \rangle$ might not be needed, but consider ‘eat’, which is an activity, and ‘eat up’, which is an accomplishment. Similarly, it is the verb that agrees with the subject, not the other way around. An edge *from* the verb *to* the subject seems to be required to establish a *dependency* (in the technical sense, see Osborne, 2005) in which a predicate selects its arguments; an edge *from* the subject *to* the verb seems to be required to guarantee agreement (another way to do that –possibly preferable to ours- would be a rule of functional application, as in Schmerling, 1983b: 35, ff.). In what follows, we will only indicate the V-S dependency (more generally, *predicate-to-argument*).

Maximizing local connectivity and eliminating transformations has other advantages. Consider the following examples:

- 22) a. A man entered who was wearing a black suit (Relative Clause Extraposition, RCE)
 b. *What colour was a man entered who was wearing? (violation of the Complex NP Constraint, CNPC; Ross, 1967: 127)

Ojeda (2005) correctly identifies discontinuity in RCE cases, which ensues the CNPC violation that renders (22 b) ungrammatical (which in turn depends on the RC forming a constituent –the *complex NP*- with the N as observed by McCawley, 1998: 451; we will come back to this particular issue in **Section 8** below). However, it is also necessary to ban the RC from being an adjunct to a phonologically null NP or a copy of the subject before raising to Spec-TP to satisfy some EPP property or any other intra-theoretical fancy. That is: we want to eliminate possible structures like (21), which contains a deleted copy of [a man]:

- 23) A man entered ~~a man~~ who was wearing a black suit

Discontinuity alone cannot do that, unless a further ban on null copies is introduced. Or, seen from the opposite side, if *copies are never introduced to begin with*. We have also more than a single derivational *cycle*, each of which corresponds to a connected irreducible single-rooted sub-graph. Let us try and make things explicit:

- 24) Cycle 1: [a man₁ entered]
 Cycle 2: [a man₂ was wearing a black suit]
 $\rho_1 = \{(\text{entered}, \text{a man}_1)\}$
 $\rho_2 = \{(\text{wearing}, \text{a man}_2); (\text{was}, \text{wearing}); (\text{wearing}, \text{a black suit})\}$
 $e = \langle \text{enter}, \text{a man}_1 \rangle; \langle \text{a man}_2, \text{a man}_1 \rangle; \langle \text{wearing}, \text{a man}_2 \rangle; \langle \text{was}, \text{wearing} \rangle; \langle \text{wearing}, \text{a black suit} \rangle$

Let us make a preliminary analysis with what we have so far. Up to now, we have identified *cycles* based on purely configurational information, namely, (i) the presence of a single root node and (ii) irreducibility. But we have given no argument as to why these properties should matter at all in the generation of descriptively adequate structural descriptions for natural language strings: now we will give such arguments. Essentially, each sub-graph (each *cycle*) in a structural description contains the following elements (based on García Fernández & Krivochen, 2017):

- a. A predicative basic expression p
 b. Temporal and aspectual modifiers of p

c. Nominal arguments of p (subjects, objects, clitics)

In García Fernández & Krivochen (2017) we referred to each structural unit containing (a, b, c) as the *extended projection* of p (in the sense of Grimshaw, 2000; Abney, 1987: 57). This idea is essentially weakly equivalent to the so-called *Condition on Elementary Tree Minimality in Lexicalized Tree Adjoining Grammars* (LTAG):

Each elementary tree consists of the extended projection of a single lexical head (Frank, 1992: 53)

In (24), we have *two* cycles: Cycle 1 [a man entered] contains an unaccusative predicate *enter*, with aspect and tense marked synthetically and its nominal argument *a man*; Cycle 2 contains a transitive predicate *wear*, with progressive aspect realized by means of the auxiliary *be* and *-ing*, and both its subject and object *a man* and *a black suit* respectively. We see that each individual sub-graph satisfies the definition of *cycle* that we introduced above, and is also compatible with the CETM. Frank (1992) correctly observes that the CETM restricts the size of a single elementary tree, because of the constraints on which ‘extended projections’ can be built: the scare quotes are required here because extended projections were originally defined in terms of X-bar theory (Grimshaw, 2000: 116, ff.), with endocentricity as a fundamental property. However, if *endocentricity* is understood as *feature percolation* (that is: the features of a head percolate upwards the projection path, such that XP is a projection of X by virtue of X’s features –categorical, say- being percolated to XP; as in Grimshaw, 2000: def. (3)), this notion clearly *does not apply* to our graphs, seeing as (a) they are not labelled, and (b) there is really no notion of ‘projection’ in the present framework. We will thus not use the term ‘projection’ to avoid misunderstandings, and interpret the CETM in a *sui generis* way. We will return to TAGs below, since this formalism will allow us to test aspects of the strong generative capacity of our own proposal. But for the time being let us return to the issue of the identity of *cycles*: we can enrich the definition, such that a *cycle* is *an irreducible single-rooted sub-graph containing a single predicative basic expression, temporal and aspectual modifiers of that predicative basic expression* (which can be analytically or synthetically expressed) *and nominal arguments of that same predicate*. This is an improved definition, although we will see that there are further considerations to make, which can simplify the conceptual apparatus.

Clearly, these structures which here we refer to as *sub-graphs* are not disconnected units: just like temporal dependencies must respect *consecutio temporum* (see, e.g., Comrie, 1986; Carrasco & García Fernández, 1994), there are restrictions over the occurrences of nominal arguments and the dependencies that can be established between occurrences of an object across sub-graphs. For instance, let $p \in X$, $p' \in Y$, $p'' \in Z$ be predicative basic expressions in sub-graphs X , Y , Z . Furthermore, if α is a nominal dependant of p in X , α must be dominated by p in X : $\rho(p, \alpha)$ must hold in X . Now, if we have an argument β which may surface as a syntactic-semantic dependant of more than a single predicate p, p', p'' , then it must be possible for β to be dominated at X , Y , and Z by p, p' , and p'' respectively. This is a condition that requires an explicit rejection of the SMC, for β has –in this example- *three mothers*: p, p' , and p'' . Note that we do not multiply the entities (we have just one β) nor the relations (because under a theory that had as many instances of β as distinct subcategorizing predicates, there would still be *three* predicate-argument relations at least); the theoretical apparatus is however simplified by the elimination of an axiom (the SMC) which arose not out of empirical necessity, but out of intra-theoretical requirements.

The previous paragraph made an essential point in terms of how cross-cycle dependencies are established; that is: how distinct cycles are related, thus yielding global compositional interpretation. These considerations must now be applied to the analysis of (23). In the structural description (24),

the relation $\langle a \text{ man}_2, a \text{ man}_1 \rangle$ serves as the ‘pivot’ between both cycles [a man entered] and [a man was wearing a black suit], since [a man] actually stands for the translation of the NP into intensional logic (IL): $\lambda PP \{^a \text{ man}\}$, or *a man*¹³. In the terms we used in Krivochen (2015b), we are dealing with two *tokens* of the same *type*: [who] and [a man] both have the same IL translation.

This last point requires further argument, and it is mostly beyond the focus of the present work. However, we do need to point out that the idea that proper and common NPs have the same kind of translation is not present in Montague (1973); rather, it is to be found in Schmerling (1988), who bases part of her argument in Sloat (1969), among others. Configurationally, the fact that a node (in this case, an NP) can be a term of more than a single grammatical relation is captured via multidominance: making the relevant node a daughter of as many mothers as predicates take it as an argument (see also Postal, 2010: 17 for some discussion). We will see below that dominance is not always to be interpreted in predicate-argument terms: that depends on properties of the relevant nodes connected by a directed edge.

But, if this is the case, then we need to make some adjustments, because we are *visiting the same node twice*, once per cycle. This is a radical departure from a *chain-based* perspective built over rooted, labelled, oriented binary-branching trees in which *paths*, rather than *trails*, are relevant (cf. Kracht, 2001; Collins & Stabler, 2016): we have not ‘moved’ anything (remember, there are no derivations), nor do we require to link two occurrences of an object¹³. We would also want to eliminate the edge $e \langle a \text{ man}_2, a \text{ man}_1 \rangle$, because if we are dealing with the same node (each visit to which materializes as a *token*) we would be having $\rho(x, x)$, which we banned above. It is a good opportunity to point out that the framework presented here differs from Metagraph grammar in an important sense: because nodes stand for the translation of basic expressions into IL, there is no need to resort to *copy (pronominal) arcs* (Postal, 2010: 40-41), because we do not have multiple terminals for ‘a man’ and ‘who’ in (23); in our scenario, there is no need to *copy* or indicate co-reference between nodes in any specific way, because there is only *one node per entity* (where ‘entity’ is to be read in model-theoretic terms, and assuming that entities are associated to a unique address in IL). Thus, we can do better than the preliminary analysis above.

Distinct cycles are related by means of nodes which are the targets for *embedding transformations* (in the sense of Fillmore, 1963) or *substitution/adjunction* (in the sense of Joshi, 1985; Kroch & Joshi, 1985), both essentially *generalized* transformations (*adjunction*, if interpreted as a general case for *substitution*). We can provide the following semi-formal definition:

25) Linking (definition):

*If a sub-graph X contains v_1 and a sub-graph Y contains v_1 (that is, if $\rho(v_X, v_1) \ \& \ \rho(v_Y, v_1)$), X and Y are **linked** at v_1*

Transitively, anything that v dominates will also be contained in all sub-graphs in which v has a mother. Note also that the relative clause is embedded in the NP, not directly depending from the root S. The revised set of relations would then be:

- 24’) Cycle 1: [a man’ [Cycle 2] entered]
 Cycle 2: [a man’ was wearing a black suit]

¹³ In Krivochen (2015b) we have argued against equating ‘tokens’ and ‘occurrences’; the present proposal supersedes that one in not requiring structure mapping at all, and thus not needing to relate proof-theoretically related phrase markers –base and derived–.

$$\rho_1 = \{(\text{entered}, a \text{ man}')\}$$

$$\rho_2 = \{(\text{wearing}, a \text{ man}')\}; (\text{wearing}, a \text{ black suit}); (\text{was}, \text{wearing})\}$$

Both cycles are linked at the node *a man'*, which is visited *twice*: once in Cycle 1 and once in Cycle 2. Of course, it is possible to have both hypotactic and paratactic relations, depending on whether adjunction occurs at the root or not: if the linking node is immediately dominated by the root node in *both* sub-graphs, we are in the presence of a *conjoining transformation*. If linking targets an embedded vertex (and its ρ -domain), we have *embedding*, as in the RCE case above.

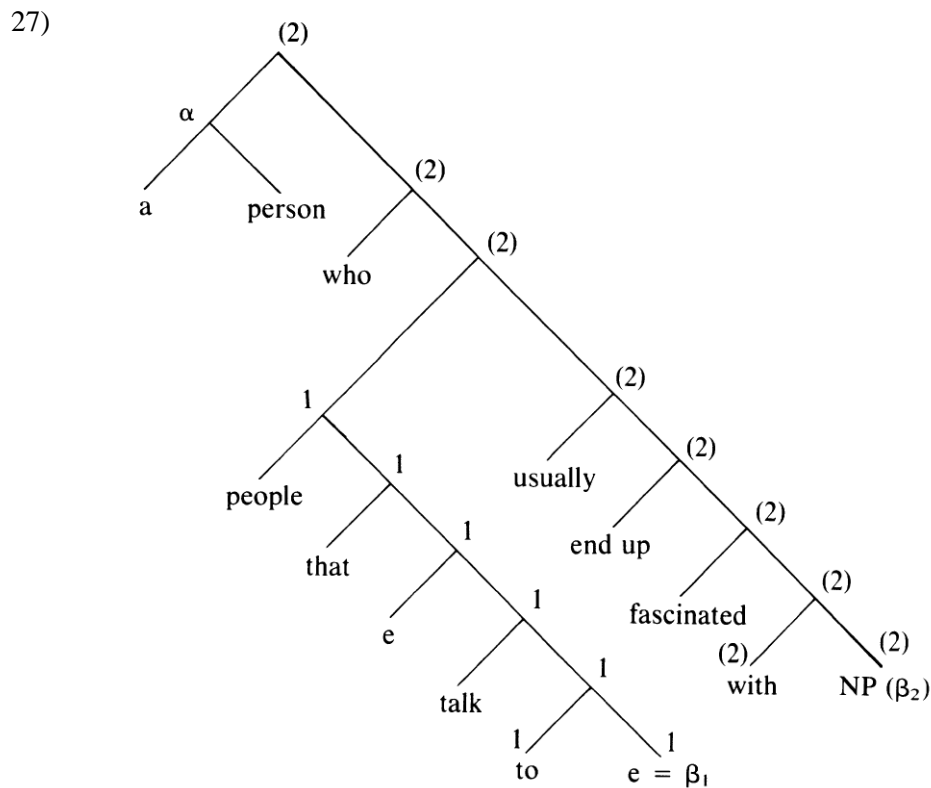
More explicit examples follow.

4. Comparing models

It would make little sense to have yet another theory for 'John loves Mary' and compare the present approach with currently available models of syntax based on these kinds of structures. If there is any merit in this work, it must be tested at the limits of empirical analysis. We have seen that we can accommodate discontinuity, which is problematic for Item-and-Arrangement-type Immediate Constituent analyses (see Schmerling, 1983a; Wells, 1947 for discussion), particularly MGG-inspired theories, in which only a single kind of local structural template $\{X, YP\}$ –where distinctness of X and Y is essential- is allowed (Chomsky, 2009, 2013: 43). Further comparison is due. To see how our model differs from the usual state of affairs in orthodox transformational generative grammar, consider the multiple-gap construction (26):

26) A person who people that talk to usually end up fascinated with (from Kayne, 1983: 230)

The GB-MP representation is (27):



Where

the NP object of *with* is (*e*), so that there are two empty categories β_1 and β_2 . The nodes that belong to G_{β_1} , and G_{β_2} [the government-projections of β_1 and β_2] have been endowed with '1' and '2', respectively, for the purposes of exposition (Kayne, 1983: 230)

This structure features multiple co-referent terminals:

- 28) a. People = *e*
 b. who = $\beta_1 = \beta_2$ = a person

However, despite sharing denotata (or, more specifically, despite having identical IL translations in the framework under discussion here), those are all *different nodes* in MGG (extra elements must be introduced in the representation to encode semantic-referential identity: indices). Note that the single mother condition is respected throughout, which forces this multiplication of terminals for what is ultimately the same entity. All differences pertain to its syntactic context (e.g., 'be the sister of P', 'bear Oblique case', etc.), that is, to the local relations it establishes with other nodes.

Moreover, cycles are stipulated in terms of 'government projections', and there is no natural way to capture predication domains in this approach (mainly, because the syntactic component is autonomous, and thus semantic relations cannot be defined over syntactic structure or vice-versa). The approach proposed here attempts to change both aspects of Kayne's approach (widely shared in MGG). We will focus on the connectivity of the relevant nodes, but the reader should bear in mind that here we operate over semantic substance: a structural description for a string is the set of all possible relations between nodes, and these nodes are proxies for their translation into IL.

In the approach presented here, we have the following *preliminary* sets of structural relations for a sentence like (26) above:

- 29) Cycle 1: [A person #]
 Cycle 2: [who [Cycle 3] usually end up fascinated with [Cycle 1]]
 Cycle 3: [people that talk to [Cycle 1]]
 $\rho_1 = \{(a, \text{person})\}$
 $\rho_2 = \{(\text{end up}, \text{who}); (\text{fascinated}, \text{who}); (\text{fascinated}, \text{with}); (\text{with}, \text{who}); (\text{fascinated}, \text{who})\}$
 $\rho_3 = \{(\text{people}, \text{that}); (\text{people}, \text{talk}); (\text{talk}, \text{to}); (\text{to}, \#)\}$

Where '#' is a placeholder to which cycles can be adjoined (*à la* Chomsky, 1955); basically, it works as a Joker to be targeted by *adjunction* or *substitution*, as appropriate (Joshi, 1985). But, what are we really dealing with here? Let us use *person'* to denote the translation of the NP [a person] into intensional logic, that is, $\lambda PP \{^a \text{person}\}$. Similarly, we will use *people'* as an abbreviation of $\lambda PP \{^a \text{people}\}$. We can thus get rid of some elements, because we have proceeded as if form was meaningful. [who] and [person] have the same translation into IL, thus, they are simply *tokens* of the same *type* (Krivochen, 2015b). There are *not* two distinct nodes, but *a single node* which can be visited as many times as necessary in a trail. Then, we can simplify (29) as (30):

- 30) Cycle 1: [*person'*]
 Cycle 2: [*people'* usually end up fascinated with *person'*]
 Cycle 3: [*people'* that talk to *person'*]
 $\rho_1 = \{\emptyset\}$
 $\rho_2 = \{(\text{end up}, \text{people}'); (\text{fascinated}, \text{people}'); (\text{fascinated}, \text{with}); (\text{with}, \text{person}'); (\text{fascinated}, \text{person}'); (\text{end up}, \text{fascinated})\}$
 $\rho_3 = \{(\text{people}', \text{that}); (\text{talk}, \text{people}'); (\text{talk}, \text{to}); (\text{to}, \text{person}')\}$

Different cycles (i.e., different subgraphs) intersect at all common nodes. Thus,

- 31) Cycle 1 \cap Cycle 2 = {P' }
 Cycle 1 \cap Cycle 3 = {P' }
 Cycle 2 \cap Cycle 3 = {Q', P' }

But {P' } is all there is to Cycle 1; thus, the intersection between Cycle 1 and Cycle 2 is trivial: Cycle 1 is *embedded* into Cycles 2 and 3, and acts as a ‘hedge’ (or ‘pivot’) between sub-graphs 2 and 3 (i.e., Cycles 2 and 3) that makes it possible to build a global compositional interpretation for (26).

The framework we present here maintains some descriptive aspects captured by the notion of ‘constituent’ from IC analyses, although with important changes: our blocks are sub-graphs, not labelled phrases. It may be observed that Categorical Grammars (which, as argued by Hockett, 1954 and Schmerling, 1983a, belong to a different class of grammars than IA models –which include IC grammars-) can isolate portions of structure by without making reference to constituency at all; rather, what matters is the set of indexed basic expressions and the operations that yield derived expressions where *indexing* is interpreted as ‘assignment to a category’.

It is important to stress now that, while the internal constitution of each sub-graph may resemble the structures one gets in a Dependency Grammar (Tesnière, 1959; Osborne et al., 2011; Kahane & Mazziotta, 2015, among many others) in rejecting binarity as a fundamental structure building principle (Krivochen, 2015a; 2017a) and giving importance to connections defined in terms of dominance (Osborne’s *catenae* and our *local trails*); there are several *crucial* differences, both in strong and weak generative power. We will now go into some of those differences, without being exhaustive.

A basic principle of DG is that ‘*dependents should be grouped together with their head*’ (Osborne & Niu, 2017). This is the principle that guarantees that a structural description for

- 32) Walk really fast

Looks like (33 a) and *not* like (33 b):

- 33) a. [walk [really fast]]
 b. [walk really [fast]]

Headedness is essential in a DG, even more so than in a PSG (which, as Hockett, 1958 and Lyons, 1968 argue, can –and should- incorporate exocentric structures). But it is not headedness we want to question now (despite the fact that there is no real sense in which sub-graphs are ‘headed’); rather, we question the segmentation that arises in a DG for a case like (32). We want to capture the fact that there is a connection between *walk* and *fast*, the presence of *really* notwithstanding. There is no need to invoke heads if all we are interested in is mapping the relations between nodes at a given point in time. We have, then, that:

- 34) $\rho = \{(fast, walk); (really, fast)\}$

Distributionally, paradigmatic choices affect either individual nodes or entire sub-graphs; in general we do not have any unit in between. It is still to be seen if the unit *catena* (a word or combination of words which is continuous with respect to dominance) as defined in Osborne et al. (2011) and Osborne & Niu (2017) serves any purpose in the present framework (if it does at all, we need to replace ‘word’ by ‘basic expression’, in any case; we will get to this shortly); it is certainly *not* a primitive in the present theory.

There are two further aspects with respect to which our theory differs from DG (and most forms of PSGs): first, as in certain forms of CG (e.g., Schmerling, 1983b), nodes need not correspond to

words. In CG terms, basic expressions need not be identified with lexical items (see also Jespersen, 1985 [1937]: 6 for a notational system that similarly assigns atomic symbols to multi-word expressions). Strictly speaking, in PSGs the correspondence is a consequence of conflating *terminals* with *lexical items*, and not a property of the formalism. An example of the kind of things we have in mind is the behaviour of *would rather* in the following examples:

- 35) a. *John wouldn't rather walk
 b. John would rather not walk
 c. There's no one here who wouldn't rather walk

Schmerling (1983b: 14) proposes that *would rather* is 'a two-word modal', assigned to the category (FC//IV)/(FC/IV), that is, a modified FC/IV which combines with an FC/IV to form an expression of category FC (i.e., Finite Clause). Note that negation cannot interrupt the unit unless under the scope of negation itself (as in (35 c), noted by Baker, 1970). A similar paradigm emerges with *would just as soon*, which Schmerling argues (p.c.) is also a multi-word basic expression. Jespersen (1985 [1937]) talks of 'composite verbal expressions' for things like *wait on* in *She waits on us* and assigns *wait on* a single primitive symbol W; similarly, he recognizes 'composite prepositions' like *on account of*, and similarly assigns them a single primitive symbol *pp*. We have discussed the problems that things like *would rather* pose for monotonic PSGs (including GB/MP grammars) in Krivochen (2016), here we just want to stress that the theory explored here does not observe either of the following requirements (from Osborne, 2005: 253):

- I. a. One wordform per node, and
 b. One node per wordform.
 II. One head per node

Requirements (I.a.) and (I.b.) are violated in the case of multi-word basic expressions: if there is a single node for *would rather* or *would just as soon* (that is, if nodes correspond to basic expressions rather than to words), there is no bijective relation between words and nodes. Requirement (II) is trivially violated if there is no meaningful notion of 'head'.

There is a third requirement formulated by Osborne, which is a more general constraint on (rooted) tree construction (see, e.g., van Steen, 2010: 109):

- III. One root node per structure

The framework exposed here violates (III) because our grammar can describe multi-rooted structures, which is inadmissible in DGs and PSGs. In PSGs, this is given by definition: the base component of a transformational grammar, as well as the level of *c-structure* in a Lexical Functional Grammar, is a Context Free Grammar. A CFG is a set $G = (V_N, V_T, P, S)$, where V_N is a set of nonterminal symbols, V_T is a set of terminal symbols, P is a production rule (also called 'transition rule') of the form $\Sigma \rightarrow F$ (where Σ is a possibly unary string in V^+ and F is a possibly unary string in V^{*14}), and S is a starting symbol (also called 'axiom'), such that $S \in V_N$. The existence of a starting symbol is a necessity since otherwise derivations would not start (see Binder, 2008 for some discussion about this scenario in

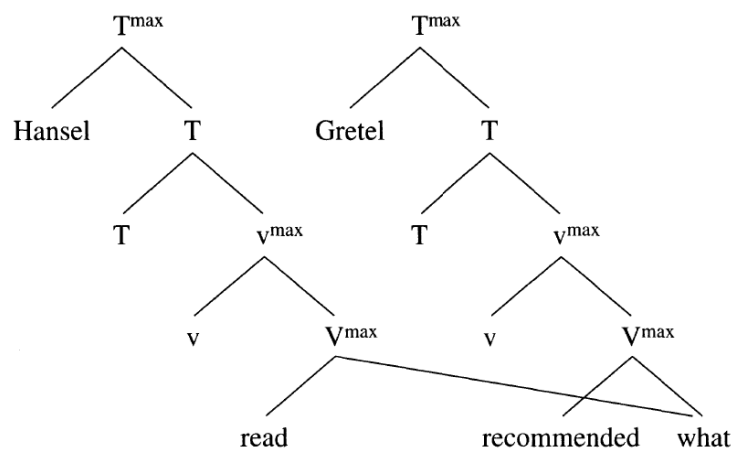
¹⁴ Hopcroft and Ullman (1969: 1) define that:

If V is an alphabet, then V denotes the set of all sentences composed of symbols of V, including the empty sentence [notated ε]. We use V+ to denote the set V* - {ε}. Thus, if V = {0, 1}, then V* = {ε, 0, 1, 00, 01, 10, 11, 000, ...} and V+ = {0, 1, 00, ...}.*

terms of a *computational frustration*): the starting symbol is a designated nonterminal which dominates but is not dominated; in other words, it is the only nonterminal that can only appear at the left-hand side of production rules. Both DG and our graph theory have the relation *dominate*, but it must be noted that if the *root* of a graph is defined *only* as a node that is not dominated by any other node in a sub-graph, then it is perfectly possible to have multi-rooted structures, the only requirement for v_x to be a root being that there is no node within the cycle which dominates v_x . As we integrate cycles with roots v_x and v_y , connections between sub-graphs can be established at nodes that v_x and v_y dominate, but keeping v_x and v_y undominated. Summarizing, there are local compatibilities between our graphs and Dependency graphs, and between our graphs and PSGs, but the systems are far from equivalent.

Outrageous though a locally multi-rooted representation might seem to some readers, it is not unheard of. Consider, for instance, the intermediate representation proposed in Citko (2005: 480) for ATB:

36) I wonder what Hansel read and Gretel recommended



T^{\max} can be safely characterized as a *root* in each sub-tree corresponding to the structural descriptions of *Hansel read what* and *Gretel recommended what*. Minus non-audible structure (T, v), our *local* description is not too far from that:

- 37) Cycle 1: [I wonder [Cycle 4]]
- Cycle 2: [Hansel read *what*]
- Cycle 3: [Gretel recommended *what*]
- Cycle 4: [[Cycle 2] and [Cycle 3]]
- $\rho_1 = \{(\text{wonder}, I); (\text{wonder}, \textit{what}); (\text{wonder}, \textit{and})\}$
- $\rho_2 = \{(\text{read}, \text{Hansel}); (\text{read}, \textit{what})\}$
- $\rho_3 = \{(\text{recommend}, \text{Gretel}); (\text{recommend}, \textit{what})\}$

A fundamental difference between our representation and Citko's is that in our representation the coordination in (36) is symmetric –i.e., *paratactic*– rather than asymmetric –i.e., *hypotactic*– (unless Gretel recommended something *after*, or *because* Hansel read it; in which case the asymmetry in semantics would betray an asymmetric phrase marker; see Schmerling, 1975; Krivochen & Schmerling, 2016 for detailed discussion). In any case, a declarative RNR sentence, like *Hansel read and Gretel recommended that book that's so popular now* would be indeed a multi-rooted structure, linked by the NP *that book that's so popular now* which is a common vertex between the two sub-graphs (as can be seen in (36) above). That node is obviously not a root because it is dominated by the relevant verbs in each conjunct. Note that in a RNR structure, single-rootedness can only be achieved

by including a designated axiom in the phrase structure rules in a top-down model: a recursive rule like (38 a) or an endocentric coordination phrase as in (38 b) (see, e.g., Progovac, 1998):

- 38) a. $S^* \rightarrow S \text{ and } S$ (and S)
 b. $\&P \rightarrow S \ \&'$
 $\ \&' \rightarrow \text{and } S$

Which generate the following structural descriptions respectively:



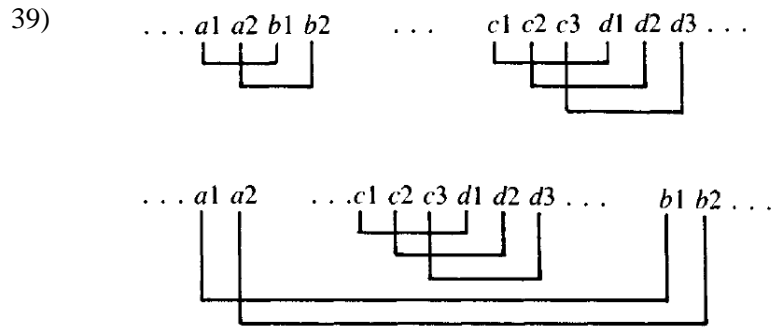
In bottom-up models (mainstream MP), the TPs corresponding to each clause are dominated by a Complementizer layer CP, and thus extra structure in the form of silent heads must be introduced to comply with *a priori* formal requirements (projection/endocentric labelling, the SMC, binarity...). Our proposal is very different from these. By ‘flattening’ the structures keeping only what is overtly present in the string, intermediate nodes (in the sense of Greibach, 1965) are eliminated: this can be seen as a radicalization of a rather basic desideratum of *economy of expression* (Bresnan, 2001: 90, ff.; see fn. 12 above), in that all ‘pre-terminal’ nodes are eliminated (recall also that there is no one-to-one correspondence between nodes and lexical items, such that a node can correspond to a multi-word basic expression). The same requirement (in a slightly stronger form) is expressed by Postal (1974: xiv):

one should assume the existence of no elements of structure, no levels of structure, and no kinds of representations whose existence is not absolutely necessary [...] I reject all so-called syntactic features, doom markers, other abstract syntactic markers, coding devices, “empty nodes”, “doubly filled nodes,” and, in short, the entire a priori unlimited set of symbolic elements available in an unconstrained system

We wholeheartedly agree with Postal’s meta-constraints on the theory of grammar, and will continue to assume that a descriptively adequate theory must meet this stronger version of FLG’s *Economy of Expression* (or any such constraint).

It seems clear that we can generate (recursively enumerate) graphs for all CF languages (as can the base component of a transformational grammar, the *c*-structure component of LFG, and GPSG). But, what is the true generative power of our theory? What kinds of structural descriptions for natural language strings can we give and what kinds are not allowed? To approximate an answer to this question, we will use Joshi’s (1985) discussion of the generative power of TAGs with links as a reference, and compare our system to his. Because the strong generative power of TAGs has been carefully studied, and because Joshi himself compares it with that of alternative frameworks, we consider that it is an appropriate measure.

TAGs with links can generate *limited* crossing dependencies, by virtue of having elementary trees annotated with links which are preserved under *adjunction*. A TAG can generate the CS language $L = \{a^n, b^n, e, c^n, d^n \mid n \geq 0\}$ iff *a*’s and *b*’s are nested, *c*’s and *d*’s are nested, and *a*’s and *c*’s and *b*’s and *d*’s are cross-serially related (or *vice versa*). That is (from Joshi, 1985: 223):



However, there are CS languages which *cannot* be generated by a TAG with links. These include languages in which *only* cross-serial dependencies are established, like $L = \{a^n, b^n, c^n, e, d^n, f^n \mid n \geq 0\}$, and double-copy languages, e.g., $L = \{w e w e w \mid w \in \{a, b\}^*\}$.

Now, because our nodes do not correspond to words or phrases or denotative expressions, but to the translation of those expressions into IL, the applicability of ‘links’ needs to be looked at very carefully. Essentially, there are two main situations in which a TAG needs links:

- I. To indicate filler-gap dependencies (in the sense of Postal, 1997; Sag, 2010)
- II. To indicate dependencies between overt elements (agreement, coreference)

Because there are no *gaps* in the present model, there will be no need to resort to links for (I). How about (II)? It seems to us that a further distinction needs to be made:

- II. a. N-V links
- b. N-N links

Links of the kind (II. b.) –the kind that is relevant for Pronominalization in a Ross-Langacker view, for instance- do not arise in our theory because those relations are formulated in terms of trails that visit a *single node* more than once, *not* as a relation between two distinct nodes. But (II. a.) is worth looking at. This is the kind of dependencies that arise in the now famous Dutch examples considered by Joshi:

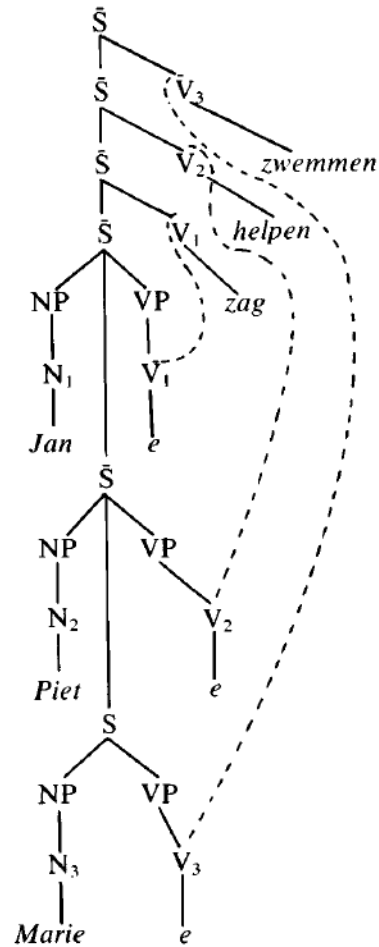
40) Jan Piet Marie zag helpen zwemmen
Jan Piet Marie saw help swim
 ‘Jan saw Piet help Mary swim’

The dependencies in (40) are as in (41):



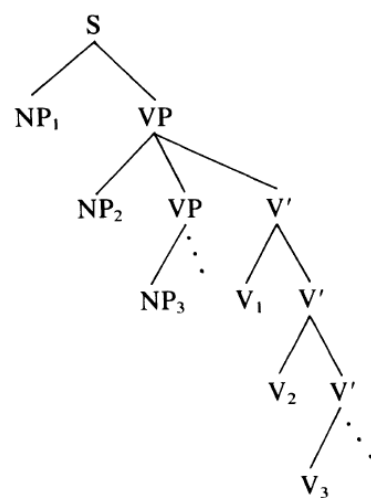
The abstract form of (40) is $a^1 a^2 a^3 b^1 b^2 b^3$, which displays crossing dependencies (indicated with superscript integers). The structure is more complex than it would seem:

42)



The structure proposed in Joshi (1985) is partly based on the discussion of cross-serial dependencies in Dutch in Bresnan et al. (1982). The latter propose the following structure (Bresnan et al., 1982: 619):

41')

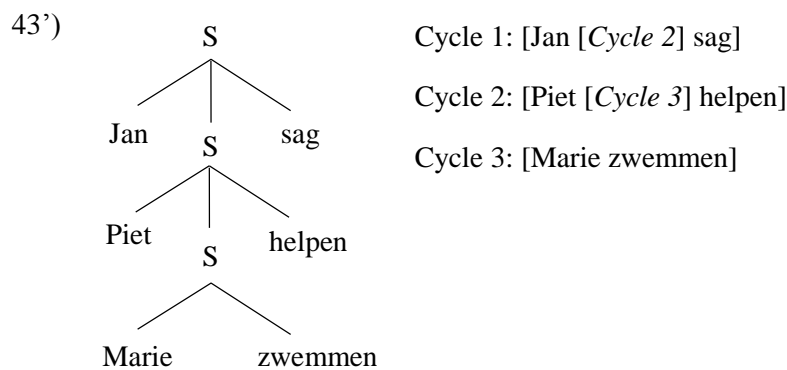


Despite some differences, both works conclude that there is no need to push the grammar to *strong* context sensitivity (i.e., to make the grammar powerful enough that it can generate *all* context-

sensitive languages); *weak* context-sensitivity, a.k.a. *mild* context-sensitivity¹⁵, seems to do the trick. Joshi’s structural description is based on the linear position of elements, *as well as* their referential dependencies. Let’s see if we can assign an adequate structural description to (40). To begin with, let us consider an incorrect, but intuitive, approach. If we grouped in a single cycle everything that a TAG joins by links (doing away with empty terminals, marked *e* in Joshi’s diagram), we would get:

- 43) Cycle 1: [Jan sag [Cycle 2]]
 Cycle 2: [Piet helpen [Cycle 3]]
 Cycle 3: [Marie zwemmen]

But things are more complex (the inadequacy of (43) is prefigured by Joshi). As it is, (43) describes the string *Jan sag Piet helpen Marie zwemmen*, which is certainly not what we have in (40). Equally, a *reduced phrase marker*-like approach based on strict center-embedding (Lasnik & Kuppin, 1977) is also inadequate (Bresnan et al., 1982: 616 argue that such structural descriptions generate a *proper subset* of the relevant structural descriptions –despite the fact that, in terms of weak generative capacity, that proper subset is ‘infinite’ [sic]-):



Both verbs *sag* and *helpen* take events (predicate-argument relations; ‘satisfied functions’) as complements; their internal argument is a complete event with external and internal argument, both of which are dominated by the relevant verb: this will become obvious in our dominance relations. Thus, we should have something more along the lines of (44), with the full set of dominance relations made explicit:

- 44) $\rho = \{(sag, Jan^?); (sag, Piet^?); (sag, helpen); (helpen, Piet^?); (helpen, Marie^?); (helpen, zwemmen); (zwemmen, Marie^?)\}$

Note that we have made no use of empty nodes, intermediate nonterminals (i.e., ‘bar levels’), or headedness. If there is something like ‘heads’, they arise from the consideration of dominance

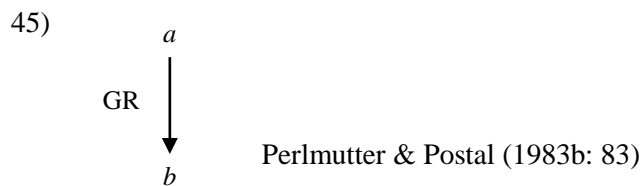
¹⁵ Bresnan et al. (1982) do not define *weak context sensitivity* formally. In Joshi’s work, ‘mild context sensitivity’ includes the following properties (as summarized in Krivochen, 2017: 65; from Joshi, 1985: 221, ff.):

- a. *Polynomial parsing complexity (for TAGs, it amounts to $O(n^6)$ in the worst case).*
- b. *Constant growth (for any string w , $|w|$ grows by a constant factor, which is a linear combination of the length of the terminal string of some initial tree and the lengths of the terminal strings of the auxiliary trees)*
- c. *Limited crossing dependencies (by virtue of a limited active search space)*

relations, but they are not primitives of the theory. A question that arises at this point is: if there is an edge from –say- *sag* to both *Jan* and *Piet*, how do we know which one is the subject and which the object? If verb-subject agreement is encoded in an edge $e \langle \text{Subj}, V \rangle$, there is no problem: there would be a cycle between every verb and its subject, but not between every verb and its object in (40). That is: there is a trail Subj-V-Subj but not Obj-V-Obj. This revised picture would yield the relations in (44’):

44’) $\rho = \{(sag, Jan’); (Jan’, sag); (sag, Piet’); (sag, helpen); (helpen, Piet’); (Piet’, helpen); (helpen, Marie’); (helpen, zwemmen); (zwemmen, Marie’); (Marie’, zwemmen)\}$

We will come back to the S-V-O relations shortly, because things are not always so clear when we have participial constructions with object agreement. Coming back to our original problem, however, what if agreement is not part of the picture? Are there other possibilities? Yes, there are. Even without adding an *a-structure* for roles of participants or *f-structure* level for grammatical functions (as in LFG; Bresnan, 2001: 19, ff.; Dalrymple, 2001: 8, ff.) or appealing to intermediate nodes to create specific configurations with respect to which functions are defined (Chomsky, 1965: 71), grammatical functions can be encoded in graphs by adding *annotations*. RG, APG, and Metagraph grammar (Perlmutter & Postal, 1983a, b; Johnson & Postal, 1980; Postal 2010 respectively) prefer to annotate *edges* rather than nodes, a perfectly viable strategy as well. For instance, a basic (and admittedly incomplete, although in ways that are at present irrelevant) general format for L-graphs could be (45):



Where (45) is to be read as ‘the linguistic element *b* bears the grammatical relation GR with the linguistic element *a*’, making no reference to linguistic levels (which would only come to play with ‘transformations’). Equivalent proposals (in the mathematical sense of ‘equivalent’) for the definition of arcs (the local vertex-edge relations in L-graphs) are made in Johnson & Postal (1980: 10, 37) and Postal (2010: 25-26). Also, Mel’čuk (2003), within a multi-layered model of Dependency Grammar, annotates arcs with integers, such that ‘*The arc between the predicate and its argument carries the number of the argument*’ (2003: 11).

Along these lines, we could add a specification to each edge, such that the grammatical function is encoded in an index (say, *s* for Subject, *o* for Object; we could also have used 1 and 2, in consonance with Relational Grammar practice) and obtain the following representation:

44’’) $e_s \langle \text{sag}, Jan \rangle, e_o \langle \text{sag}, Piet \rangle, e_o \langle \text{sag}, helpen \rangle \dots$

However, (44’’) is not quite it either. We find this notation cumbersome, and not really useful: adding diacritics is only justified if there is no other way to encode grammatical distinctions. If said distinctions have any real descriptive value, however, there should be a way to encode them in the grammar *qua* system, rather than as elements of the alphabet that the rules of the grammar manipulate.

Above, in **Definition 4**, we defined the ρ -set of a graph *G* as the full set of dominance relations between vertices in that graph. Then, we remained purposefully agnostic with respect to whether a ρ -set was an ordered or an unordered set: now, we will make use of that initial ambiguity to model grammatical relations in terms of our graph-theoretic approach. We can capture the relevant

functional asymmetries (Subject vs. Objects, and Direct Object vs. Indirect Object) by introducing an *order* into the set of dominance relations following an independently motivated functional hierarchy (see Dalrymple, 2001: 9 for a summary of different approaches to grammatical function hierarchies). Recall that a vertex v_1 dominates v_2 if and only if there is an edge $e \langle v_1, v_2 \rangle$ in a directed graph. Then, we can make the inherent asymmetry of a directed graph an asset: make that asymmetry correspond to an external hierarchy that can help us in the description of relations without the need to appeal to annotations (to either edges or vertices). We will take Relational Grammar's (RG) hierarchy of grammatical functions (e.g., Perlmutter & Postal, 1983b: 86 in terms of 'nuclear' and 'object' set-defining features; Bell, 1983: 148) as a reference. The motivation for this choice is that in RG grammatical functions are primitives (as opposed to the *Aspects* model, in which they are read off configurations, with labels for non-terminals and relations like *sisterhood* or *motherhood* being primitives instead (see e.g. Chomsky, 1965: 71). RG's definition of *clause* makes this clear:

A clause consists of a network of grammatical relations (Perlmutter & Postal, 1983a: 9)

In this context, grammatical functions configure a structured system: they are organized in a *hierarchy*. The RG hierarchy, as presented in Bell (1983: 148) is as follows:

46) 1 (Subject) >> 2 (Direct Object) >> 3 (Indirect Object)

Where terms outrank non-terms, and where *nuclear* functions outrank *object* functions: 1 is a *nuclear, non-object* function; 2 is a *nuclear, object* function, and 3 is a *non-nuclear, object* function (see also Johnson & Postal, 1980: 250 for an APG translation of the RG hierarchy). Now, how can we capture that hierarchy in the ρ -set of a graph? Suppose that we make the ρ -set of a sub-graph an *ordered* set, thus getting the following inequality to hold (square brackets are used to disambiguate the formula, and bear no particular theoretical significance):

47) $[\rho = \{(a, b), (a, c)\}] \neq [\rho = \{(a, c), (a, b)\}]$

We can notate the *ordered* ρ -set of a graph using angular brackets, such that the ordered set of dominance relations between vertices a, b, c is as in (48):

48) $\rho = \langle (a, b), (a, c) \rangle$

This means that the relation (a, b) is ordered before the relation (a, c) . What have we gained? If the order between dominance relations is made to correspond with the order between grammatical relations, we have the following relations, more or less formally¹⁶:

i) 2 (Direct Object) >> 3 (Indirect Object) >> 4 (Subobject) >> 5 (Semiobject)

However, things are rather complicated. The interpretative rule in (49) assumes that there is no Indirect Object unless there is a Direct Object, in line with traditional accounts; that assumption links the hierarchy of grammatical functions with the Case hierarchy, in which ACC >> DAT. Moreover, it is also not clear whether the relations 4 and 5 are 'as primitive' as the others, or depend on more fine-grained aspects of lexical semantics. For example, consider the following examples that Postal (2010: 73) uses to illustrate the relation 4 (Subobject), with the relevant NP(4) italicized:

$$49) \rho = \langle (a, b), (a, c), (a, d) \rangle$$

Let $a = V, A$, and $b, c, d = N$ (and $b \neq c \neq d$), then b is the Subject of a , c is the Direct Object of a , and d is the Indirect Object of a .

That is: the order of dominance relations corresponds to the order of grammatical relations, such that if a relation (x, y) is ordered before a relation (x, z) in terms of dominance, the grammatical relation (x, y) is also ordered before (x, z) in the hierarchy of grammatical functions (the two orders are isomorphic, which means that there is a function f with an inverse that relates both sets of orders). If there is a single NP in a structure, the hierarchy determines that it will be interpreted as the highest possible grammatical relation licensed by V , *modulo* lexical semantics. Note that b, c , and d are sister nodes by virtue of sharing a mother a ; the diagram of this graph depicts a ‘flat’ structure that is common in frameworks that aim at eliminating empty nodes, silent structure, and bar-levels (e.g., Culicover & Jackendoff, 2005: Chapter 4).

In this light, let us revise the set of dominance relations in (44), as (50):

$$50) \rho = \langle (\text{sag}, \text{Jan}'); (\text{sag}, \text{Piet}'); (\text{sag}, \text{helpen}); (\text{helpen}, \text{Piet}'); (\text{helpen}, \text{Marie}'); (\text{helpen}, \text{zwemmen}); (\text{zwemmen}, \text{Marie}') \rangle$$

The structure of (50) has *Jan* as the subject of *sag*, and *Piet* as its object. The relation *sag-helpen* is not contemplated in the functional hierarchy in (49) because we are not dealing with a V-N relation: the functional hierarchy in (46) and the interpretation of (46) from ordered relations in (49) pertain to functions that NPs or constructions with their functional potential (say, S') play in the A-structure of Vs. We can use a DG notion to make this somewhat clearer: *valency* (or *valence*). Tesnière (1959: Chapter 97, §3) presents the notion with a chemistry analogy (already used by Jespersen, 1937) as follows:

The verb may therefore be compared to a sort of atom, susceptible to attracting a greater or lesser number of actants, according to the number of bonds the verb has available to keep them as dependents. The number of bonds a verb has constitutes what we call the verb's valency

Vs do not count for satisfying the valence of a V, NPs do: this is easily illustrated in a contrast like **John saw run_V*, but *John saw Peter_{NP}*. In ECM alternations of monotransitive Vs, it is the whole event (say, *John saw [Peter run]*) that satisfies the valence of the V: we have indicated that by establishing a dominance relation between the matrix V and the embedded V (e.g., the edge $e \langle \text{sag}, \text{helpen} \rangle$). But it is important to note that *helpen* does *not* receive a grammatical function in the minimal sub-graph that contains *sag*. We will come back to these issues in the next section, which will deal with the grammar of complement constructions.

-
- ii) Herb neared *the vampire* / Herb wanted *pizza*

Even if the italicized NPs could be otherwise grouped, it seems strange to us that the same relation can be established with two verbs as different as *near* and *want*. As an example, *near* does not require an agentive subject, which suggests that it is not a transitive V at all (but rather, an unaccusative V):

- iii) The train neared the station / *The train wanted *pizza*

Considerations of this kind fuel our skepticism about the putative advantages of including Postal's extended typology of objects in (49) as is.

In any case, and coming back to the problem of *strong generative capacity*, the fact that we can provide a structural description for cross-serial dependencies pushes our system to *mild* context-sensitivity. Does it go all the way up to context-sensitivity? If our system did indeed go up to CS, it would mean that we could provide structural descriptions for strings that have been generated by CSGs. But our system is *not* unrestricted in the kinds of crossing dependencies it allows: in particular, as we will see in the next section, there are constraints with respect to when cross-cycle dependencies can be established. These constraints over the dependencies that can hold between elements in different cycles after *substitution/adjunction* make our system weakly equivalent to a TAG with links, in generative power, with the only difference that under specific conditions we do allow for *new* links to be created, joining nodes that belong to different cycles (roughly, different *elementary trees* using TAG terms). Sections 5 and 6 will come back to this constraint and the kind of structures it allows for.

5. On English Complement Constructions

In this section we will deal with the structural descriptions that our graph-theoretical approach assigns to different kinds of transitive constructions in English. In this section we are concerned with those cases in which a verb takes a non-finite complement: strings of the type $V_{\text{fin}} + \text{NP} + V_{\text{inf}}$. The classification of so-called ‘predicate complement constructions’ has been a concern since the early days of Generative Grammar: Rosenbaum (1965) distinguishes two kinds of such constructions, classified essentially in terms of VPs taking either NPs dominating Ss or VPs as complements; and Perlmutter (1968) and Postal (1974) expand and modify the proposal. These foundational works differentiate between the following kinds of non-finite complements:

- 51) a. Subject-to-Subject raising
- b. Subject-to-Object raising (a.k.a. Exceptional Case Marking ECM)
- c. Subject-controlled Equi
- d. Object-controlled Equi

These terms are used descriptively, without presupposing any theoretical analysis (despite the fact that the nomenclatures that we have used include the names of two transformations, *raising* and *Equi NP deletion*). In any case, what we require of a grammatical theory is that it adequately distinguishes between these structures and assign distinct structural descriptions where applicable. We will proceed in the order indicated in (51), starting from Subject-to-Subject raising.

Consider the following example:

- 52) John seems to have read the book

One of the first questions we can ask ourselves is ‘*how many cycles are we dealing with here?*’ The traditional view in transformational generative grammar is that in (52) the NP *John* receives thematic role from *read* and then moves (*raises*) to the matrix clause to receive Case from *seem*. The analysis of these structures is essentially *biclausal*: each verbal predicate *seem* and *read* is contained in separate clauses (dominated by distinct S nodes); the NP *John* originates as the subject of the embedded one, and moves to the position of subject of the matrix one. But transformational analyses for Raising to Subject have not been uncontested, and the same is the case with bi-clausal analyses. Jacobson (1992) and Bresnan (2001) present non-transformational bi-clausal analyses; Torrego (2002) and Krivochen (2013) present a transformational, monoclausal analysis. Here, we try to advance a non-transformational, monoclausal analysis (see also Gazdar, 1982: 152-155; Pollard & Sag, 1984: 133, ff. APG does not have ‘transformations’ as such, with raising constructions involving *immigrant* arcs which presupposes more than one predicate –but not more than one ‘clause’ in any meaningful

sense, see Johnson & Postal, 1980: 134). The reasons for assuming a monoclausal analysis (which translate here into having a single sub-graph, thus a single cycle) are essentially *semantic*: raising verbs are taken here to be assimilable to the class of auxiliary verbs, specifically, *modals* (see Wurmbrand, 1999 for the exact opposite conclusion). The map of connectivity relations in (51) is as follows:

- 53) Cycle 1: [John seems to have read the book]
 $\rho = \langle \langle \text{seems, } John' \rangle; \langle \text{to have, read} \rangle; \langle \text{read, } John' \rangle; \langle \text{read, } book' \rangle \rangle$

The relations between *read*, *John'*, and *book'* illustrate the interpretation rule in (48). *John* is the subject of *seem* and of *read*, but, crucially, there is *no relation* between *John* and the perfective auxiliary *have*: this is consistent with the analysis of the perfective perfect auxiliary as a *functional auxiliary* in Bravo et al. (2015) and subsequent work (see also **Section 6** below). *Have to* does not take *John* as an argument, nor does it impact on its semantic interpretation or assign it Case. It just makes sense not to establish a relation between the auxiliary and the NP. We can capture the fact that there is a relation between *seem* and *John* as well as *read* and *John* (the former, via agreement and Case, the latter, via theta-marking, however these are conceived of in the reader's framework of choice –but favouring lexical specification theories like LFG-) in a monoclausal structure (having a single sub-graph as the structural description of the whole sentence), and without multiplying the nodes (thus, no copies or traces).

Given this analysis, and anticipating part of the discussion in **Section 7** the reader may wonder what happens if there is a *lexical auxiliary* in the chain. That is, would we have an edge between, say, the modal *need* and *John* in *John seems to need to work extra time to pay the bills*, given that modal and phasal (e.g., *start*, *continue*, *finish*) auxiliaries are identified as *lexical auxiliaries* in Bravo et al. (2015)? Things are not clear. In principle, if lexical auxiliaries define local domains as claimed in Krivochen & García Fernández (2017) we would be dealing with *three* cycles: one for *seem*, one for *need*, and one for *work*. But we have been careful in the past when extending our analysis of the Spanish auxiliary system to English: the English auxiliary system *does not work like the Spanish one* (a point that is anticipated in the categorial system, in its Montagovian version as well as the Schmerlingian reworking). It is essential to note that Spanish auxiliaries can appear in different orders (given some constraints based on the type of auxiliary and the information it conveys), yielding different meanings, whereas the ordering of English auxiliaries is rigid. The internal segmentation of auxiliary chains is thus different in English and Spanish, as well as the relations that auxiliaries establish with other elements in the sentence, like the subject and the lexical verb. This is one of the factors that yield inter-linguistic variation, and it can be readily captured in a Montague-Schmerling approach. While there are arguments to have modals and aspectual auxiliaries form a constituent with nominative subjects in Finite Clauses FC and Inverted Finite Clauses IFC in English (Schmerling, 1983a), those *do not* apply to Spanish:

- Spanish auxiliaries inflect like finite verbs for TAM and agreement (whereas only two inflect for TAM in English: 'have-has-had' and 'be-is-was'). For example, we see in (53) below that the modal *poder* has an infinitive form and an imperfective past form, just like the lexical verb *trabajar*

- 54) a. Juan trabajaba_{IMPF/PAST} los fines de semana
J. worked_{ImpfPast} the weekends
 b. Juan podía_{IMPF/PAST} trabajar_{INF} los fines de semana
J. was able_{ImpfPast} to work the weekends

c. Juan tenía_{IMPF/PAST} que poder_{INF} trabajar_{INF} los fines de semana

J. had to_{ImpfPast} be able to_{Inf} work_{Inf} the weekends

- Spanish *lexical* verbs do invert in interrogative contexts, as well as *auxiliary chains* (including the lexical verb); this contrasts with English having a single ‘operator’ (the auxiliary that inverts with the subject) *per* clause (Quirk et al., 1985: 121)¹⁷

55) a. Juan *dijo* que... → ¿Qué *dijo* Juan?

J. said that... *What said J.?*

b. Juan *había* tenido que decir que... → ¿Qué *había* tenido que decir Juan?

J. had had to say that... *What had had to say J.?*

- Spanish auxiliaries are not restricted to specific clause types (i.e., they have full paradigms)
- Spanish does not have English-like VP ellipsis in its treatment of auxiliaries, including *stripping* (Schmerling, 1983a; 2016: Appendix B. We will come back to this property below)

56) a. -Who wants to come along?

-I do [~~want to come along~~]! (Culicover & Jackendoff, 2005: 233)

b. Robin ate a bagel for breakfast, and Leslie *will* [~~eat a bagel~~] too (adapted from C&J, 2005: 283)

c. *Juan comió un bagel de desayuno, y María también va a [~~comer un bagel~~]

A different categorial segmentation for the Spanish auxiliary system with respect to the English one seems to be thus not without strong empirical motivation. In this scenario, a deeper analysis of the English auxiliary system is required before jumping to conclusions. We will come back to aspects of the Spanish auxiliary system in **Section 7**.

We have seen some basic aspects of Raising-to-Subject constructions. Next up in (51) are Raising-to-Object structures, which we have briefly dealt with in the previous section. Let us briefly recapitulate with an example and its full set of dominance relations (its ρ -set):

57) Mary saw John run

$\rho = \langle\langle \text{see, Mary} \rangle\rangle; \langle\langle \text{see, John} \rangle\rangle; \langle\langle \text{see, run} \rangle\rangle; \langle\langle \text{run, John} \rangle\rangle$

This is a much simpler example than the crossing dependencies we saw in (40) (the famous *Jan Marie Piet sag helpen zwemmen*) and the discussion that follows. What we need to note here is that the matrix V *see* dominates *both* its nominal Object *John* (which we can identify as such by the hierarchy in (48)) *and* the embedded V *run*. As we said above, *run* does not satisfy the valency of *see*, rather, *John* does (in terms of getting a well-formed structure), however, because *John* is the subject of *run*, and it is this event (*John run*) that is the complement of *see*, we have an edge *from* the matrix V *to* the embedded V. We may note here that the formal properties of structural descriptions for ECM

¹⁷ Moreover, it is possible for the subject to surface in positions *internal* to the chain (see García Fernández & Krivochen, 2017); however, the only possible positions are those depending from *lexical auxiliaries* (see also Torrego, 1984):

- (i) ¿Qué *había* tenido Juan *que* decir?
 What had had J. to say?
 ‘What had John had to say?’

structures are very similar to those of structural descriptions assigned to some small clause complements, as in the examples in (58):

- 58) a. They consider John an idiot
 b. John wants Bill dead
 c. I like my tea strong

The ρ -sets for the sentences in (58) are given in (58'):

- 58') a. $\rho = \langle (\text{consider}, \text{they}); (\text{consider}, \text{John}'); (\text{consider}, \text{idiot}); (\text{idiot}, \text{John}') \rangle$
 b. $\rho = \langle (\text{want}, \text{John}'); (\text{want}, \text{Bill}'); (\text{want}, \text{dead}); (\text{dead}, \text{Bill}') \rangle$
 c. $\rho = \langle (\text{like}, \text{I}); (\text{like}, \text{tea}'); (\text{like}, \text{strong}); (\text{strong}, \text{tea}') \rangle$

We follow Rosenbaum (1965) and particularly Postal (1974) in considering the accusative NP in ECM and small clause (secondary predication) structures to be an argument of the embedded predicate, which gets *raised* to an object position in the matrix clause. Our model for Raising-to-Object, then, maintains the relation between the embedded predicate (V, A, P...) and its NP subject, *adding* an edge between the matrix predicate and that NP; the Rosenbaum approach, through Ross' (1969b) rules of *tree pruning*, adequately yields a structure in which there is no bounding node between the main predicate and the embedded predicate.

In this sense, -and jumping to the kind of last structure of the list in (51)- structural descriptions assigned to ECM structures are different from those assigned to Object-controlled Equi: the ρ -set for an Object-controlled Equi sentence like (59) is given below:

- 59) Mary ordered John to shut up
 Cycle 1: [Mary ordered John [Cycle 2]]
 Cycle 2: [John shut up]
 $\rho_1 = \langle (\text{ordered}, \text{Mary}'); (\text{ordered}, \text{John}') \rangle$
 $\rho_2 = \langle (\text{shut up}, \text{John}') \rangle$

Note that the matrix V *order* only dominates *John'*, not the embedded predicate *shut up*. The structural description for an Object controlled Equi sentence also features multiple cycles, as opposed to the single cycle in ECM: this correlates with well-known asymmetries between these constructions, like the following:

- 60) a. John wanted the guests to leave \rightarrow *What did John want?*
 * *Who did John want to leave?*
 b. John ordered the guests to leave \rightarrow ?? *What did John order?*
 * *Who did John order to leave?*
 c. John saw Mary cross the street \rightarrow *What did John see?*
 * *Who did John see cross the street?*
 d. John told Mary to finish the report \rightarrow * *What did John tell?*
 * *Who did John tell to finish the report?*

ECM constructions allow for the whole clausal complement of V to be made into Wh- and questioned, but not for the extraction of the subject of the embedded infinitive. In traditional IC terms, we capture the intuition that the accusative NP is a constituent of the same clause as the embedded infinitive, and that whole clause is what can be questioned. Object-controlled Equi, on the other hand,

features an accusative NP which is part of the embedded clause containing the infinitive, but *also* of the matrix VP. The latter structure, in GB/MP, requires the introduction of a null category PRO as the subject of the infinitive, which is controlled by the object of the matrix clause. The relevant IC segmentations are as follows, for a string $V_{fin}+NP_{Acc}+V_{inf}$ (see Rosenbaum, 1965 for a classic argument in terms of distinct *Deep Structures* for ECM and Object-controlled Equi; McCawley, 1998: 133, ff. for descriptive discussion and examples galore):

- 61) a. [... V_{fin} [NP_{Acc} V_{inf}]] (ECM)
 b. [... V_{fin} NP_{Acc} [PRO V_{inf}]] (Object-controlled Equi)

By allowing trails to visit a node more than once, we do not require the addition of a new terminal in the form of an empty category PRO (which, in turn, requires the introduction of a whole set of conditions ruling its distribution: Control Theory, including the rather controversial *PRO theorem*; see Bouchard, 1982: Chapter 5 for useful early-GB discussion).

In sum, ECM and Object-controlled Equi receive a different treatment in our theory, as desired (and this desire, it is worth pointing out, arises from the consideration of empirical facts).

We have deliberately left Subject-controlled Equi (a.k.a. just ‘Control’) for last, because of its problematic nature. During the Standard Theory and GB days, Raising vs. Control was considered to be an essential distinction which made use of two different mechanisms: movement vs. deletion in Equi-based theories; movement vs. lexical insertion of a null pronoun in GB and lexicalist theories (see Davies and Dubinsky, 2004: Chapter 1 for an empirical overview; Rosenbaum, 1965 and Postal, 1974 are classical references). But more recently, the distinction in the mechanisms involved in Raising and Control has been blurred. Hornstein (1999) and much subsequent work has argued that there is a single underlying mechanism, which is Movement, unifying Raising and Control. The differences between these structures would reside only in the fact that NPs acquire further theta-roles in the course of the derivation in Control, but not in Raising structures (because Raising verbs are not theta-assigners); this proposal entails the elimination of PRO and the proliferation of copies/traces. On the other hand, non-transformational frameworks reject a movement analysis (naturally): in LFG, Raising Vs take XCOMP (‘open’) complements: the SUBJ of XCOMP is also an argument of the matrix verb, and this is specified in the matrix verb’s *f*-structure (Bresnan, 2001; Dalrymple, 2001), and the relevant relation in Raising is actually *functional control*: the subjects of the matrix and embedded Vs need to have the same *f*-structure in Subject-to-Subject raising and the object of the matrix V and the subject of the embedded V need to have the same *f*-structure in Subject-to-Object raising (Dalrymple, 2001: Chapter 12). In the case of Equi Vs (a.k.a. Control Vs in more recent times), the complement is not open: we don’t have XCOMP, but COMP, with a PRO in the subject position of the embedded predicate and a relation of *anaphoric control* between the subject or object of the matrix V and PRO (in Subject- and Object-controlled Equi respectively), see Bresnan (2001: 297, ff.) and Dalrymple (2001: 325, ff.) for an accessible introduction to *anaphoric* vs. *functional control*.

The reason we have spent some time with LFG’s analysis is that in the framework sketched here there seems to be no reason to have different structural descriptions for Subject-controlled Equi and Raising if the difference can be codified in lexical specifications (at least in English): these specifications pertain to the interpretation of nodes, but not to the connections they establish with other nodes in a structural description. What these lexical specifications look like depends on the kind of lexicon that we want to have. This idea is compatible with so-called *Lexical entailment equi* theories (e.g., Dowty, 1978, 1985) and its development in Jacobson (1990): there is no movement involved in Raising or in Equi, but their differences arise as a matter of *lexical semantics*. Dowty’s take on lexically governed rules is particularly germane to the proposal advanced here, since

The key to this method [formulate transformations as lexical rules rather than as syntactic rules] is to treat Dative Shift, Passive, the Raising rules, and other rules that "change grammatical relations" as rules operating on individual verbs rather than on the full sentence phrase markers in which these verbs occur (Dowty, 1978: 393)

We disagree with Dowty in grouping Dative Shift and Raising with Passive, because it seems to us that only Passive is a true RCT, whereas Dative Shift and Raising do not *change* grammatical relations, they merely *add* a new relation without modifying existing connections (*contra* Dowty, 1978: 400). However, the analysis of (some) lexically governed transformations (e.g., Raising) in terms of lexical specifications is precisely the kind of framework that complements the one sketched here perfectly. The conception of the lexicon in Categorical Grammar (particularly in the versions of Dowty, 1978, 1979) seems to us to be a very good candidate for a theory of nodes, which –recall– correspond not to lexical items, but to basic expressions in the algebra of a language L (in our case, English).

Coming back to the comparison of Equi and Raising structures, for (62) and (63) we have the ρ -sets in (62') and (63') respectively:

62) Mary *tried* to finish the paper (Subject-controlled Equi)

63) Mary *was likely* to finish the paper (Subject-to-Subject raising)

62') $\rho = \langle (\text{try}, \text{Mary}'); (\text{finish}, \text{Mary}'); (\text{finish}, \text{paper}') \rangle$

63') $\rho = \langle (\text{be likely}, \text{Mary}'); (\text{finish}, \text{Mary}'); (\text{finish}, \text{paper}') \rangle$

Note that the Raising predicate *be likely* is a multi-word basic expression, such that it is *one node*, not two. There is, in this theory, no *structural* difference between Raising and Subject-Equi sentences: the differences arise when we consider the properties of the lexical items that govern the structures. This view is clearly incompatible with *endoskeletal* models (primarily, GB), in which syntactic structure *projects* from lexical requirements (this is crucially different from having *a-structure*, *f-structure*, and *c-structure* as *parallel* levels, as in LFG) via *Satisfy*, the *Projection Principle*, or some such operation (see, e.g., Chomsky, 1995: 187). There are differences in lexical specification, but *not* in syntactic configuration: in this way, we simply avoid the *movement vs. construal* debate surrounding Equi (see the papers in Davies & Dubinsky, 2007 for a good overview of that debate) because, essentially, the grammar in this theory is *not* generative; there *is* a single format for both Equi and Raising, but it is not a structural template nor a projection of lexical properties.

Raising structures present an interesting interaction with the licensing of polarity items, which we will only comment on here, much research pending. Negative Polarity Items (NPIs) like 'any' or 'ever' need to be licensed by a negative expression, be it an adverb like 'no', 'never', 'seldom', a quantifier like 'none', or an operator like *interrogation*, where –in the simple cases– *licensing* requires c-command such that an NPI must be within the c-command domain of (the minimal XP containing) a licenser (Horn, 1989, 1997; Ladusaw, 1980). Ladusaw (1980: 112) formulates the so-called *Polarity Hypothesis* as follows:

A NPI must appear in the scope of a trigger (a downward entailing element). If the trigger appears in the same clause as the NPI, the trigger must precede the NPI

There are two conditions in Ladusaw's hypothesis, one pertaining to structural order and another to linear order. Here, we are only interested in the former. In the context of this work, we will express the structural order condition as follows:

64) A node *a* appears **in the scope of** a node *b* iff:

- i. *There is a trail t such that b dominates a in t (as per **Definition 3**), and*
- ii. *There is no c , such that c is the root of a sub-graph including a but excluding b , and*
- iii. *Both a and b belong to the same graph at the point of determining scope*

May the following contrast suffice as an illustration (with licenser and licensee in **bold**):

- 65) a. *A man in love was **ever** happy (no trigger; thus, NPI is not licensed)
- b. **No** man in love was **ever** happy (**No** as a local, accessible trigger; the NPI is thus licensed)
- c. Which man $\text{COMP}_{[+Int]}$ in love was **ever** happy? (the interrogative operator $\text{COMP}_{[+Int]}$ as a local, accessible trigger; the NPI is thus licensed)

We will now analyze some situations in which we get unexpected licensing under standard assumptions about phrase structure. These assumptions, in what pertains to the present point, are the following:

- I. Non-complements are internally opaque (Chomsky, 1977; Huang, 1982; Uriagereka, 2002)
- II. A negative polarity expression *within an adjunct* cannot legitimate an NPI in object position (as a consequence of (I) plus the additional assumption that material in an adjunct to XP never c-commands material in object position in XP)

Assumption II is related to the fact that polarity scope relations are restricted to clausal domains, meaning, rooted sub-graphs (see Sternefeld, 1998a, b; more on this below). Non-monotonically assembled nonterminals, including *all non-objects*, need to be Chomsky-adjointed (Chomsky, 1955), thus, when we consider a *kernel sequence*¹⁸, they are still not there (see also the relative timing between *constituent structure rules* and *conjoining and embedding generalized transformations* in Fillmore's 1963 cycle).

It is thus crucial to determine what is an adjunct and what is not (which in turn implies a certain phrase structural configuration). In general, the distinction can be captured in IA as well as IP frameworks; Dowty (2003) formulates the essential features that distinguish *complements* from *adjuncts* in an admittedly simplistic fashion (but which captures the core aspects of the syntax and semantics of both grammatical functions):

A constituent Y in a phrase $[X Y]$ (or in $[Y X]$) is an ADJUNCT if and only if (i) phrase X by itself (without Y) is also a well-formed constituent, and (ii) X (without Y) is of the SAME syntactic category as phrase $[X Y]$. (X is in this case the HEAD of the phrase $[X Y]$.)

Then, a constituent Y in $[X Y]$ is a COMPLEMENT if and only if (i) X by itself (without Y) is not well formed, or else (ii) if it is grammatical, then X standing alone not have the same category as in $[X Y]$ (and does not have exactly the same meaning as it has in $[X Y]$)

If Y is an adjunct, the meaning of $[X Y]$ has the same kind of meaning (same logical type) as that of X , and Y merely restricts $[X Y]$ to a proper subset of the meaning/denotation of X alone.

Where Y is a complement in $[X Y]$, (i) the meaning of X by itself, without Y , is incomplete or incoherent. Else, (ii) X must be understood elliptically [...]

¹⁸ As a reminder, *kernel sequences* are strings of terminal nodes derived by means of phrase structure rules exclusively (Chomsky, 1955: 481). In contrast, *kernel sentences* are those which result from applying *obligatory transformations* to the terminal strings of the Σ, F grammar (Chomsky, 1957: 45).

Also, the same adjunct combined with different heads affects their meaning in the “same” way semantically (e.g. walk slowly vs. write slowly). But the same complement can have more radically different effects with different heads (e.g. manage to leave vs. refuse to leave).

(Dowty, 2003: 2-3)

Consider now the following examples, in which we will analyze the conditions for NPI licensing:

- 66) a. **No** policeman seemed to all of the reporters to have **any** chance of solving the case
- b. The police seemed to **none** of the reporters to have **any** chance of solving the case
- c. *The police seemed to have **any** chance of solving the case to **none** of the reporters
- d. *To **none** of the reporters, the police seemed to have **any** chance of solving the case¹⁹

The traditional picture, complying with the condition that an NPI (here, **any**) needs to appear within the scope of its trigger (here, **no/none**), where scope is defined in terms of c-command, is satisfied in (66 a), at least if we assume that a quantified NP of the kind [Q_{NEG} + N] counts as a trigger as a whole despite the fact that, strictly speaking, Q_{NEG} does not c-command the NPI. The opposite scenario arises in (66 c), in which (depending on where we assume the *to*-phrase is adjoined, more on this in a moment) either the NPI c-commands the trigger or the NPI is directly free (i.e., unbound). But it is the grammaticality of (66 b) that results of particular interest: note that the trigger appears within a phrase that does not seem to c-command the NPI; however, the NPI appears to be somehow (non-canonically) licensed. The status of the *to*-phrase in raising constructions is in general assumed to be that of an adjunct, in phrase structure grammars (and in the ‘mixed’ approach in Dowty, 2003): note that in general it can be omitted and when it is materialized its linear position is not as restricted as with complements. Let us illustrate the situation with a specific kind of copy-raising construction, *Richard* (Rogers, 1971):

- 67) a. (to me) it looks (to me) like Richard is in trouble (to me)
- b. (to me) Richard looks (to me) like he’s in trouble (to me)

The *to*-phrase in (66 a) above presents similar positional freedom:

- 66’) a. (to all of the reporters) **No** policeman seemed (to all of the reporters) to have **any** chance of solving the case (to all of the reporters)

This is to be expected, since the *to*-phrase does not intervene between the licensor **no** and the licensee **any**. However, when the polarity expression that licenses the NPI occurs within the *to*-phrase, things get more restrictive. Summarizing (66 b, c, d) in a single example, we get both peripheral positions yielding severely degraded or directly ungrammatical results:

- 66’’) (*to **none** of the reporters) The police seemed (to **none** of the reporters) to have **any** chance of solving the case (*to **none** of the reporters)

¹⁹ For ease of contrast, compare (66 d) with the ‘inversion-via-focus’ version, which sounds much better:

- i) To none of the reporters did the police seem to have any chance of solving the case

We will come back to this contrast briefly.

In the terms we have been working with here, *licensing* is a condition requiring the existence of a trail between licensor and licensee where both are accessible. In short, we seem to have elements to formulate the following condition in (68):

68) Licensing (first preliminary formulation)

A node a may license a node b iff $(a, b) \in \rho^$*

Which entails that there must be a *trail* between *a* and *b*. While this condition seems to be sensible, it is not enough: a trail between the peripheral occurrences of **none** and the NPI **any** is certainly formulable, but would yield incorrect empirical predictions. What else is needed, then? It seems that here the property that graphs are rooted comes in handy: **Definition 7** establishes that graphs contain at least one node which is not dominated by any other node within that subgraph; that is the *root*. This means that the traditional transformational distinction between *root* and *non-root* operations (e.g., Emonds, 1970) can be captured in the present framework. And how exactly does this help? To begin with, we propose that the relevant structure of (66'') is (69), using *S* to denote the root node for convenience:

69) (*to **none** of the reporters)_{S-Adjunct} The police seemed (to **none** of the reporters)_{VP-Adjunct} to have **any** chance of solving the case (*to **none** of the reporters)_{S-Adjunct}

That is: both left and right peripheral adjuncts are assumed here to be adjoined to the root, whereas the intermediate adjunct is somewhere in the VP, in phrase structure terms. Here, what matters is simply the root vs. non-root distinction, because if the adjunct is a sub-graph, we want to know in which positions the nodes that the sub-graph contains are accessible for operations outside that sub-graph, including licensing. Thus, we seem to need to reformulate the condition on licensing above to incorporate the nuances we have briefly discussed in this paragraph. The result goes along the following lines (with condition (iii) to be reformulated below):

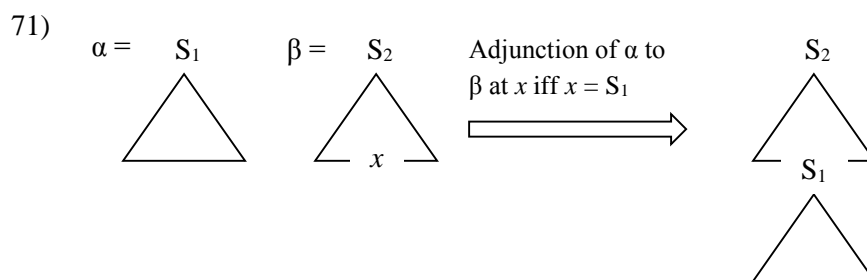
70) Licensing (second preliminary formulation)

$a \in A$ may license $b \in B$ iff

- i. $(a, b) \in \rho^*$, and*
- ii. B is not adjoined to the root of A , and*
- iii. There is at least a node in A that is identical to a node in B*

Where A and B are sub-graphs and a and b are nodes

Note that we have incorporated a few corrections: on the one hand, licensing is a node-to-node relation. Recall that nodes correspond to basic expressions, not lexical items: NPIs like *lift a finger* or *give a damn* are indeed multi-word basic expressions. Also, we added the condition that the licensee cannot be contained in a sub-graph that is adjoined to the root of the graph containing the licensor: left or right periphery is a matter of linear order, not of syntactic connections. It is essential to bear in mind that *root node* does not mean that we are dealing with a matrix clause, for there are embedded roots in the case of the application of generalized embedding transformations (in the sense of Fillmore, 1963; these could also be understood in terms of *substitution* in Joshi & Kroch, 1985); for instance:



S_1 is still the root of the sub-graph α (whose internal structure does not concern us now), which is what matters (see Joshi, 1985 for a detailed account of *adjunction* in phrase structural terms). Strictly speaking, (71) illustrates the *substitution* of x by S_1 under identity, should the TAG definition of *adjunction* be used.

Now, let us consider a contrast that we mentioned in passing above: that between (66 d), repeated here for convenience as (72 a), and the new example (72 b)

- 72) a. *To **none** of the reporters the police seemed to have **any** chance of solving the case
 b. To **none** of the reporters did the police seem to have **any** chance of solving the case

We will refer to the process that gives us the frontalization *without* inversion in (72 a) as *topicalization*, and frontalization *with* inversion as *focalization*. This choice of terminology is not innocent: the kind of argument we put forth here contrasts with Rizzi's (1997: 295-296) pertaining the mechanisms of *topic* and *focus*. In the light of the conditions (70 i-iii) above for licensing, we can interpret the contrast in (72), in which an NPI is licensed under frontalization + inversion (*focus*) but not under mere frontalization (*topic*).

Rizzi (1997: 295) observes that there may be more than a single *topic* per clause, but only one *focus*. He initially considers (but promptly proceeds to reject) the idea that the derivation of these structures differ in that only *topic* involves adjunction, but *focus* does not. There are, however, reasons to think that initial view may be correct for these instances. If *topic* involves adjunction, that means that we are dealing with (at least) two distinct sub-trees which are related by the creation of an edge between nodes in the adjoined object and the target of adjunction. Furthermore, the target of adjunction in *topic* structures seems to be the root, at least in the cases we have observed here. If this is so, then conditions (70 i-iii) are sufficient to filter out NPI licensing from a *topic*. But then what happens with (72 b)? Two options are logically possible at this point:

- I. *Focus* is adjunction, like *topic*, but to a non-root
- II. *Focus* is not adjunction at all

In a non-transformational framework, in which things do not move around literally leaving copies and traces, the relevant difference between adjunction to the non-root and no adjunction is that between *non-monotonic* structure and *monotonic* structure respectively: are we in the presence of more than a single derivational space, and thus, more than a single sub-graph? Here, we will argue that we are *not*. Much research pending, the polarity expression in the focalized constituent generates an intrusion effect with *focus*, but not with *topic*, such that something that should not be able to license an NPI, *de facto* does. We attribute this intrusion to the possibility that *focus* only involves a single sub-graph to begin with, which would be equivalent to proposing that the specific kind of *foci* that we see in (72 b) are *base-generated* (but this does not extend to other kinds of *foci*, see Jiménez Fernández, 2015 for a classification of *foci*, and Krivochen & García Fernández, 2018 for arguments in favour of a

multidominance analysis for *verum focus* in Spanish that is compatible with the *monotonic* view of *focus* proposed here, although under slightly different assumptions).²⁰

In this section we have presented a sketch for a treatment of English clausal complement constructions which seems to us to adequately capture relevant empirical generalizations while at the same time simplifying the theoretical apparatus. We have also taken advantage of the analysis of raising constructions and looked at aspects of NPI licensing under specific structural conditions. The next section will be devoted to the dynamics of VPs as well, but considering the interaction between auxiliary verbs and lexical verbs in Spanish. We will also take a look at the way in which dependencies with clitics work across sub-graphs, focusing on the ‘transparency’ or ‘opacity’ of parenthetical clauses, which by definition are not monotonically derived. We will argue that a strictly configurational theory of (non-)monotonicity in the grammar is too restrictive when it comes to predicting possible dependencies across sub-graphs.

6. Some considerations about Clitic Climbing and Spanish auxiliary chains

By and large, the model presented here has been devised as a descriptive tool in order to have fully explicit maps of node connectivity in English sentences. We have indeed warned the reader that extensions and applications to other languages are not automatic, because the focus of this model is not universality, but rather description of specific features of particular natural languages. In the present section we will, however, show that in principle there are problematic aspects of dependencies in Spanish sentences that can be captured using a natural extension of the model that we have sketched so far, an extension that we will then proceed to apply to further English data.

The graph-theoretic approach pursued here is strongly cyclic, and at the same time can straightforwardly describe trans-cyclic relations. We can take a sub-graph and embed it in another graph, provided that there is a node in the target graph that either dominates or is identical to a node in the embedded sub-graph. This much is not very different from *substitution / adjunction* in a Tree Adjoining Grammar (Joshi, 1985; Kroch & Joshi, 1985), but because there is no restriction that the target for embedding is the root of a sub-graph, we can speak of *generalized adjunction*, which also sounds rather cool. Recall now that we started with a distinction between *relation changing transformations* (RCT) and *relation preserving transformations* (RPT). Assume now that we have a string *abcd* that displays the following dominance relations:

$$73) \rho = \{(a, b); (a, c); (a, d); (c, d)\}$$

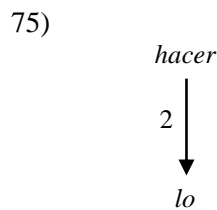
And there is an embedding transformation inserts the substring *efg* at *c* via *adjunction*, yielding *ab[efg]d* (‘pushing’ *d* downwards; Joshi, 1985: 209). A question arises: is *efg* opaque for purposes of operations at (i.e., triggered by an element in) *abd*? That is, can we create an edge from any individual element in the original string to some individual element in the adjoined string? If so, under which conditions?

²⁰ As we have emphasized above, the present framework does not directly pertain to or relate with morpho-phonology. However, a tangential piece of evidence could be called upon to strengthen the idea that only topics, but not foci, involve more than one sub-graph: in the contrast between (72 a) and (72 b), only in the former do we have a separate tonal unit for the frontalized phrase. This follows directly if we have root-adjunction under a principle like Emonds’ (1970: 9): ‘a root *S* immediately dominated by another *S* is set off by commas’.

We would like to propose, initially, that the insertion of an element that *only disrupts linear order* without changing constituency does *not* count as defining a cycle. Thus, we suggest the following preliminary constraints (from Krivochen et al., in preparation):

- 74) $[X... [...\alpha...]...Y]$ yielding $R(X, Y)$ is a legitimate configuration iff
- i. $[...\alpha...]$ has been introduced non-monotonically, **and**
 - ii. The rule introducing $[...\alpha...]$ is an order-changing rule (in the sense of McCawley, 1982: 94) [a relation-preserving transformation in the terms we use here]; that is, there is no element in α that either dominates or is dominated by a non-root node in the target of embedding

We will proceed to exemplify now. Let X be an argument clitic and let Y be its governing V ; thus, we have the following relation: $R(X, Y) = R(\text{Clitic}, V)$. What exactly is R ? Here we argue that a clitic-governing V structure is modelled in terms of dominance, thus, something like *hacerlo* ($\text{do}_{\text{INF}} \text{it}_{\text{CL}}$) displays the dominance relation $\rho(\text{hacer}, \text{lo})$. In Relational Grammar terms, we would have the following preliminary representation:



Or, using our *ordered* notation:

$$75') \rho = \langle (\text{hacer}, \text{PRO}); (\text{hacer}, \text{lo}) \rangle$$

These considerations hold at a very local level, and for *a-structure*. More globally, we need to consider V -clitic relations in a wider syntactic context, and constrain the possible dependencies that clitics can establish with otherwise *in abstracto* potential hosts. To this end, we need to make some assumptions explicit. To begin with, we will assume that a clitic is *ordered* (via direct dominance or transitive dominance) with respect to *all* other nodes within its *derivational current* (understanding this term as in Uriagereka, 2002; a local monotonically assembled domain / a minimal sub-graph containing the clitic and a suitable host, in this particular case), as *per* the definition of *order* above (**Definition 11**), to be substantially expanded on below (in **Section 7**). We assume that it is the ordering imposed over nodes in a tree that allows for a componential interpretation to be assigned to those nodes. If a clitic is not ordered with respect to a given node v , that node v (which may be the root node of an auxiliary tree) does not disrupt any relation within the minimal sub-graph containing the clitic and its *closest* host (where the length of a trail is measured as the number of edges in that trail²¹). This means, as a provisional observation, that parentheticals (if derived along the lines of Emonds, 1979; McCawley, 1982, via adjunction) should not count as intervening nodes for clitic climbing. This is indeed verified empirically:

²¹ The space in which our graphs exist is a(n Euclidean) *metric* space, thus, the familiar definitions hold (for d a metric over a set of points and x, y points):

- i. $d(x, y) \geq 0$ and $d(x, y) = 0$ iff $x = y$ (separation axiom)
- ii. $d(x, y) = d(y, x)$ (symmetry axiom)
- iii. $d(x, z) \leq d(x, y) + d(y, z)$ (metric inequality)

76) a. */?? Los hermanos se la dejan a Ana preparar *e* algunas veces (example from Emonds, 2007: 200)

The brothers CL_{DAT} CL_{ACC} let Prt Ana to-prepare some times

b. Los hermanos se la dejan, a Ana, preparar *e* algunas veces

(76) can only be acceptable if *a Ana* is derived in parallel, not monotonically. If this is the case, the clitic is ordered with respect to [*dejar*] and *preparar*, but *not* with respect to *a Ana* at the point of mapping dependencies, which is in fact predicted by our framework.

However, this is not (it cannot be) the full story: what counts as a ‘parenthetical’ (in the sense of ‘adjoined opaque domain’, whose opacity is, in traditional IC-PSG terms, given by the fact that they are not subcategorized for nor monotonically assembled) is not always clear. Initially, we would expect that (77 b) below would not work, but surprisingly, it does:

77) a. Juan puede, en realidad tiene que, hacerlo

J. may, in reality has to, do-CL_{ACC}

b. Juan puede, en realidad lo tiene que, hacer

c. *Juan lo puede, en realidad tiene que, hacer

The reason why we would expect (77 b) to be ungrammatical under a general view of what a ‘parenthetical’ is is that, if *en realidad tiene que* is derived in parallel and inserted by an order-changing rule (*a la* McCawley, 1982), then it should *not* be a suitable host for the clitic: either the clitic cannot move *into* the domain derived in parallel (because it is opaque, following Chomsky, 1977 in a Subjacency-inspired view, or Uriagereka, 2002 for a different perspective), or, if the clitic is base-generated, it cannot be thematically interpreted as an argument of the lexical verb (because opacity is a double edged sword: nothing comes in, nothing goes out). Further refinement is thus required.

In this context, speaking of ‘order-changing rules’ seems to be a more felicitous term than ‘parentheticals’, in the light of the following contrasts pertaining to the positional freedom of the relevant adjoined sub-graphs:

78) a. Los hermanos se_i la_j dejan, a Ana_i , preparar e_j algunas veces

b. Los hermanos se_i la_j dejan preparar e_j algunas veces, a Ana_i

c. A Ana_i , los hermanos se_i la_j dejan preparar e_j algunas veces

d. Los hermanos, a Ana_i , se_i la_j dejan preparar e_j algunas veces

79) a. Juan puede, en realidad lo tiene que, hacer

b. *Juan puede hacerlo, en realidad tiene que

c. *Juan lo puede hacer, en realidad tiene que

d. *Juan puede hacer, en realidad lo tiene que

In contrast with the positional freedom that we see in (78) with respect to the possible sites where the parenthetical *a Ana* can appear in the string, it seems to be the case that the rule that adjoins *en realidad (lo) tiene que* in (79) is more than simply an order-changing rule: the possible adjunction sites are very much restricted, depending on the semantic relation between the auxiliary in the adjoined phrase and the matrix clause. In this case, the lexical auxiliary in the adjoined clause is related to that in the matrix clause on a scale, such that ‘tener que’ means obligation, whereas ‘poder’ only means possibility. In this sense, linear order is relevant: it represents the scale along which the meanings of the modals are ordered, from ‘weak’ to ‘strong’. Note that when there is no scalar relation between the auxiliaries involved (because, say, we have a modal auxiliary *poder* and a phasal aspectual auxiliary *empezar a*), the construction ceases to be acceptable:

80) ??Juan puede, en realidad lo empieza a, hacer

J. may, in reality CL_{ACC} starts to, do

But even this slightly more nuanced account cannot be the whole story, even though it does give a partial account of the kinds of auxiliaries that can be adjoined in this manner. We still have unanswered questions, even in the descriptive front. Structurally, why is the parenthetical *a Ana* in (78) not an intervening object between the ‘gap’ and the coindexed clitic, but the adjoined auxiliary in (79) *is*? The possibility we will explore here is that if the adjoined object is a ‘self-contained’ unit, *it does not count as intervening* for operations at the target of adjunction. In other words, and appealing to TAG terminology: if an Auxiliary Tree is a *self-contained unit*, it does not intervene for purposes of operations at the Initial Tree to which the AT is adjoined. Of course, making what ‘self-contained’ means explicit is paramount. We define the notion as follows:

81) Self-containment (definition)

A graph G is a self-contained syntactic object iff $\nexists(x), x \in G$ such that

i. $\rho^(y, x)$ holds for $y \in G'$ and $G' \neq G$, and*

ii. x is an argument in G

That is: a graph is self-contained if it does *not* contain any node corresponding to an argument (i.e., a node that establishes with a predicate in G one of the relations in (49)) that is dominated by a node outside that graph (i.e., an argument of a predicate in G' , where $G' \neq G$). Note that *self-containment* is the notion that we appealed to informally in condition (70 *iii*) for licensing above, which we repeat here:

There is at least a node in A that is identical to a node in B

In this context, then, the final version of licensing conditions is as follows, including condition (70 *iii*) formulated in terms of *self-containment* (which we need anyway):

82) Licensing (final formulation)

$a \in A$ may license $b \in B$ iff

i. $(a, b) \in \rho^$, and [alternatively, there is a trail T that contains a and b]*

ii. B is not adjoined to the root of A , and

iii. Neither A nor B are self-contained

Where A and B are sub-graphs and a and b are nodes

If there is *no* domination relation, then the nodes in the ‘self-contained’ sub-graph are *not strictly ordered with respect to the nodes at the target of adjunction*, because there is no trail communicating those. In (79), the adjoined ‘parenthetical’ contains a node such that $\rho(\text{tener que}, \text{Cl})$ holds, as well as $\rho(\text{hacer}, \text{Cl})$. We can give the relevant set of *local* dominance relations for (79 a) as follows:

83) Cycle 1: $\rho = \langle (\text{poder}, \text{Juan}'); (\text{poder}, \text{lo}); (\text{poder}, \text{hacer}); (\text{hacer}, \text{Juan}'); (\text{hacer}, \text{lo}) \rangle$

Cycle 2: $\rho = \langle (\text{tener que}, \text{Juan}'); (\text{tener que}, \text{lo}); (\text{tener que}, \text{hacer}); (\text{hacer}, \text{Juan}'); (\text{hacer}, \text{lo}) \rangle$

The adjoined object, which corresponds to Cycle 2 in (79), is not ‘self-contained’ in the sense that it dominates a node that is also dominated by an element in another domain, in this case, the clitic *lo* is dominated by *poder* in Cycle 1, as well as *hacer*, which belongs to *both* cycles; *as per* the

McCawley/Levine approach to RNR. This last claim requires some unpacking: were we working within a transformational framework, the D-Structure of (79 a) would need to look like (79 a') in order to get a proper semantic interpretation:

79 a') Juan puede hacerlo, en realidad tiene que hacerlo (*D-Structure*)

In this case, we are interested in the fact that the clitic is a node which belongs to two subgraphs playing an argumental role in both of them (it is the direct object of the lexical verb *hacer*), but the clitic is part of a bigger sub-graph, which is *not* self-contained: the arc (*hacer*, *lo*). That is not the case with [a Ana], which is *the multidominated node itself*: it is thus free to 'move around', changing just the linear order between nodes (but not, we stress again, grammatical relations).

A crucial point (which takes us to the second part of this section's title) is that so-called 'clitic climbing' occurs across sequences of verbal predicates under specific conditions (for a Relational Grammar treatment, which is now classical, see Aissen & Perlmutter, 1983); in the case that interests us now, chains of auxiliaries (in the sense of Bravo et al., 2015). Here we use the expression *clitic climbing* in a purely descriptive manner, to denote strings in which a clitic's host is not the lexical item that takes it as an argument (without implying that the clitic has *moved* from one position to the other; see Ordóñez, 2012 for a general perspective). In previous works we defined an 'auxiliary chain' as follows:

84) An auxiliary chain CH_{AUX} is a string $\{x \wedge y \wedge z \dots n\} \wedge VP$ where

i) $\{x, y, z \dots n\} \in \text{Auxiliary Verb}$

ii) $n > 2$

(Bravo et al., 2015: 75)

In Bravo et al. (2015); García Fernández et al. (2017) and Krivochen & García Fernández (2017) we argued that there are empirical reasons to claim that auxiliary chains are not uniform in terms of their cyclic properties: some auxiliaries *block* the transmission of temporal and aspectual information, whereas others *let that information go through*. In Bravo et al. (2015) we referred to the former as 'Lexical auxiliaries', and to the latter as 'Functional auxiliaries'. More generally, lexical auxiliaries can *modify* lexical verbs and *be modified* themselves by functional auxiliaries (the latter of which carry temporal/aspectual information), whereas functional auxiliaries can *only modify* lexical heads (be these auxiliaries or verbs), but *not be modified* themselves. The following table summarizes the proposal:

Not affected (by virtue of being temporal information themselves)	Transparent / functional	Opaque / Lexical
<i>Ir a</i> (Eng. <i>be going to</i>), <i>acabar de</i> (in its 'recent past' reading, Eng. <i>have just -ed</i>)	Progressive <i>estar</i> 'to be', passive <i>ser</i> 'to be', perfective <i>haber</i> , Eng. <i>have -ed</i> .	Phasals (<i>empezar a; terminar de...</i>), second-position modals, first-position auxiliaries (<i>soler, haber de</i> 'have to'), <i>tardar en</i> 'take a long time'.

The distinction is empirically motivated, but it has far-reaching consequences for the format of phrase markers, which we have analysed in previous work, from both Immediate Constituent and Categorical

Grammars. We provided arguments that prove that a monotonic approach to sequences of auxiliary verbs (as in Cinque, 1999 within MGG and Bach, 1983 within CG) run into empirical problems when it comes to appropriately delimiting the domains across which modification takes place. As an alternative to the monotonic view of structure building we proposed that the syntax of chains of auxiliary verbs require us to move up and down the hierarchy of formal languages if we are to provide strongly adequate structural descriptions to strings featuring such chains. Here we simply want to stress that the empirical coverage achieved by the dynamical approach to chains of auxiliaries can be straightforwardly captured in the graph framework. Consider the following examples:

85) Juan tiene que empezar a trabajar

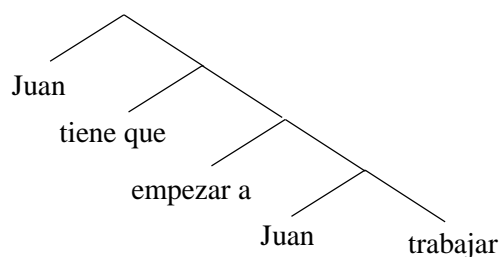
J. has to start working

86) Juan va a haber sido regañado

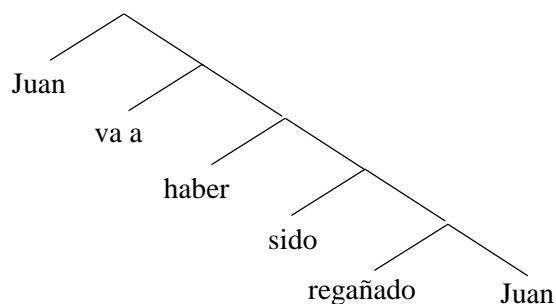
J. will have been nagged

(85) features a sequence of *lexical auxiliaries*: the modal *tener que* and the phasal aspectual auxiliary *empezar*, followed by the main lexical verb *trabajar*. (86), contrastingly, features a sequence of *functional auxiliaries*: *ir a* (which is a temporal mark), the perfective aspectual *haber*, diathesis *ser*, and finally the main lexical verb *regañado*. The phrase markers that a monotonic theory (which need not be IC; see Bach, 1983: 111 for a monotonic CG approach to auxiliary chains in English) assigns (85) and (86) is presented, just for the sake of comparison, in (87 a) and (87 b) (and whose empirical inadequacy has been discussed in detail in Bravo et al., 2015 and Krivochen & García Fernández, 2017). The full set of dominance relations for (85) and (86) is presented in (85') and (86') respectively:

87) (a)



(b)



85') $\rho = \{(Juan, tener\ que); (tener\ que, empezar); (empezar, trabajar); (Juan, trabajar)\}$

86') $\rho = \{(Juan, va\ a); (va\ a, regañado); (haber, regañado); (sido, regañado); (regañado, Juan)\}$

The first thing to note is that the dominance relations vary between (81) and (82), which –were we committed to a monotonic view of phrase structure– would be unexpected: this means that auxiliary chains manifest more than a single kind of computational dependency. The dependencies in (82') do not follow the kind of dominance relations that we would expect in a stepwise IA PSG. Rather, as pointed out in Bravo et al. (2015: 78), the dependencies resemble the situation in *unification*

grammars (Shieber, 1986), of ample use in non-transformational models (see also Jackendoff, 2011). The modification relations in (86), in predicate(argument) notation, are:

86'') va a(regañado); haber(regañado); sido(regañado)

Unify: (Aux 1(VP)) U (Aux 2(VP)) U (Aux 3, (VP))

Output: (Aux 1, Aux 2, Aux 3(VP))

Where the unification of two feature structures D' and D'' as the most general feature structure D , such that $D' \subseteq D$ and $D'' \subseteq D$. We notate this $D = D' \cup D''$ (Shieber, 1986: 14). Because the minimalist stepwise binary generative operation Merge (Chomsky, 1995 and much subsequent work) does not involve the identification of common feature structures, if we applied Merge to the modification sub-structures ('treelets') in (85''), we would expect to have three VPs in the output, contrary to fact (see Jackendoff, 2011: 276 for some discussion). Naturally, the identification of common components in treelets which are input to *Unification* is equivalent to visiting the same node more than once in a trail, as seen in (86'). This feature, together with the strong cyclicity entailed by manipulating minimal rooted sub-graphs, makes our graph proposal appropriate to generate structural descriptions for the output of Unification grammars, but it is also flexible enough to accommodate monotonic PSG-like structures should that be needed to account for semantic reasons. An example of such monotonic dependencies is (85') –a sequence of lexical auxiliaries–, in which the PSG tree does adequately represent modification relations: *tener que* modifies *empezar*, and *empezar* modifies *trabajar*. The phrase marker does grow monotonically, and the corresponding semantic representation is monotonic as well: we are dealing with a single rooted graph, rather than with the result of *Unifying* or *adjoining* more than one local structural domain.

Descriptively, then, our graph model can adequately represent the difference between the structural patterns yielded in auxiliary chains by lexical and functional auxiliaries; that is, we can provide descriptions which capture the heterogeneity of the class of auxiliary verbs that was empirically identified and theoretically accounted for in previous works (and, importantly, from *different* kinds of grammatical frameworks: IA and IP). In doing so, we formulated a further condition over cross-cycle relations (the notion of a *self-contained* sub-graph), which we will make use of in the analyses that follow.

7. On unexpected binding effects: a connectivity approach to Binding Theory

The point of the previous section was that a one-size-fits-all, purely configurational theory of parentheticals cannot derive adequate structural descriptions for all the empirical cases. There are instances in which an adjoined clause *does* intervene in the dynamics of the matrix clause, and there are also instances in which adjoined parentheticals are 'invisible' for operations at the target of adjunction; we proposed that the crucial factor to take into consideration was whether the adjoined domain is a *self-contained unit* or not in the sense of (81) above. Here we will present further evidence in favour of this 'mixed' approach to parentheticals (and other adjoined sub-graphs), which follows directly if the framework sketched here is adopted.

Consider the following example (taken from Bresnan, 2001: 81), with indexes and gaps added for expository purposes:

88) The one_j he_i should be spoken to e_i by e_j, for God's sake, is his_i mother_j

Here we have an interesting mix between a multiple gap construction like (25) above (Kayne's example *A person who people that talk to usually end up fascinated with*), and the syntax of

parentheticals. In Bresnan's example, [for God's sake] is a self-contained unit, and thus has its own syntactic and semantic independence. We can modify the example to show not only that parentheticals can be accessible from the matrix clause they adjoin to if they are *not* self-contained (i.e., if they contain a node that dominates a node also dominated by a node in the matrix clause or if they contain a node immediately dominated by a node in the matrix clause), but that there are restrictions over the materialization of nodes (i.e., their morpho-phonological exponent) which would be unexpected under an 'adjoined sub-graphs are always opaque' theory (cf. Uriagereka, 2002, in the most radical interpretation). Let us take a look at the following paradigm:

- 89) a. The one he_i should be spoken to by, for his_i sake, is his_i mother
 b. *The one he_i should be spoken to by, for $John's_i$ sake, is his_i mother
 c. *The one he_i should be spoken to by, for his_i sake, is $John's_i$ mother
 d. The one $John_i$ should be spoken to by, for $John's_i$ sake, is his_i mother

The cases we are interested in are (89 b) and (89 c). The traditional literature on Binding Theory would talk of a Principle C violation, in which the R-expression [John] is bound within its governing category. That proposal would indeed work for (89 c), in which the pronoun [he] and the R-expression [John] co-exist in the same derivational space. (89 b), on the other hand, poses the following problem: in order to blame its ungrammaticality on a Principle C violation, we need to be able to claim that the R-expression is bound, but how can it be? The R-expression is contained within a parenthetical, which cannot be monotonically derived. Two possible 'solutions' appear:

- I. The parenthetical is visible for purposes of dependencies between elements in the matrix clause by the principles of Binding Theory because these apply late in the syntactic computation, after adjunction (possibly, at LF; see Lebeaux, 2009)
- II. The parenthetical is visible because it is *not* a self-contained unit

Note that only I. requires a multiplication of levels of representation (to at least two: a syntactic level and LF, or, more generally, one where adjunction takes place and one where indexing takes place). But even if indexing and the computation of reference does indeed take place at a very late stage of the syntactic derivation, it is not clear how to appropriately filter out the cases in which parentheticals are completely opaque, for instance (90):

- 90) *What_i did John, who asked for e_i , get a book for his birthday?

In this case, the non-restrictive relative clause is opaque for purposes of extraction, and the ungrammaticality of (90) could be blamed on a violation of the Complex NP Constraint (Ross, 1967: 127); however, the CNPC in and of itself does not explain *why* S complements to NPs are opaque (and why in this case it is impossible to repair the island violation). In the derivational proposal advanced in Krivochen (under review), the opacity of an adjoined domain is a consequence of the combination of two factors: (a) *derivational timing* (relative to the ordering of embedding and singular transformations in a Fillmorean architecture) and (b) whether the adjoined sub-graph is a *self-contained unit*. In that work, which advanced an expansion of the syntactic-semantic machinery assumed in Generative Semantics, we proposed the following constraint on trans-derivational dependencies:

Let γ and β be two sub-trees such that γ contains a node X that corresponds to the root of β . A singular transformation T_S triggered from γ can affect β iff T_S is intrinsically ordered²² after an embedding transformation that adjoins β to γ at X .

What singular transformations (e.g., Wh-movement) *cannot have access to*, we argued, is *elements embedded within β* ; only β as a whole can be affected by a singular transformation at γ ordered after *adjunction* of β to γ . We committed ourselves to a model of syntax with multiple cycles (building on Fillmore), which crucially does *not* entail any commitment to a multi-layered model with several levels of representation and corresponding rules of interpretation.

Here we go farther in that line: we propose that a graph can be assigned a model-theoretic semantic interpretation *if and only if* it is self-contained; it should be evident that graphs corresponding to independent simple sentences are indeed self-contained (we will come back to this below, when raising questions about the place of *deletion* in this model). Note that we said ‘graphs’, not ‘sub-graphs’: the reason is that getting a graph G none of whose nodes is dominated by a node outside G may require the composition of more than a single sub-graph. If an interpretation for a graph implies (at least) walking a trail in that graph, we need all relevant nodes to be ordered with respect to one another. Informally, we say that if a node or a set of nodes is not ordered with respect to the graph G that we are walking at some point, then it is not possible to interpret that node or set thereof compositionally with respect to G (or any node or set thereof in G). We can formulate a general condition to this effect (much research pending):

91) Full Order Condition

A node v is ordered with respect to all and every other nodes in its ρ -domain within a cycle (i.e., within its minimal single-rooted sub-tree)

Ojeda (1987) formulates a ‘Full Leaf Order’ condition which can be thought of as the ‘linear precedence’ counterpart of our order condition, as follows:

Full Leaf Order: *For all leaves $u, v, u < v$ or $v < u$. [where ‘<’ is a binary relation ‘precedes’]* (Ojeda, 1987: 258)

The importance of an *order* imposed over a set of nodes in a graph being *strict* and *total* is also emphasized in Sternefeld’s connectivity-based approach to Binding Theory, which we will deal with below. Crucially, a binary relation $R(x, y)$ defines a strict *order* over a set or class $D(x)$ if and only if:

for all a, b, c ,
 $R(a, b) \rightarrow D(a) \wedge D(b)$; [irreflexivity]

²² A weak version of this prediction would refer to *extrinsic ordering* (a proposal along these lines is in effect suggested in Lakoff & Ross, 1967 [1976]; McCawley, 1968 opposes it). Ringen (1972: 267) presents the distinction between *intrinsic* and *extrinsic* ordering very clearly:

*If in a grammar, G , rule X is ordered before rule Y , then X and Y would be **extrinsically** ordered if G restricts how these rules can apply; that is, if these same rules could apply in the order Y before X in some derivations if not restricted by G . X and Y would be **intrinsically** ordered if there is only one order in which these rules could ever apply in any derivation; that is, if it would be impossible for these rules to apply in the order Y before X .*

For purposes of this work we will maintain the strong hypothesis (*intrinsic ordering*), but it is crucial to bear in mind that the problem requires very careful comparative and typological empirical analysis.

$\neg(R(a, b) \wedge R(b, a));$ [antisymmetry]
 $R(a, b) \wedge R(b, c) \rightarrow R(a, c).$ [transitivity]

$R(a, b)$ may again be written $a < b \text{ (mod. } R\text{)}$. (Krivine, 1971: 7)

The condition in (91) should be *completed* with the additional requirement that the *ordering* of a graph be *total* and *strict* (which in turn implies that for any nodes a, b or any sub-graphs G, G' , the order is *irreflexive*, *antisymmetric*, and *transitive*). The idea we put forth here is that when nodes are disconnected from a sub-graph (and are thus not ordered with respect to any other node in the relevant sub-graph) they *cannot be assigned a compositional interpretation*. In Ojeda's view, which is possibly the other side of the coin, a node that is not ordered with respect to any other node does not precede or follow any other node, thus it cannot be 'located' in a string. We are concerned with *trails* rather than *strings*, but it is rather clear how these conditions are at least compatible.

The conundrum we are facing with the paradigm in (89) seems to resist traditional accounts. In sum, the problem is that delaying the application of Binding Theory derivationally (say, to LF, after the 'overt' syntactic computation and covert operations) does not solve the essential, underlying problem: how come sometimes parentheticals are visible, and sometimes they aren't? However, if we combine the observations made within the context of BT with the non-monotonic perspectives on computation advanced here and in past works, we can enhance the descriptive and explanatory power of the grammar. In (89 b), the parenthetical [for John's sake] can form a chain with the other two occurrences of the same *type* (namely, [he] and [his]) because it contains a multidominated node, such that the following relations hold (we use { } and not ⟨ ⟩ for this ρ -set because it is *not* an ordered set):

92) $\rho = \{(\text{to, John}), (\text{John}, \text{should}), (\text{for, John}), (\text{John}, \text{mother})\}$

Because it is not a self-contained unit, the sub-graph corresponding to the structural description of *for John's sake* is visible for binding purposes (independently of *when*, in derivational time, BT principles are supposed to apply), and thus the materialization of the occurrence of *John* in the parenthetical as [John] does trigger a Principle C violation. Equally, we can see parentheticals (more specifically, non-restrictive relative clauses) triggering crossover effects²³:

- 93) a. Now that President Trump_{*i*} has been offered Mexico's help in the wake of Hurricane Harvey, he_{*i*} may be accepting assistance from a country full of 'bad hombres'.
 b. *Who_{*i*}, now that President Trump_{*i*} has been offered Mexico's help in the wake of Hurricane Harvey, *t_i* may be accepting help from a country full 'bad hombres'?

The ungrammaticality of (93 b) is unexpected (and remains unaccounted for) if non-restrictive relative clauses (a.k.a 'apositive relatives') are assumed to be opaque, and this holds regardless of the derivational timing of adjunction to the root (the 'Main Clause Hypothesis' of Emonds, 1979) or the establishment of a discontinuous dependency (McCawley, 1982, based loosely on Wells, 1947). Here, we combine considerations about derivational timing (essentially following Fillmore, 1963) with the relativization of the notion of representational opacity: a sub-graph is opaque if and only if it is a self-contained unit. We follow Postal (2004: 206) in rejecting any *explanatory* value in Principle C: the Principle does *not* explain why crossover effects appear, and it only *weakly* captures the facts descriptively. In an exhaustive account of these effects, not only issues of node connectivity can be

²³ We are indebted to Susan Schmerling for her help in constructing these examples.

considered: conditions like Wasow's (1972) *Novelty Condition*²⁴, which incorporate an element of linear order in the determination of the modes of presentation of *denotata* (what we could refer to the nodes' *Sinn*, utilizing Fregean terminology). If the theory exposed here is read with Harrisian glasses (or in the context of Lees and Klima, 1964), then it should be possible to formulate a rule of *pronominalization* which makes reference to the ordered occurrences of a particular node when walking a graph²⁵. Because edges are directed, we can indeed capture both forward and backward pronominalization (Ross, 1969a), both of which are formulated by Ross as *cyclic rules* (that is, they apply within sub-graphs). It must be explicitly said, however, that such an extension of the graph approach (that is, including aspects of the morpho-phonological exponents assigned to nodes in the graphs) requires the addition of a memory stack that can keep track of which nodes are visited when, and whether, at any given node, we have visited that node already. As of now, the consequences of adopting this view are still unknown. Constraints on forward and backwards pronominalization, which seem to interact with cyclic conditions (see, e.g., Ross, 1969a: 192), would have to be explicitly formulated, etc.²⁶ Personally, I would prefer to stick to the *vanilla* version of the theory, in which linearity in phono-morphology (i.e., the 'strings' generated by the grammar) does not really matter because what we are doing is formalizing connectivity in syntactic dependencies (creating a snapshot of a dynamical system, more specifically), but it is good practice, I think, to point towards alternatives when they could exist.

At this point, the reader can ask: is it possible to formulate a theory of Binding in graph theoretic terms? It is, and, at least partially, it has been done before (under different assumptions, we will come back to this momentarily). Consider, for instance, the definition of *Binding Tree* in Sternefeld (1998a: 158):

Given a tree Σ and a node $\alpha \in \Sigma$, the Binding Tree for α is the smallest subtree $T \subseteq \Sigma$ that satisfies the following conditions:

- a. $\alpha \in T$
- b. The root of T is the root of Σ
- c. If $\beta \in T$ and γ is the local trace of β , then $\gamma \in T$ **unless**

²⁴ Formulated as follows: 'An anaphorically dependent element cannot have more determinate reference than its antecedent'.

²⁵ For instance,

An NP₁ may pronominalize an identical NP₂ if NP₁ is to the left of NP₂ under any conditions, or if NP₁ is to the right of NP₂ and NP₂ does not command NP₁
(Bach, 1970: 121)

Where 'to the left of' simply means 'visited before'. Command, in the context of Bach's discussion, is Langacker's (1969) notion, which is based on the asymmetry of ρ : the ρ -domain of a vertex is an *ordered set*.

²⁶ There is also the obvious fact that pronouns can occur in contexts which do not satisfy the structural description for any *pronominalization*-like rule. Compare:

- i. John_i thinks that he_i is smart
- ii. He is smart

There is no reason to believe that 'He' in (ii) has been derived transformationally (or by visiting a node for an n^{th} time, say). This point, as far as I know, was first made by Postal (1969: 202).

- (i) α is an *R-expression*
- (ii) β reflexively dominates an adjunct that dominates α , and
- (iii) γ is not a reconstruction site

(Highlight in the original)

Sternefeld's definition in turn owes much to Barss (1986) and Lebeaux (1994), and is formulated in a late-GB kind of framework. We thus see a prominent role for the notion of *reconstruction*, which plays no role in the theory sketched here: on the one hand, we have *trails*, not *paths* (this means that we constrain the number of times we walk through each edge, not through each node). On the other, *reconstruction* is an essentially derivational operation, and there are no derivations in the framework explored here. Moreover, our nodes do not correspond to lexical items, as we have stressed before: rather, they are *IL translations of basic expressions*. This last consideration is related to Montague's (1974: 194) approach to the interpretation of basic expressions:

*a model should assign to a basic expression not a denotation but a **denotation function**, that is, a function that maps each infinite sequence of individuals onto a possible denotation of the expression* (Highlight in the original)

In this context, to formulate principles of binding that make reference to the morpho-phonological form of a node makes little sense. It is well known, besides, that the distribution of referential expressions does not always coincide with what is expected from binding principles: a clear example is the non-anaphoric interpretation of *himself* in the sentences that follow:

- 94) a. Mary_i complained that [the teacher gave extra help to everyone but herself_i] (from Baker, 1995: 64)
- b. Himself a man of science, John had to verify Bill's results (from Krivochen, 2017b: 20)

The point is, *anaphors*, *pronouns*, *R-expressions* are defined *by* and *for* binding theory in GB/MP, and it is not always clear (a) at which level each principle applies, and (b) whether binding primitives are morpho-phonological, semantic, or syntactic (i.e., if an anaphor is always a SELF object, or a semantically reflexive object, or an object that establishes a local syntactic relation with some other object). Here, things work slightly differently. Note that, of the three conditions defining a Binding Tree in Sternefeld (1998a), only conditions (c (i), (iii)) make explicit reference to non-configurational aspects: R-expressions in (c-i) and reconstruction (which in turn depends on movement) in (c-iii). Conditions (a) and (b) are locality conditions, which ensue that the Binding Tree for α is a minimal rooted tree, and condition (c-ii) can be thought of as a monotonicity condition: if β binds α in T if the trail between α and β does not visit a root R that includes α and excludes β ; or, to use Sternefeld's condition adapted to current notation, if $(R, \beta) \notin \tau$. If these considerations were to be incorporated, we can indeed formulate some interesting generalizations:

95) Binding Graph (definition)

Let the Binding Graph G_B of α be the minimal graph G that contains α and a root R

If α is bound within its Binding Graph, call α an 'anaphor' regardless of its morpho-phonological form

If α is bound outside its Binding Graph, call α a 'pronoun' regardless of its morpho-phonological form

Where

*β may bind α iff $IL(\beta) = IL(\alpha)$ and **either***

i. $(\alpha, \beta) \in \rho^*$, **or**

ii. $(\beta, \alpha) \in \rho^*$

We will see in **Section 9** that there are properties of the nodes (rather than of the pure configuration) that need to be considered when analyzing scope opening and closing, which introduces a further level of granularity into something like Sternefeld's (1998a: 151) *Scope Condition*, by acknowledging the need to distinguish between elements that can open scope and elements that can close it. These properties, however, do not configure an independent and separate system; rather, they need to interact with configurational constraints in order to be descriptively adequate and part of a consistent formal system.

All in all, the take-home message of this section and the previous one is that adjoined clauses are not always as opaque as MGG would have us believe. A configurational approach combined with an appropriate definition of 'opacity', here given in (81) in terms of *self-containment* seems to be preferable to purely configurational approaches in which the opacity of certain labelled domains is defined *a priori*. In connection to these considerations, the following section will deal with aspects of complementation within the NP, more specifically, aspects of the syntax of non-restrictive (a.k.a. *appositive*) relative clauses and their similarities with what we have called *parentheticals*.

8. Complementation within the NP: on appositive relatives

In this section we will deal with some aspects of the syntax of appositive relative clauses. We will follow Emonds (1970, 1979) in considering appositive relatives to be 'main clauses' (i.e., initially independent sub-graphs, adjoined non-monotonically to the sub-graph that contains the antecedent of the relative), but unlike Emonds, we do not assume that the rule of Parenthetical Placement changes grammatical relations. We thus side with McCawley in that '*parentheticals are placed by that changes word order without changing constituent structure*' (1982: 95); however, in the next section we will see that there are exceptions to his claim that '*all grammatical phenomena to which the constituency of the [target of parenthetical placement] is relevant behave as if the parenthetical were not there*' (1982: 96), which reveal some interesting aspects of the conditions we can impose over relations among sub-graphs.

This section begins with an uncontroversial observation, which is that appositive relative clauses are non-recursive. The point is illustrated by Emonds and McCawley with examples like (96 a), to be contrasted with (96 b):

- 96) a. *John, who goes to MIT, who likes math, will get a job
b. People who go to MIT who like math will get jobs (Emonds, 1979: 222)

Note that it is possible to save (96 a) in a strictly paratactic reading, but in that case the structural description is different. In Emonds' own terms,

the string John, who goes to MIT is not a constituent coreferential with the pronoun following (that is, the string is not even a constituent) under the MCH [Main Clause Hypothesis] (Emonds, 1979: 222)

This means that an appositive relative cannot take as an antecedent a structure of the kind [NP [Appositive]]. In a word, *appositive relatives are non-recursive*. An appositive, however, can take as an antecedent an NP modified by a restrictive relative clause (see (97 a), below).

Restrictive relative clauses (both Wh- relatives and *that*-relatives), on the other hand, *are* monotonically recursive, in the sense that they can take an [NP [Restrictive]] structure as their antecedent. (97) illustrates this fact, with all possible combinations of Wh- and COMP:

- 97) a. [[Every psychologist [**that** Emily talked to]]_i **who**_i insisted on helping] ended up making a disaster
 b. [[Every psychologist_i [**who**_i Emily talked to]] **that** insisted on helping] ended up making a disaster
 c. [[Every psychologist_i [**who**_i Emily talked to]]_j **who**_j insisted on helping] ended up making a disaster
 d. [[Every psychologist [**that** Emily talked to]] **that** insisted on helping] ended up making a disaster

McCawley (1998: 435) explicitly claims that because restrictive relative clauses (RRC) are sister-adjoined to N',

The structure with the restrictive relative modifying an N' yields the correct prediction that restrictive relative clauses can be stacked, since there is nothing to prevent the N' of an [N' N' Comp'] [read: S'] from itself having the form N' Comp'

Perhaps more important is the observation that [N' Comp] behave, under certain tests, like *constituents*: transformations targeting *N* can (and in some cases *must*) equally target *N Comp*. Pronominalization is one such transformation:

- 98) Tom has [[a violin]_i which once belonged to Heifetz]_j, and Jane has one_{?i/j} too (example taken from McCawley, 1998: 445; indexes and judgments are ours)

Stripping (a.k.a. VP ellipsis) is another one ((99 a) is taken from McCawley, 1982: 96; (99 b) is taken from McCawley, 1998: 450):

- 99) a. John sold Mary, who had offered him \$600 an ounce, a pound of gold, but Arthur refused to Ø. (Ø = refused to sell Mary a pound of gold; ≠ refused to sell Mary, who had offered him \$600 an ounce, a pound of gold; ≠ refused to sell Mary)
 b. John sold Mary, who had offered him \$600 an ounce, a pound of gold, and Arthur did Ø too (Ø = sold Mary a pound of gold; ≠ sold Mary, who had offered him \$600 an ounce, a pound of gold; ≠ sold Mary)

Also relevant for our purposes is the observation that an N' plus a restrictive relative can be conjoined with another N' plus restrictive relative (100 a); that is not possible with appositive relatives (100 b) (examples taken from McCawley, 1998: 446):

- 100) a. [Some violins [_{RRC} that have been owned by famous performers]] and [flutes [_{RRC} that were played by Frederick the Great]] are going to be sold at auction.
 b. [*These violins, [_{Appos} which were made by Cremonese masters]], and [pianos, [_{Appos} which were made in nineteenth-century Paris]], are expected to fetch high prices.

Appositive relative clauses can co-appear with restrictive relative clauses, albeit under specific conditions, for instance:

a restrictive can follow an appositive if it is the only constituent that follows (Emonds, 1979: 222)

Examples of this structure are presented in (101):

- 101) a. We found that movie, which cost plenty, that you so highly recommended (Emonds' ex. (22))²⁷
b. It was Fred, who you met at my party, that I was just talking to on the phone (McCawley, 1998: 449, ex. (19 a))

We may add that given the adjacency condition imposed over Bare Relatives (a.k.a *abridged* relative clauses, or S' subjected to *Whiz-deletion*) we will only deal with Wh-relatives and *that*-relatives in this section, because we are interested in the recursive properties and relative positioning of restrictives and appositives, and bare relatives are neither recursive nor can they be in any position that is not strictly adnominal (although, as usual, they are attested; see e.g., McCawley, 1998: 433, ex. (15 b)).

An appositive can follow a restrictive relative clause even if it is not the final constituent:

- 102) a. The children that you brought, who were charming, got sick later (adapted from Emonds, 1979: 222)
b. *The children, who were charming, that you brought got sick later (Emonds' ex. (21))

Is there any way in which the syntactic approach advanced here can shed light on the distribution and properties of restrictive versus appositive relatives? We believe there is, and that the differences in distribution as formulated by Emonds (and also illustrated by McCawley) can be indeed accounted for if we assume that (a) structural descriptions for English strings can take the form of maximally connected graphs, and (b) the relevant conditions over graph well-formedness from syntactic and semantic points of view are expressed in terms of possible and impossible connectivity patterns between nodes and between sub-graphs. In this sense, recall that above we introduced the notion of graphs being *ordered*, and the condition that order be *strict* and *total* (see (91) and the discussion that follows); these conditions apply to nodes as well as sub-graphs. This means that, given a string featuring a number of relative clauses (note that we have not qualified 'relative clauses'; in principle they could be either restrictive or appositive), the graph G corresponding to the structural description of that string will be well-formed if and only if there is an unambiguous (i.e., unique) strict ordering $O(G)$.

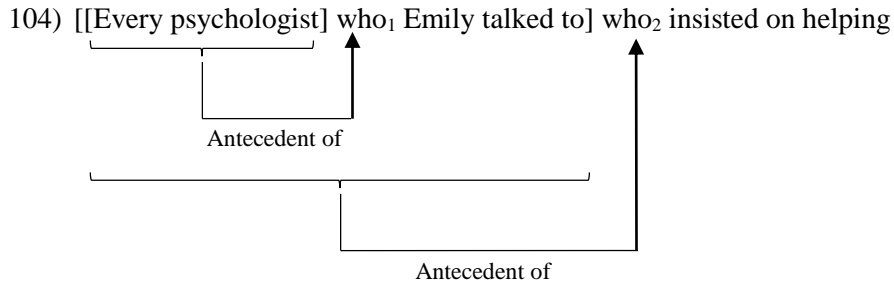
A notion of ordering is inherent to McCawley's observation that a restrictive relative clause can take a structure [N' COMP] as an antecedent (however the reader chooses to represent those nodes, we will just use the term CNP henceforth). Let us consider (97 c) again, repeated here as (103):

- 103) Every psychologist who₁ Emily talked to who₂ insisted on helping ended up making a disaster

²⁷ Not all our informants were happy with this sentence, but we are keeping Emonds' own judgment nonetheless. If the antecedent NP appears in subject position, the acceptability of the sentence decreases to the point of being ungrammatical, as noted by Emonds:

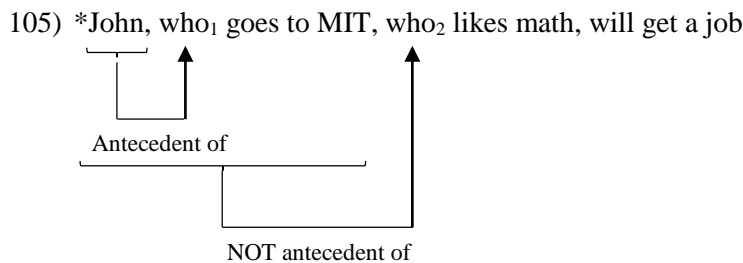
- i) ??/*That movie, which cost plenty, that you so highly recommended ended up being a disappointment

The antecedent of who_1 is the NP *Every psychologist*; the antecedent of who_2 is the CNP *Every psychologist who Emily talked to*. The ‘size’ of the antecedent grows monotonically with the introduction of new RRC, and the ordering between these is unambiguous. Let the interpretation of a definite description proceed in a Russellian fashion, such that *Every psychologist* is a function $f(x)$; then, the interpretation of who_1 must also be $f(x)$, because *every psychologist* and *who* are the same node visited twice in a trail. Then, we have who_2 , which is a second order function insofar as its interpretation is a function over $f(x)$, call it $g(y)$ where $y = f(x)$. The point here is that the size of the antecedent grows linearly, and so does the function corresponding to the interpretation of antecedents. Graphically,



The monotonic growth in the structure gives us precisely the kind of strict ordering that we require as a well-formedness condition on graphs. In this case, the order depends on n -order functions corresponding to the interpretation of the relative operators.

The question now is, what happens in the case of appositives? The argument we put forth here is that appositive relatives are *not* monotonically recursive *because there cannot be a strict ordering for a set of recursive appositive relatives*. Note that the diagram that we used for (103), with its monotonically growing antecedent size does *not* correspond to the interpretation of (96 a), repeated here:



It is obvious that the antecedent of who_2 is *not* *John, who goes to MIT*, but just *John*. But that is the same antecedent as who_1 : whatever relation exists between *John* and who_1 also exists between *John* and who_2 , and neither of these takes the other as an argument; there are no first- and second-order functions in the correct representation for (104), as opposed to the situation in (105). This means that there cannot be a strict ordering (total, antisymmetric, irreflexive, intransitive) between these two nodes who_1 and who_2 in the appositive case; in turn, this means that there is no strict ordering between the sub-graphs which contain these nodes, because if there was, then the nodes would be transitively ordered (recall that *transitivity* is a condition for *strict ordering*). That is: if $G \ni who_1$ was ordered with respect to $G' \ni who_2$, then who_1 would be ordered with respect to who_2 , contrary to fact.

We said that there cannot be a strict ordering over a set of appositives such that we get a monotonically recursive interpretation (following Emonds and McCawley); it must be noted that the reading in which multiple appositives following an antecedent NP are strictly paratactic receives a different structural description. In computational terms, recursive restrictive relative clauses can (and in fact must) take as antecedents ever-growing structures [N' COMP]: it is *not* possible to have a

sequence of restrictive relatives in which all take the same N terminal as their antecedent, ignoring other relatives. To give a concrete example, the segmentation and indexing in (106) is impossible to obtain:

106) [[Every psychologist]_i [who_i Emily talked to] [who_i insisted on helping]] ended up making a disaster

But that is precisely the kind of structural description that we have in the case of several *stacked* appositives, as in (107):

107) Donald Trump is a man who will spare no effort to get different parts of the country to hate and fear each other, who will do everything he can to damage the U.S. position in the world, who will set things up so his family members benefit financially from his presidency, ...

The only possible structural description is one in which dependencies are strictly paratactic; computationally, a *finite-state* sequence:

107') Donald Trump is [a man]_i [who_i will spare no effort to get different parts of the country to hate and fear each other], [who_i will do everything he can to damage the U.S. position in the world], [who_i will set things up so his family members benefit financially from his presidency], ...

What we have here is simple *head-tail* recursion (Uriagereka, 2008: 228; see also his finite-state treatment of iterated small clauses in 204, ff.). In the terms of Krivochen & Schmerling (2016), the appositive relatives in (107) and examples of the sort are *que-coordinated*, each a state in a *Markov chain*. The order among the appositives in cases like (107) is strictly linear, with parataxis being the only structural option to save the representation.

Let us now turn to ‘mixed’ cases, in which we have an appositive and a RRC to be ordered. Relevant examples are like (101 a) and (102 a), repeated here as (108 a) and (108 b) respectively (with minor annotations added):

108) a. We found that movie_i, which_i cost plenty, that you so highly recommended e_i
b. The children_i that you brought e_i, who_j were charming, got sick later

As noted in fn. 24, (108 a) elicited mixed responses from our informants. Emonds’ condition for a restrictive to follow an appositive was that the restrictive clause was string-final (i.e., if there is no other constituent after that). The reason why (108 a) is not fully acceptable may have to do with the fact that there is no growth in the size of the antecedent for the restrictive relative: both the appositive and the restrictive relative have exactly the same antecedent *that movie*. However, there are semantic differences between the two clauses which allow for an ordering (but *not a strict ordering*) to be imposed in interpretation. To the extent that (108 a) is grammatical and acceptable, that acceptability is accounted for outside of the grammar as it is conceived of here; at least if the grammar includes a condition on *strict node ordering*.

The example (108 b), however, receives a simpler analysis, due to the fact that there is indeed a possible strict ordering between the restrictive and the appositive clauses. In this case, the antecedent of the restrictive clause is the NP *the children*, whereas the antecedent of the appositive clause is the complex [NP S’] (note: *not* [N’ COMP], as in McCawley’s quotation above). The condition imposed by Emonds seems to be relevant for these cases as well, for we *cannot* have (109 b) as an extraposed version of (108 a) –which does not mean at all that (109 b) is ungrammatical-:

109) a. The children that you brought which your sister loved, who were charming, got sick later

- b. The children that you brought, who were charming, which your sister loved, got sick later

The relevant interpretation for (109 a) is that there is a set of children, a subset of which were brought, and a subset of this subset were also loved by someone's sister: we have a monotonically recursive restriction over the extension of the set of children. Call the set resultant from this double restriction *s*. Then, a property is assigned to the members of *s*, namely, that of being charming. Syntactically, the antecedent of the first restrictive relative clause, *that you brought* is simply the NP *the children*; the antecedent of the second restrictive relative is (as we would expect), the CNP *the children that you brought*. So far so monotonic. The antecedent of the appositive clause is the whole CNP *the children that you brought which your sister loved*; there is a straightforward ordering to be imposed among the sub-graphs (and transitively, all the nodes in each of these sub-graphs) in (109 a). But the same interpretation procedure is not available for (109 b): the relative clause *which your sister loved* does not receive a restrictive interpretation; rather, it is interpreted as another appositive: the set of children who were charming and the set of children that someone's sister loved are strictly co-extensive. In this case, the finite-state syntax for stacked (*head-tail* recursive, not *truly* recursive) appositives that we proposed above is called upon again.

In this section we have analysed aspects of complementation within the NP: we have derived empirical restrictions on the distribution and combinatorial properties of restrictive and appositive relative clauses (originally noted by Emonds, 1979 and McCawley, 1983, among others) from independently motivated requirements of strict ordering to be imposed over a graph in order to assign a compositional interpretation to that graph. Monotonic growth (the case of truly recursive restrictive relative clauses) yields a straightforward order by having relative operators be functions of n^{th} order in Context-Free (or higher) semantic representations. Contrastingly, appositive clauses can only be *stacked* in finite-state sequences, displaying strictly paratactic dependencies.

9. Islands and lexical restriction: opening and closing scope²⁸

We briefly dealt with Wh-interrogatives above, but in those accounts we treated Wh-phrases as an internally unanalysable whole. Thus, Wh-phrases like *what* and *which of the books* would have received the same treatment in the initial presentation of Wh-dependencies. While that uniformity assumption ('all Wh- are created equal') helped us simplify the exposition, there are interesting data whose analysis requires further refinement. The graph-theoretic proposal advanced in this work interacts in interesting ways with paradigms concerning Wh-movement, islandhood, and lexical restriction. We speak of lexical restriction in Wh-phrases, when we have structures of the type [Wh-NP], in which the Wh-word quantifies over an overt NP. Following usual practice, we will refer to the NP as the 'lexical restrictor'²⁹.

Consider now (110) and (111):

110) What did who prefer?

111) Which class did which student prefer?

²⁸ This section owes much to Julie Franck, Whit Tabor, and Sandra Villata. Mistakes and misinterpretations are mine alone, though.

²⁹ In some varieties of LFG, like the one exposed in Bresnan et al. (2016), structures like [Which class] are referred to as 'operator complexes'. The Wh-element is the operator, and the whole [Wh-NP] structure containing the operator is the operator complex.

In this paradigm, (110) is usually considered degraded with respect to (111) (see Villata et al., 2016 for a psycholinguistic perspective). Rizzi (1990) –and much subsequent work– analyzes Wh-islands like (110) as violations of Relativized Minimality³⁰, but accounting for the contrast between (110) and (111) requires the introduction of further complications in the theory, particularly pertaining to the internal composition of Wh-phrases and the level of granularity at which RM applies (and raising the question of what the syntactic computation has access to); furthermore, ‘displacement-as-movement’ and appeal to the Copy Theory of movement are unavoidable in the MGG view (including Rizzi’s; see for instance the analysis of the paradigms that Villata et al. worked with in Rizzi, 2017). We would want to avoid those complications while remaining faithful to our goal to construct an explicit (*but not necessarily psychologically real*³¹) model of the grammar of English (and fragments of Spanish). Can we say something about this contrast?

First, let us make explicit the domination and connectivity relations in the graphs corresponding to the structural description of each sentence (as a reminder, we use $\langle \rangle$ for *ordered* ρ -sets and $\{ \}$ for *unordered* ρ -sets):

110’ $\rho = \langle (\text{prefer}, \textit{who}); (\text{prefer}, \textit{what}); (\text{did}, \text{prefer}) \rangle$

Note that: $(\textit{what}, \textit{what}) \notin \rho$

111’ Cycle 1: $[\text{Which class}] \rightarrow \textit{class}'$ (i.e., $\lambda\text{PP}\{\wedge\textit{class}\}$)

Cycle 2: $[\text{which student}] \rightarrow \textit{student}'$ (i.e., $\lambda\text{PP}\{\wedge\textit{student}\}$)

Cycle 3: $[\textit{student}' \text{ prefer } \textit{class}']$

$\rho_1 = \{(\textit{which}_1, \textit{class}'); (\textit{class}', \textit{which}_1)\}$

$\rho_2 = \{(\textit{which}_2, \textit{student}'); (\textit{student}', \textit{which}_2)\}$

$\rho_3 = \langle (\text{prefer}, \textit{student}'); (\text{prefer}, \textit{class}') \rangle$

Note that: $(\textit{which}, \textit{student}') \in \rho$; $(\textit{student}', \textit{which}) \in \rho$

Why do we want there to be an edge *from* the Wh- *to* the NP restrictor and another *from* the restrictor *to* the Wh- element? Essentially, to capture the two relations that arise here: the Wh-element quantifies the NP, whereas the NP semantically restricts the Wh-element³².

Now, let us further assume that Wh-elements open scope, and that said scope needs to be closed, an assumption that is not alien to computer languages (so, for instance, [needs to be closed by] to have a well-formed code in languages like C and Java; also *if-then-else-end* sequences in pseudocode). Initially, this is not very different from Koopman & Sportiche’s (2000) Bijection Principle. The gist is that *open scope counts as vacuous quantification*, and as such needs to be

³⁰ We can formulate the principle as follows: in a configuration like (i), a local structural relation cannot hold between X and Y if Z is a potential bearer of the relevant relation and Z intervenes structurally between X and Y (X c-commands Z, Z c-commands Y, and X c-commands Y transitively):

i) X...Z...Y...

³¹ This statement may sound contentious to some, but it is actually a very common assumption in theories of language that do not identify themselves with the goals of the so-called ‘biolinguistic’ enterprise, for instance, Ajdukiewicz- and Montague-style Categorical Grammar (see Dowty, 1978: fns. 2 and 3).

³² In LFG, an operator can inherit the syntactic rank of its operator complex (Bresnan et al., 2016: 225). For that inheritance to take place, we can hypothesize, the operator and its complex must be connected at *c-structure*.

lexically bound. The conditions for closing scope in linguistic structures, in this preliminary version, are the following:

- 112) i. For some v' to close the scope opened by v , then $(v', v) \in \rho^*$, must hold
- ii. For some v' to close the scope opened by v , it must be the case that $v \neq v'$

All in all, the two conditions in (112) can be understood in relation to what Sternefeld (1998a: 151) refers to the *Scope Condition* (based on much previous work, see Ladusaw, 1980; Barss, 1986; Lebeaux, 1994, among many others), which he formulates as follows:

NPIs as well as bound variables must be [in] the scope of (i.e., c-commanded by) the operators they depend on

There are, however, two crucial differences between Sternefeld's condition (which we take to be representative of a whole tradition, the GB-MP approach, also shared by non-transformational frameworks like LFG; see Darlymple, 2001: Chapter 11 §2; Bresnan, 2001: 212, ff.). One is that dominance in our framework is *not* defined in terms of c-command in L-trees, but rather in terms of the existence of a directed edge (or a set thereof, configuring a *trail*) between the relevant nodes. The other is that we establish a further condition on *what kind of element* can close scope. This condition is based on a classical distinction between two classes of grammatical objects: *categorematic* and *syncategorematic*.

The closing element cannot be a Wh- element or a relational element (by the condition noted in (110') and by (112 ii), which amount to saying that an operator cannot bind itself), so we do have certain restrictions over the nature of v' . At this point, we need to introduce a very useful distinction which has a long history:

...partes igitur orationis sunt secundum dialecticos duae, nomen et uerbum, quia hae solae etiam per se coniunctae plenam faciunt orationem, alias autem partes συγκατηγορήματα, hoc est consignantia, appellabant (Priscian, *De Partibus* §54.5)

Quite more recently, Schmerling (2016) very clearly summarizes the notion of *syncategoremata* as follows:

The designation syncategorematic pertains to material that has no meaning on its own and does not itself belong to any syntactic category but gets its meaning from the words in its context

Wh-elements do *not* have a meaning of their own, yet every expression in which they occur *does* have a meaning. They are thus *syncategorematic* in the Priscian tradition. Lexical NPs, on the other hand, are *categorematic*, by virtue of having a descriptive meaning of their own; and so are Vs and (possibly) *lexical auxiliaries* (but not *functional auxiliaries*). A syncategorematic element cannot close scope, but a categorematic element can, by virtue of its lexical content. Thus, NPs can close scope, but so can VPs, or, rather, *argumental positions* licensed by Vs (because these correspond to categorematic elements) within VPs. Lexical predicates select for categorial *and* semantic information, thus establishing contextual restrictions with respect to their co-occurring elements. In a theory that encodes displacement as constituent movement, these restrictions also pertain to the gaps left behind after movement. Let us now see how this machinery can assist us in the analysis of the contrast in (110)-(111).

In (110) (*What did who like?*), *what* open scope, and so does *who*. The 'gap' licensed by the verb *like* (that is, a position required to satisfy the predicate's *valence*) can close only *one* of those.

Thus, by the time we finish walking the trails, there is an unsolved open scope, which degrades the sentence.

But that is not the case if there is a lexical restriction, because the categorematic element can close the scope opened by the Wh-element *in situ*: within the minimal sub-graph in which *both* elements appear, *students* can close the scope of *which*₂, and *class* can close the scope of *which*₁. This means that the structural conditions required for semantic interpretation are divorced from scope, such that there are no ‘filler-gap’ dependencies: there are only vertices which we can visit twice or more for different purposes. In the case of (111), we go from *prefer* to *class*’ for thematic reasons (and, possibly, Case).

The position licensed by the V, should there be one, is required for thematic interpretation *but* not for syntactic reconstruction (i.e., filler-gap dependencies understood transformationally). In this state of affairs, we can ask about a sentence like (113):

113) *Which building do you wonder which engineer slept?

Sleep, as an unergative V, does *not* give us an appropriate theta-/a-structure that ensures that (113) can be mapped onto any ‘conceivable’ state of affairs. It also doesn’t license a position for [which building], which in the terms we have introduced here means that there is no edge from *slept* to *which building*. Scope is closed locally within the treelet *which building*, but that treelet can receive no semantic interpretation. The problem is thus one of semantic selection, not of hard ungrammaticality.

In sum, *requirements of scope are divorced from thematic/a-structural requirements*. Scope can be satisfied locally in Wh-phrases with lexical restriction (in LFG terms, the scope of an operator can be satisfied within that operator’s complex; see Bresnan, et al., 2016). If we go back to Ladusaw’s (1980) *Polarity Hypothesis*, paraphrasing and adapting to the present context, we have the following:

114) Wh-hypothesis

A variable must appear in the scope of a trigger, where:

- i. There is a trail T including both the variable and the trigger, and*
- ii. The variable is ordered with respect to the trigger in T such that (trigger, variable) ∈ ρ**
- iii. The position associated with the variable is assigned an independent meaning by semantic interpretation rules*

Consider again example (111), *which class did which student prefer?* In order to have a well-formed graph, first we need to ensure that there is a way to get to the variable from the trigger, then, that the trigger dominates the variable. Both these conditions are satisfied in the operator complex (or ‘trigger complex’, adapting LFG terminology to the one used here); since we have ρ(which, *student*’) and ρ(*student*’, which); same with *which class*. Now, by ‘independent meaning’ in condition (114 iii) we mean nothing else than the requirement that the variable be categorematic. In this case, the *semantic interpretation rule* to be invoked is (49): that requires us to have access to the strictly ordered set of dominance relations corresponding to the graph and establish, based on the position of the relevant NP in that set, what its grammatical function is. Given *class*’ and *student*’ as shorthands for the translations of *which class* and *which student* respectively, we have:

115) ρ = ⟨(prefer, *student*’); (prefer, *class*’)⟩

Now, it is clear that because we are drawing a map, there are inherent limitations to this proposal. For instance, the experimental preferences in Wh-island extraction found by Villata et al. (2016) fall outside the scope of this theory: our graphs and trails are not designed to encode gradience or, as a matter of fact, any aspect of *performance*. The focus here is to try and characterize possible dependencies, and separate them from impossible dependencies. The ‘grammar’, here, is *not* the ‘parser’: the ‘grammar’, in the limited scope of this theory, is *a map of connections*. This interacts with previous work of ours in the distinction between two different kinds of locality conditions:

- A. Islands that emerge from aspects of the computation: these, in turn, fall into two categories
 - I. The local ‘oscillation’ between levels in the Chomsky Hierarchy that we have described in Krivochen (2015a, 2016, 2017a); Krivochen & Schmerling (2016); Saddy (2017); Bravo et al. (2015), among other works: if the computational system can shift between levels locally, the change in computational dependencies defines opaque domains (see also the Multiple Spell-Out model of Uriagereka, 2002)
 - II. Lexically-induced islands (e.g., factive islands, ‘inner islands’, among others; see Ross, 1984, Rooryck, 1992)
- B. Islands that emerge from aspects of the *embodiment* of the computation (attention, memory, garden paths...)

The proposal presented here, as it simply makes all relations between nodes in a representation explicit, does not really *account* for locality effects, but it seems to me to do good job at *characterizing* them and the conditions under which cycles of the type A(I) arise (for instance, via *adjunction/substitution* of a sub-graph, as above). Needless to say, an account of A(II), and B. (which we have grossly oversimplified) is necessary for a psychologically plausible theory of grammar.

10. On MIGs and prizes (and embedding, and complements)

A theory that maximizes connectivity and allows for transitive loops can provide new insights in the analysis of MIG-sentences (a.k.a Bach-Peters paradoxes; see Bach, 1970; Karttunen, 1971; McCawley, 1967):

- 116) The man_i who shows he_i deserves it_j will get the prize_j he_i desires
- 117) Every pilot_i who shot at it_j hit the MIG_j that chased him_i

There are some interesting properties that we want to call the reader’s attention to. First and foremost, we need to note that an adequate structural description for Bach-Peters paradoxes should capture the fact that there are dependencies of varying computational complexity (in terms of the Chomsky Hierarchy of formal grammars and languages). Specifically, we proposed in past works (Krivochen, 2016, 2017; Krivochen & Saddy, 2017) that there are several *local* structural layers here, each displaying varying levels of computational complexity within the Chomsky Hierarchy, and that the assignment of a *strongly adequate structural description*³³ to these sentences *requires* the system

³³ We understand this notion in the sense of Joshi (1985: 208):

*A grammar G is weakly adequate for a string language L if L(G) = L. G is **strongly adequate** for L if L(G) = L and for each w [a string in L] in L, G assigns an ‘appropriate’ structural description to w* (Joshi, 1985: 208)

In this context, the question has been raised to us whether *strong adequacy* is equivalent to *strong generative capacity* (Chomsky, 1965: 60). The difference is crucial: the strong generative capacity of a grammar is the set

to be sensitive to local changes in computational complexity. Let us illustrate this point using (116) as our example:

- a) The – man; the – prize; he – deserves – it \Longrightarrow **Finite-state**
- b) [The man *S'* [will get [the prize *S'*]]] \Longrightarrow **Context-Free**
- c) [*the man* [who shows [he deserves **it**]]] will get [**the prize** [*he* desires *e*]] \Longrightarrow **(Mildly) Context-Sensitive**
-

Note that the computational differences arise *within local derivational units* and pertain to relations of co-reference within or across treelets. A strongly adequate grammar, needs to be flexible enough to accommodate for oscillatory dependencies in assigning structural descriptions to sentences, going *up and down the Chomsky Hierarchy* in local domains. In this perspective, we can define *cycles* as *emergent properties* from a computational system that does not commit to a ‘one-size-fits-all’ theory of phrase structure, but to one that models language as a non-linear, dynamical system (Saddy & Uriagereka, 2004; Saddy, 2017, and related work). In such a system, computational uniformity would in fact have to be *stipulated*. In (a) we have uniformly binary branching minimal treelets, following Greibach (1965: 44) and Uriagereka (2012: 53), a FS grammar can capture all relevant dependencies within these substrings. (b) features phrasal ‘constituents’ related within structures containing placeholders for root nodes; center embedding pushes us up to CFG. The appearance of crossing dependencies in (c) forces us up at least ‘half’ a level, to ‘mild’ CSG (Joshi, 1985 and much related work).

One of the reasons why Bach-Peters paradoxes bear a particular syntactic and semantic interest is that, under a transformational view of how pronouns come to be in Surface Structure, they must be assigned Deep Structures of infinite complexity (Bach, 1970). While it is possible to save the transformation by restricting its applicability to specific contexts (e.g., pronouns in root clauses need not be transformationally derived, as argued by Lakoff and Postal, among others), that modification increases the computational power of the grammar, and creates a potential overgeneration problem. One of the advantages of the framework we have sketched here is that Pronominalization need not make reference to *two* nodes (as in the Lees-Klima version), but only *one*, which is visited (at least) *twice*. In the theory proposed here, the lexical items ‘the man’, ‘who’, and ‘he’ in (116) are *the same node*, because that node in the graph does not correspond to a constant in the language, not even to a variable in that language, but to *a translation of the indexed basic expression in that node into intensional logic* (in the sense of Montague, 1973; Partee, 1974). As mentioned above, while it is possible to recast the Pronoun Rule³⁴ of Lees & Klima (1964: 23) in terms of trails, where a node *x*

of structural descriptions it generates; the definition given by Chomsky says nothing about that set being internally heterogeneous. The grammar that generates structural descriptions in Mainstream Generative Grammar is computationally uniform, and thus the set of structural descriptions is computationally uniform as well (falling within Type 1 languages). In contrast, Joshi’s requirement for *strong adequacy* for structural descriptions can incorporate aspects of mixed computation if the grammar is made sensitive to semantic dependencies between syntactic objects (which, in Chomsky’s architecture for a linguistic theory, cannot be formulated given the fact that the syntactic component is conceived of as an autonomous component). Chomsky’s *strong generative capacity* makes sure there is a set of structural descriptions, Joshi’s *strong adequacy* makes sure that set is (under present assumptions) minimally adequate (not assigning any more structure than strictly necessary), but not procrustean. See Krivochen (2015a; 2016); Krivochen & Saddy (2017) for discussion.

³⁴ *Pronoun Rule: X-Nom-Y-Nom’-Z → X-Nom-Y-Nom’+Pron-Z where Nom = Nom’, and where Nom is in a matrix sentence while Nom’ is in a constituent sentence embedded within that matrix sentence* (Lees & Klima, 1964: 23).

pronominalizes in context C if there is a trail t in which $\rho^*(x, x)$ holds –for example- we will not be concerned with the morpho-phonological exponents that the nodes in our graphs are assigned.

Let us now give the cyclic analysis and ρ -set for (116): recall that each *cycle* contains a single categorematic / lexical predicate, plus its temporal / aspectual modifiers and nominal arguments:

- 118) Cycle 1: [*man*' [Cycle 2] will get *prize*' [Cycle 3]]
 Cycle 2: [*man*' show [Cycle 4]]
 Cycle 3: [*man*' desires *prize*']
 Cycle 4: [*man*' deserves *prize*']

Putting it all together, then, we get:

- 119) [*man*' [*man*' show [*man*' deserves *prize*']] will get *prize*' [*man*' desires *prize*']]

Unifying all cycles, and identifying common nodes, we get the following ρ -set

- 120) $\rho = \langle\langle$ (show, *man*') ; (show, deserve); (deserve, *man*') ; (deserve, *prize*') ; (get, *man*') ; get, *prize*') ; (get, desire); (desire, *man*') ; (desire, *prize*') $\rangle\rangle$

Some clarifications are due. Perhaps the most important of them pertains to the fact that the object position (arc 2) of *show* is identified with the root of the graph corresponding to the subordinate clause *he deserves it*, which is *deserve* by virtue of not being dominated by any other node within that sub-graph. However, that is not the only possible analysis. Arguably (and this is completely orthogonal to the transformational/non-transformational debate), there is a COMP position in that clause, which we assumed in our discussion of relative clauses; if that is indeed the case, then COMP is the root. Thus, locally, instead of

- 121) $\langle\langle$ (show, *man*') ; (show, deserve) $\rangle\rangle$

We should have (122):

- 122) $\langle\langle$ (show, *man*') ; (show, COMP) $\rangle\rangle$

And the sub-graph for the subordinate clause would then be:

- 123) $\langle\langle$ (COMP, deserve); (deserve, *man*') ; (deserve, *prize*') $\rangle\rangle$

It is, however, not clear what the status of COMP is, apart from its descriptive and notational convenience. Note that in interrogatives we do not need to introduce a further 'projection' to host Wh-phrases (see the discussion in **Section 9**); it is not required semantically or syntactically (where 'syntactically' means 'necessary for the well-formedness of the graph'). Also, when dealing with relative clauses we have only looked at Wh- and *that* relatives, not contact relatives or abridged relatives; and we have not analysed finite clausal complements at all.

In this context, there are some things to bear in mind. First, it really should not matter whether we have a node morpho-phonologically realized or not, because we are only mapping syntactic connectivity (which is, in our view, semantic in nature; this view owes much to Generative Semantics). Second, observing that we do not need an extra nonterminal CP to deal with Wh-movement (just like Dependency Grammar doesn't; see e.g., Groß & Osborne, 2009; Mel'čuk, 2003) does *not* mean that no clausal domain has a COMP position: *a priori* structural uniformity is a feature of Minimalism, not of the framework we sketch here. Thus, in order to claim that there is a node n in graph G , G being the structural description assigned by the grammar to a sentence S we need to have *independent* evidence for n in G which comes exclusively from S . Think of this as a radical version of

LFG's *Economy of Expression*, as pointed out above. Having *root* phenomena does not equal having COMP positions, for roots are properties of graphs, and COMP is the identity of a specific node; the former pertains to the format of locally cyclic, directed, rooted graphs, whereas the latter is a substantive hypothesis pertaining to what a specific node is to be called. By definition, our graphs are indeed *rooted*, that is, they contain *at least* one node which is not dominated by any other node. But the inclusion of a COMP node in specific subordinate positions requires extra –empirical– justification. Just to be clear, the contexts we are considering now are:

- 124) a. Contact and abridged relative clauses (CRC)
- b. Contact noun clauses (CNC)

Both these constructions are present in (116), as indicated in the annotated (116'):

116') The man who shows [_{CNC} he deserves it] will get the prize [_{CRC} he desires]

What we need to ask at this point, is whether the relevant construction requires any of its parts to be dominated by an element outside the sub-graph corresponding to said construction; as an example, do we want the structural description for *shows he deserves it* to have $\rho(\text{show, deserve})$, as in ECM structures (see the analysis of (57) above)? Not really: there is no dependency between *show* and *he*, whereas there *is* a dependency between an ECM verb and the subject of the subordinate clause (which is arguably manifested as Accusative case marking in the embedded subject). There seems to be a root for the sub-graph corresponding to the structural description of the CNC which is *not* the verb in that CNC; we will assume that there is indeed a COMP *that* which can be left unpronounced. Thus, (122) is the local dominance set that we will consider for the CNC in (116).

At this point, we need to ask a further question: does the preceding argument mean that we need a covert COMP for CRC as well? Not really: the relative clause has an antecedent (*the prize*) and its subject *he* is obtained via Pronominalization; a single node is walked on more than once. In the case of the CRC there seems to be no need to resort to a distinct COMP position to serve as the root.

Let us consider some aspects of licensing in Bach-Peters sentences more carefully. When we consider a modified version of (117), in (125) below, we may ask how it is possible to get all relevant connections to hold:

125) [Every MIG_i [that chased a pilot_j [who shot (at) it_i]]] was hit by him_j

Let us proceed carefully. First, note that the referent for *him* in (125) is in the object position within a restrictive relative clause: *that chased a pilot who shot at it*. The reading notated in (125) with indexes should not be under a Subjacency-inspired view of locality: to begin with, *him* appears within a *by-phrase* adjunct; then, there are at least two bounding nodes on top of *a pilot*: S' (*that...*) and NP (*Every MIG...*). Moreover, all trails are *crossing*, there's literally *no embedding dependency* between indexes: that is, we do not have *i...j...j...i*, but rather *i...j...i...j*.

Furthermore, *a pilot* must be accessible to *him* at the point of establishing a dependency, despite the presence of a potential governor *it* (corresponding to *every MIG*), in flagrant violation of Relativized Minimality (Rizzi, 1990). How can we derive the correct reading? To begin with, the fact that we do not require for each node to be visited only once allows for *Every MIG* and *it* on the one hand, and *a pilot*, *who*, and *him* on the other to be superficial morpho-phonological realizations of just two nodes: *MIG'* and *pilot'*. But we still need to be able to create trails across sub-graphs, which then lead to the identification of common components. A crucial point here is that *none* of the sub-graphs in (116)-(117) is *self-contained*, because all those contain nodes that are dominated directly by nodes in other

graphs as well. We repeat the cyclic structure of (116) as (126 a) and that of (117) is represented as (126 b):

- 126) a. Cycle 1: [*man*' [Cycle 2] will get *prize*' [Cycle 3]]
 Cycle 2: [*man*' show [Cycle 4]]
 Cycle 3: [*man*' desires *prize*']
 Cycle 4: [*man*' deserves *prize*']
- b. Cycle 1: [*pilot*' [Cycle 2] hit *MIG*' [Cycle 3]]
 Cycle 2: [*pilot*' shot *MIG*']
 Cycle 3: [*MIG*' chased *pilot*']

Let us focus on (126 a) first. We see that *man*' and *prize*' appear in all four cycles, linking them. This means that none of them is self-contained in the technical sense, because some of its internal vertices are also dependents of other sub-graphs (i.e., are dominated by nodes in other sub-graphs), as can be seen in the ρ -set for (116) given in (120), above. The same happens in (126 b): *pilot*' and *MIG*' are dominated by nodes in all cycles (i.e., in all sub-graphs). The corresponding ρ -set for (117) is (127):

- 127) $\rho_1 = \langle (\text{hit}, \textit{pilot}'); (\text{hit}, \textit{MIG}') \rangle$
 $\rho_2 = \langle (\text{shot}, \textit{pilot}'); (\text{shot}, \textit{MIG}') \rangle$
 $\rho_3 = \langle (\text{chase}, \textit{MIG}'); (\text{chase}, \textit{pilot}') \rangle$

The combination of sub-graphs (or 'cycles') allows for all conditions required for licensing to hold, in both the structural descriptions of (116) and (117). We repeat the definition of licensing for convenience:

128) *Licensing (final formulation)*

a $\in A$ may license *b* $\in B$ iff

- i. $(a, b) \in \rho^*$, and [alternatively, there is a trail *T* that contains *a* and *b*]
- ii. *B* is not adjoined to the root of *A*, and
- iii. ***Neither A nor B are self-contained***

Where *A* and *B* are sub-graphs and *a* and *b* are nodes

And where *self-containment* is defined as follows (also repeated from above):

A graph *G* is a self-contained syntactic object iff $\nexists(x), x \in G$ such that

- i. $\rho^*(y, x)$ holds for $y \in G'$ and $G' \neq G$, **and**
- ii. *x* is an argument in *G*

Thus, whereas a purely configurational approach to cyclicity (e.g., Subjacency / Superiority-based approaches) may wrongly predict that the relevant referential dependencies are not possible; however, all trails are well-formed, and the graph-approach can provide us with an exhaustive description of relations between nodes.

In connection to this last point, the description of conditions over possible dependencies between nodes within and across graphs, the following section will focus on some aspects of the structure of coordination, and the formulation of admissibility conditions for relations across conjuncts. We will argue that a single structural template for coordinated structures is descriptively inadequate, and

attempt to capture the advantages of a computationally mixed approach to coordination (particularly, the case made in Krivochen & Schmerling, 2016) in terms of possible relations between elements belonging to separate *cycles*.

11. Two kinds of ‘and’ and the Coordinate Structure Constraint

So far we have been analyzing either simple sentences or complex sentences displaying different kinds of embedding; we have not, however, dealt with conjoining yet. We will do that now. In Krivochen & Schmerling (2016a, b) and Krivochen (2015a, 2016c) we argued that, syntactically, true coordination³⁵ is not a unified phenomenon. Rather, structural descriptions assigned to strings of the general form [X and Y] need to take into consideration both syntactic and semantic features, which cluster coordinate structures in two classes. We will refer to these classes as *et-coordination* vs. *que-coordination* (adopting Latin terms due to their descriptive resemblance, as we will see shortly). The empirical specifics of this distinction are currently under research (see Krivochen and Schmerling, 2016a for extensive discussion and examples), but we can summarize the main characteristics of each:

Que-coordination:

- Strictly paratactic dependencies between terms
- The arguments are interpreted as a single entity (*perfective*); thus no probing into a conjunct is allowed
- Triggers singular agreement when it is NPs being coordinated due to internal opacity

Et-coordination:

- Hypotactic dependencies between terms
- Each argument is a separate entity (*inflective*), allowing for inner probing into a conjunct
- Triggers plural agreement when it is NPs being coordinated due to accessibility

Let us give an example (the reader can find many more in Krivochen & Schmerling, 2016a)

³⁵ ‘True’ coordination is to be distinguished from what Krivochen & Schmerling (2016b) call ‘mirage coordination’. These are sequences of the kind [V...and VP], involving finite Vs -most frequently two-, the last of which is a fully fledged VP. We argued that these structures

appear to enter into verb coordinations—but we will argue that on close examination these can be seen to involve something other than coordination. The fact that the structures we will consider appear at first to be coordinations but in fact are not gives us our name for them: mirage coordinations (Krivochen & Schmerling, 2016b: 1)

Mirage coordination examples are, by virtue of not being real coordinations, exempt from the usual constraints on coordinate structures, including the CSC, and display strong restrictions (e.g., only two Vs can appear in mirage coordination, as opposed to the initially unbounded nature of true coordination; moreover, only a very limited number of verbs can appear as the first ‘mirage conjunct’ in these structures). Examples of mirage coordination include the English *go and*, *try and*, *up and*, *take and* in examples like:

- i) She’s *gone and ruined* her dress now. (Ross’ 1967 (4.107a))
- ii) She’s *up(ped) and ruined* her dress (note that there is no RNR interpretation available)
- iii) She *took and replaced* the hose (RNR interpretation irrelevant)

Mirage coordination is not exclusive to English, of course. Spanish features a number of mirage coordinated constructions, including *agarrar y*, *coger y*, and *ir y*.

- 129) a. The sudden rise and (the) equally sudden fall of the stock market have economists worried.
 b. The sudden rise and (the) equally sudden fall of the stock market has economists worried.
- 130) a. La abrupta subida y la igualmente abrupta bajada de la bolsa preocupan al Gobierno.
the sudden rise and the equally sudden fall of the stock market worry_{PL} the Government
 ‘The sudden rise and the equally sudden fall of the stock market worry the Government.’
- b. La abrupta subida e igualmente abrupta bajada de la bolsa preocupa al Gobierno.
the sudden rise and equally sudden fall of the stock market worries_{SG} the Government
 ‘The sudden rise and equally sudden fall of the stock market worries the Government.’

The English example (129 b) and the Spanish example (130 b) are both instances of *que-coordination*. In each of these examples the conjoined NP’s are understood as having a single referent, albeit a complex one: stock market fluctuation. Sentences (129 a) and (130 a), exhibiting *et-coordination*, are not semantically equivalent to their *que-coordinated* counterparts. Consider the *et-coordinated* example (129 a): this sentence could describe a situation where particular economists were worried about the sudden rise of the stock market but not about its sudden fall, whereas others were concerned about its fall and not its rise. This interpretation is not possible for the *que-coordinated* example (129 b), where economists must be worried about the combination.

In this context, we argued that the CSC, as formulated in Ross (1967: 428)...

In a coordinate structure, no conjunct may be chopped, nor may any element contained in a conjunct be chopped out of that conjunct.

...applies only to *et-coordinated* structures. Summarizing much discussion, the reason is relatively clear: the CSC is formulated as a condition over chopping transformational rules³⁶, and these rules need to have access to whatever they reorder. If *que-coordination* yields an internally opaque unit, then no term of a *que-coordination* may be analyzed by a transformational rule. What follows, insofar as we will be concerned with the CSC, will refer exclusively to *et-coordinated* structures.

As a principle quantifying over rules, Ross’ formulation of the CSC has no place in a model-theoretic framework. However, its importance cannot be underestimated: it is possibly the most problem-free principle ever formulated in generative grammar (Postal, 1997: 52), and its applicability goes well beyond English (and even beyond Indo-European languages; e.g., Georgopoulos, 1985: 87-88), which is an increasingly rare feature in MGG constraints and principles.

We have two main challenges, then: first, to provide adequate structural descriptions for coordinate structures; and second, to capture the valid content of the CSC as a condition over admissible graphs.

³⁶ Where ‘chopping’ is defined as follows:

*If the structural index of a transformation has n terms, a_1, a_2, a_n , it is a **reordering** transformation if its structural change has any a_i as its k^{th} term, or if a_i is adjoined to its k^{th} term, where $i \neq k$. If a transformation reorders a_i , and its structural change substitutes the identity element or some a_k , $i \neq k$, for the i^{th} term of the structural index, the transformation is a **chopping** transformation. Other reordering transformations are called **copying** transformations. (Ross, 1967: 427. Emphasis ours)*

As for the first challenge, it is useful to take a look at what other graph-theoretic frameworks have proposed. APG, for instance, recognizes a class of relational Arc dubbed *Con*, a Structural, Non-nominal R-sign (Johnson & Postal, 1980: 198). Further conditions specify that the heads and tails of *Con* arcs must be labelled with the same major category and *Con* arc heads must bear the same category label (essentially, a condition on the identity of *coordinantes*). The relevant definitions, pertaining to heads and tails of *Con* arcs, are the following (Johnson & Postal, 1980: 209):

- 131) a. $\text{Coordinate}(a) \leftrightarrow (\exists A) (\text{Con arc}(A) \wedge \text{Tail}(a, A))$
 b. $\text{Conjunctive}(a) \leftrightarrow (\exists A) (\text{Con arc}(A) \wedge \text{Head}(a, A))$

For example, in *The fiend shot and knifed its victim* (example taken from Johnson & Postal, 1980: 222), the coordinated Vs are the heads of *Con* arcs. In and of itself, we think, the APG treatment of coordination does not capture the difference between *et-* and *que-coordination*, which is empirically motivated at its very core. Let's continue the survey.

Treatments of coordination within Dependency grammar are not quite homogeneous: Mel'čuk (1988: 26-28) argues that coordinated structures are headed (otherwise, there could be no dependency relation), based on the purported fact that '*In the majority of cases there is no reversibility in coordinated structures*' (Mel'čuk, 1988: 26); in these asymmetric structures, the left conjunct is the head, with the right conjunct *depending* on it. We think that the reversibility argument is only (partially) valid for a specific kind of coordinate structures, specifically, *et-coordinations*; furthermore, the lack of reversibility need not imply headedness, it only requires hypotaxis. Syntactically, there is no argument or reason to make all strings containing *and* belong to the same class; semantically and pragmatically, the heterogeneity of coordination only becomes more evident. Consider the following example, which we used in Krivochen (2017) as part of an argument in favor of mixed computation (the idea that the grammar oscillates up and down the Chomsky Hierarchy when assigning structural descriptions to sub-strings):

- 132) (talking about afternoon tea) John got the milk and Bill brought some cookies

Note that (132') is perfectly acceptable (since there is no particular order in which the purchase of milk and cookies should be presented; the events could very well take place simultaneously for all we care):

- 132') Bill got the cookies and John bought the milk

The reversibility of conjuncts, we argued in Krivochen (2017), points towards a paratactic structural description (essentially, with both S terms being states in a Markov chain). But let's go one step further, and consider (132) embedded in a bigger structure:

- 133) Bill got the cookies and John bought the milk, and we all had a wonderful afternoon tea

The terms of coordination are two once again, and the relation seems to be asymmetric: we had a wonderful afternoon tea *after* (and probably also *because*) John and Bill contributed with milk and cookies respectively. A successful syntactic treatment of coordination must be capable of capturing the distinction between what Schmerling (1975) refers to as *symmetric* and *asymmetric conjunction*.

A different view from Mel'čuk, but within DG, is expressed by Tesnière (1959: 80, ff.). Tesnière assigns coordinating conjunctions to a category *junctives*, which are *functional elements* (or *empty words*, 'mots vides'). Junctives are structurally between the terms they conjoin (called *nuclei*), but

remain outside these terms. In brief, coordinated structures are not only not headed structures, but they are not even dependency-based.

We argue, against MGG, that coordination is *never* an endocentric structure, there is no ‘head’ in coordinated structures; thus, there cannot be a dependency in DG terms, even though there is dominance. We are interested in capturing the contrast between *et-* and *que-coordination* in a way that maintains the core aspect of the analysis: both paratactic and hypotactic dependencies can be established in coordination. That is: to a string *X and Y*, where X and Y are variables over strings, there can correspond two structural descriptions, roughly along the lines of (134):

- 134) a. $X \rightarrow (\text{and}, Y)$
 b. $\text{and} \rightarrow (X, Y)$

(134 a) corresponds to the *et-coordinated* case, in which the root of the sub-graph X dominates the root of the sub-graph Y. There lies the essence of *hypotaxis* in the present view. (134 b) corresponds to *que-coordination*, in which case the roots of the conjuncts are both dominated by *and*, which is thus the root of the whole graph. In this latter case, both conjuncts’ roots are *sisters*, in the former case, the second conjunct’s root is within the ρ -domain of the first conjunct’s root.

Let us consider the ρ -sets of the two conjuncts in (136), an instance of *asymmetric et-coordination*:

- 135) John had a beer and fell asleep
 136) $\rho_1 = \langle (\text{have}, \text{John}'); (\text{have}, \text{beer}'); (\text{have}, \text{and}) \rangle$
 $\rho_2 = \langle (\text{and}, \text{fall}); (\text{fall}, \text{John}'); (\text{fall}, \text{asleep}); (\text{asleep}, \text{John}') \rangle$

The root of the first conjunct is the verb *have*, because there is no node that dominates it. In the asymmetric case, the first conjunct is strictly ordered with respect to the second such that the first conjunct always precedes the second conjunct. Moreover, the second conjunct is completely within the ρ -domain of the root of the first conjunct.

The *que-coordinated* case is different, as we anticipated. Let us consider first the case in which we coordinate full clauses both (137 a) and (137 b) are well-formed:

- 137) a. Bill got the cookies and John bought the milk
 b. John bought the milk and Bill got the cookies

What we would like to propose with respect to *que-coordination* is that both conjuncts share the same root, which is the coordinating conjunction. It is crucial to note that this does *not* mean that ‘and’ is the label, or head, of the coordinated structure, *because our structures are neither labelled nor endocentric*. The same happens when we *que-coordinate* simple Ns, for instance:

- 138) a. John had red beans and rice for dinner
 b. The rise and fall of the stock market has economists worried

What we have in (137) and (138) is, we argue, something along the lines of (139 a) and (139 b), respectively:

- 139) a. $\rho = \langle (\text{and}, \text{get}); (\text{and}, \text{buy}); (\text{get}, \text{Bill}'); (\text{get}, \text{cookies}'); (\text{buy}, \text{John}'); (\text{buy}, \text{milk}') \rangle$
 b. $\rho = \langle (\text{have}, \text{John}'); (\text{have}, \text{and}); (\text{and}, \text{red beans}'); (\text{and}, \text{rice}') \rangle$

Note that *que-coordination* does not overlap with *symmetric coordination*: the following is an example of *symmetric et-coordination*:

140) John drinks beer and whisky

In Krivochen & Schmerling (2016) we argued that the possibility of having structures of the kind *both X and Y* or *either X or Y* is a test for *et-coordination*, since *both* requires access to the terms of the coordination. If *que-coordination* yields internally opaque units, then it is impossible to access the terms separately. Note, then, that:

- 141) a. ?John had both red beans and rice for dinner (? In the interpretation in which ‘red beans and rice’ is a single dish)
b. ?John had either red beans or rice for dinner (same as above)
c. John drinks both beer and whisky
d. John drinks either beer or whisky

Regarding the structural properties of *que-coordination* and its interaction with the CSC, Krivochen & Schmerling (2016); Krivochen (2018) defend the idea that the internal opacity of *que-coordination* makes it impossible for it to be targeted by extraction transformations. Moreover, if the terms are indeed interpreted as a single, internally unanalyzable unit, there is no rescuing the structure with ATB (which, again, would require access to the internal structure of each conjunct). As such, the CSC applies to *et-coordinated* structures, not *que-coordinates* ones. How can we formulate a relevant admissibility condition that captures the descriptive power of the CSC? The formulation in Krivochen (2017a) could prove useful in this respect:

- 142) In an *et*-coordinated structure of the form $[K...SO_i \text{ [and } [L...SO_j]]]$, where
- $SO_{(i,j)}$ are in parallel structures (where ‘parallelism’ is defined over semantic-syntactic construal, see McCawley, 1968 and much related work),
 - K and L are terms, and
 - $L \subset K$;
- a mapped phrase marker $[SO_{(ij)}...[K...SO_i \text{ [and } [L...SO_j]]]$ is legitimate *iff* $i = j$

This condition straightforwardly yields the well-formedness of ATB constructions (Williams, 1978), but it also predicts that canonical examples of the applicability of the CSC are indeed outside the phase space for a specific language. Note that this formulation makes *no reference* to situations of the following kind,

143) $[SO_i...[K...SO_i \text{ [and } [L...]]]$

in which the term $[L...]$, of arbitrary complexity, is left untouched. Let us give an example of such a construction, taken from Goldsmith (1985) –traces and indexes added for expository purposes only–

144) $[[[K \text{ [How many lakes]}]_i \text{ can } [K \text{ we destroy } t_i]] \text{ and } [L \text{ not arouse public antipathy}]]?$

It is crucial to point out that the nature of the term L is not relevant: Schmerling’s (1983: 17-18) argument that [and not] is actually a constituent (a terminal, in the terms we have been managing here, the composition of *et*-coordination + negation), and thus the possibility that [not arouse...] is not itself a constituent, makes no difference for the purposes of (194). The key factor here is that there is *no variable* (in the sense of Ross, 1967) within term L. Extraction only targets K, and the extended phrase marker after chopping is still K (calling it K’ would make no difference: the dependency is still established *within* a single cycle).

Much research pending, we have tried to capture some of the aspects of *et*- and *que-coordination* with the tools available to us in the theory of syntax sketched in the present paper. Even if the specifics of the analysis turn out to be only partially on the right track, we think that there are aspects of the study

of coordinated structures on which a graph-theoretic approach can shed light, combining insights from PSGs and DGs with the model-theoretic machinery.

12. A small collection of ‘transformations’

We will now attempt to classify some well-known transformations from the *aetas aurea* of generative grammar in terms of whether they change existing grammatical relations or they just create new relations on top of what was already there. The following examples must be understood as referring *only to English*, and encoding traditional insights on what reordering transformations do (*qua* descriptive devices). We are essentially assuming that, if we are providing a map of all the relations in a snapshot of a dynamical system, under a transformational approach what transformations do is make the system evolve and change, thus, we can have a pre- and post-transformational snapshots. There is *no* aspiration of aprioristic universality. All in all, what is important here is to see how many ‘transformations’ actually change grammatical relations, which directly impacts on the descriptive adequacy of our theory. In this sense, Epstein et al. (1998: 3) say that

From Syntactic Structures through the Extended Standard Theory framework, transformations were both language-specific and construction-specific: the linguistic expressions of a given language were construed as a set of structures generable by the phrase structure and transformational rules particular to that language. Transformational rules thus played a crucial role in the description of particular languages.

They go on to make the contrast with P&P-based theories (GB/MP), which focus on universal restrictive principles. We think that the descriptive power of the REST, which was due to the close attention paid to constructions and the explicit formulation of specific transformational rules, is something to *recover* rather than to abandon in favour of putative ‘universal’ principles and rules (e.g., *Affect-a*) which greatly obscure grammatical description (and do not provide *explanations* in any meaningful sense; see Postal, 2004: §§ 9, 12, 13). What follows is a classification of well-known transformations in terms of whether the processes create new relations while maintaining existing relations or whether no new relations are created, only linear order; McCawley’s (1982) RCT vs. RPT. This classification –by no means exhaustive!–, which summarizes processes that we have referred to in previous sections, must be understood as referring *only to English*, and transformations –once again– are to be understood in this context as *strictly descriptive proxies*.

I. **Relation-preserving transformations:**

I.1 Create *new* relations (leaving old relations intact):

- a) Interrogative formation (**Section 9**)
- b) Raising-to-Subject (**Section 5**)
- c) Raising-to-Object (**Section 5**)
- d) Topicalization (**Section 5**)
- e) Focalization (**Section 5**)
- f) NPI fronting (**Section 5**)

I.2 Maintain all existing relations *only* changing linear order:

- a) RNR (see McCawley, 1988; also (35) above)

- b) Wrap (as in Bach, 1979; **Section 3**)
- c) Dative Shift (*contra* Dowty, 1979)
- d) Right dislocation, including Heavy NP Shift (but see below)
- e) Location fronting
- f) Though- / as- preposing (see Ross, 2012)
- g) Parenthetical insertion (see McCawley, 1982 and **Sections 7 and 8**)
- h) Clitic climbing (**Section 6**)
- i) Pseudocleft formation (*as per* Ross, 2011)

II. Relation-changing transformations:

- a) Passive (see Williams, 1982 and Bresnan, 1983; Dowty, 1978; Postal, 1986 respectively for for- and-against views on whether Passive is indeed a transformation. Müller, 2000 provides arguments against an Object-to-Subject raising analysis of the German passive, but it is not clear whether those arguments can be extrapolated directly to English)

We have classified Passive as a relation-changing transformation (in the descriptive sense in which we have been thinking about transformations all throughout) because it requires not only the advancement of a 2 to 1 (1-advancement), but also that the relevant NP is *not a 2 anymore*. Note: this does not mean that the NP under consideration is no longer an argument of the V, just that its grammatical function changes: this is fully compatible with RG's view that passivization (in English) is an entirely relational process (Postal, 1986). Again limiting our discussion to English, there cannot be an Accusative pronoun as the subject of a passive sentence:

145) *Him was killed by Bill

A Movement-based PSG (ST, GB, MP) derives a passive via movement, such that an NP is base-generated in the object position (sister to V) at Deep Structure (or whatever level of representation or intermediate derivational step is there before Move- α) and then moved to Spec-Infl / Spec-T / Spec-AgrS (depending on the framework of choice) leaving a trace or copy behind. We have briefly commented on some inadequacies of both approaches to displacement here and in past works, and it is also a thoroughly researched topic in non-transformational analysis. What we want to emphasize here is that the framework we have developed here applies straightforwardly to Passives, with no additional stipulations needed. Consider the following pair:

- 146) a. Bill killed John
 b. John was killed

The ρ -sets for (146 a) and (146 b) are given as (146 a') and (146 b') below:

- 146') a. $\rho = \langle (\text{kill}, \text{Bill}'); (\text{kill}, \text{John}') \rangle$
 b. $\rho = \langle (\text{be}, \text{John}'); (\text{be}, \text{killed}); (\text{killed}, \text{John}') \rangle$

We assume two fundamental properties of passive *be* in English: (i) has no a-structure; and (ii) it forms a derived expression with the subject of the lexical verb (Schmerling, 1983b: 26-27). Only the first of these properties holds for Spanish *ser* (see also Bravo et al., 2015; García Fernández et al., 2017 for discussion about the functional nature of Spanish passive *ser*), as pointed out above, Spanish does not have English-style VP ellipsis:

- 147) a. Mary wasn't killed, John was $\forall P$
 b. *María no fue asesinada, Juan fue $\forall P$

We propose that it is the existence of a relation (be, *John*') licenses examples like (131 a), and presumably that relation does not hold in Spanish: *ser* is much more of an inflectional element than *be* in English, which would translate into the claim that passive *ser* actually configures a two-word basic expression with the participle: the ρ -set for something like *María fue asesinada* would be simply $\rho = \langle \langle \text{ser asesinada, María} \rangle \rangle$, not $\rho = \langle \langle \text{ser, María} \rangle; \langle \text{ser, asesinada} \rangle; \langle \text{asesinada, María} \rangle \rangle$. There is no relation between *ser* and *María*, which readily blocks the possibility of having a passive treelet *María fue* in VP ellipsis contexts (cf. (131 b)): passive *ser* is not a basic expression in Spanish.

This proposal finds some additional support in the analysis of examples like the following:

- 148) *María no fue asesinada, pero Juan lo fue*
M. NEG was murdered, but J. CL was

In (148), the pro-form *lo* stands for the VP *asesinado*, deleted under sloppy identity (because the agreement features in the participle must change, since *María* is a feminine N and *Juan* is a masculine N). It is also a phonological clitic, in the sense that it is not a well formed phonological word on its own. Now, assume that *ser asesinado* is not a two-word basic expression, but rather a derived expression: for instance, let *ayudado* be assigned to the category FC/NP (i.e., the category of expressions that must concatenate with an NP to form a finite clause). Then, passive *ser* (should it be a basic expression of its own), could be (FC/NP)/(FC/NP) (see Bach, 1983: 111 for such an analysis). But that would predict that a rule of functional application (or a 'transformation') can affect the participle (or the pro-form) independently, for instance, *clitic climbing*. A model of Spanish passives in which *ser* does not form a basic expression with the participle it selects could generate the ungrammatical (149):

- 149) **María no fue asesinada, pero Juan lo tuvo que ser*
M. NEG was murdered, but J. CL had to be

If, on the other hand, there is a basic expression (of category FC/NP, in our example, but not necessarily) *ser asesinado*, (149) is adequately filtered: the pro-form *lo* cannot be reordered on its own, and of course *ser+lo* cannot be targeted by *clitic climbing*.

By adopting the CG distinction between basic and derived expressions, plus the Schmerlingian idea that basic expressions can be multi-word, and allowing nodes in our graphs to correspond to basic expressions rather than lexical items, we can capture an empirical difference which cannot be straightforwardly accounted for (or even described) in structurally uniform, *a priori*-based PSGs.

The rather surprising (well, that depends on the reader) provisional conclusion that we tried to illustrate in this brief section is that most 'transformations' either create new relations *while preserving existing ones* or only change linear order, thus not impacting on node connectivity at all. We think that pursuing this path could lead to a radical simplification of the apparatus required in the study of the syntactic phenomena in natural languages, attending to their specific connectivity patterns.

13. Some open problems and questions

The graph-theoretic approach to grammatical description that we have presented in this work is still in its infancy. Thus, it is only to be expected that there are many problems and questions still to be addressed in order to make this theory a competitive one. In this section, we present some of those as

a roadmap for future research, together with some provisional answers and mostly speculative explorations.

13.1 A brief note on Left and Right extractions

Interrogative formation and right-dislocation are different kinds of ‘transformations’ in the framework presented in this paper: their structural descriptions are *not* isomorphic. That might shed some light on why leftwards movement is unbounded (creates new relations, cumulatively) whereas rightwards movement –extraposition, Heavy NP shift- is heavily constrained (only changing linear order, but not constituency). Now, what is not clear is how to encode the specific filters (e.g., Right Roof Constraint) as graph admissibility conditions. A possibility is that those are not constraints at all, if dependencies are cycle-bounded in self-contained objects *and* if cycles are single-rooted sub-trees. If right-raised elements sister-adjoin to the root, as opposed as dominating it, the distinction between unbounded left-raising (i.e., ‘successive cyclic movement’ in MGG, or percolation of SLASH features in HPSG) and clause-bound rightwards raising can be hypothesized to derive not necessarily from the node they establish a relation with (in both cases it can be the root), but the kind of relation that is established: ρ vs. τ .

We will now illustrate some of these points. Consider now the distinction between *et-coordination* and *que-coordination* which we introduced in **Section 11** here, and in Krivochen (2015a, 2016, 2017a) and Krivochen & Schmerling (2016), according to which coordinated structures can be hypotactic and phrase-structural (*et-coordination*) or paratactic and finite-state (*que-coordination*). RNR arises in *et-coordinated* structures (symmetric or asymmetric), which are linked by virtue of having (in the sense of ‘not self-contained’). Thus, we are dealing with the usual state of affairs (‘usual’ from p. 1 onwards):

- 150) Bill bought, and John washed, those china dishes
Cycle 1: [Bill bought *dish*’]
Cycle 2: [John washed *dish*’]
 $\rho = \langle (\text{buy}, \text{Bill}'), (\text{buy}, \text{dish}'), (\text{wash}, \text{John}') (\text{wash}, \text{dish}') \rangle$

In this sense, as long as we keep *et-coordinating*, we can RNraise without boundaries: *et-coordination* is indeed recursive and monotonically so. Thus, the availability of cyclic RNR is only to be expected. For example:

150’) Bill bought, John washed, Mary dried, and Susy carefully put away, those china dishes

In the specific case in which RNraised NPs are *self-contained* units, island effects are straightforwardly captured (examples from Postal, 1997: 102. Traces and indexing are his):

- 151) a. Mary buys t_1 and Bill knows a man who sells t_1 [pictures of Elvis Presley] $_1$.
b.*Who $_1$ does Mary buy t_2 and Bill know a man who sells t_2 [pictures of t_1] $_2$?

An account of the islandhood effects in (151) which appeals to *self-containment* differs from constituency-based ones, like Wexler & Culicover’s (1980: 301) (also cited in Postal, 1997: 102)

[A] raised node [by RNR] always behaves, vis-à-vis all constraints on analyzability, just as it would if it were in its original underlying position. Hence, whereas it is apparently possible to apply RNR to a constituent of a relative clause, if we then try to analyze this raised node, we find that it acts as though it were still within the relative clause

This generalization is indeed compatible with the McCawley/Levine view, in which RNR is only an order-changing rule: if the constituency structure is not modified throughout the derivation, (135) is indeed a violation of the Complex NP Constraint.

But RC extraposition and Heavy NP shift do not work in the same way (examples from McCawley, 1998: 529. Annotations are ours).

- 152) a. That someone exists [who can beat you up to a pulp] is a foregone conclusion
 b. *That someone exists t_i is a foregone conclusion [who can beat you up to a pulp]_i (via Relative Clause Extraposition)
- 153) a. That John sent to his mother [the money that you wanted him to give us] is understandable
 b. *That John sent to his mother t_i is understandable [the money that you wanted him to give us]_i (via Heavy NP Shift)

What do the structural descriptions of these sentences look like, and how do they differ from RNR? Postal (1997: 97) assumes that RNR is basically ATB to the right. More specifically, he assumes, against McCawley, Ojeda, Levine, and others, that leftwards extractions (Wh-movement) and RNR are derived by the same means (which for him are ‘movement’ rules). We agree with Postal that L-extractions and RNR are derived *by the same means*, but those means are *multidominated nodes*, not relation-changing movement (that is, we disagree with Postal as to what those means are). It is not clear whether Postal’s arguments indeed show that L-extractions and RNR change constituency, particularly considering that in APG there are no constituents as such. Johnson & Postal (1980: 618), for instance, make use of an R-disloc arc.

In any case, what matters (descriptively) is that the unboundedness of RNR is captured by a strongly cyclic multidominance theory, as is the unboundedness of L-extractions. In contrast, we get heavily constrained displacement in extraposition and Heavy NP shift (HNPS), which are both cycle-bounded operations. RNR preserves local relations across cycles (each conjunct is a cycle by virtue of being a single-rooted sub-graph), whereas RCE and HNPS change order within a single cycle (and this is the core observation behind the Right Roof Constraint). In proof-theoretic terms, the problem seems to be one of derivational timing (pre- vs. post-cyclic rules), and it is not clear how to formulate those under the present assumptions: because there are no derivations in model-theoretic syntax, there is no ‘timing’ or ordering. A different approach must be pursued, which reformulates the relevant conditions in terms of admissibility. There is one possibility: because we have kept *root nodes* (as nodes that are not part of the ρ -domain of any other node; there being no reason to ban these to begin with), it is possible to specify that certain ‘transformations’ apply within the ρ -domain of a root. In other words, certain admissibility conditions make reference to a designated node and possible relations that can be established with said node in graphs that belong to the grammar. It is a good opportunity to re-evaluate Emonds’ (1970) typology of transformations, given the importance of root phenomena. In this sense, we could informally propose the following descriptive classification:

- 154) a. Processes targeting the closest root
 b. Processes targeting the matrix root
 c. Structure-preserving processes

Bear with me on this one. Recall that, in a proof-theoretic, transformation-enriched framework,

A phrase node X in a tree T can be moved, copied, or inserted into a new position in T, according to the structural change of a transformation whose structural description T satisfies, only if at least one of two conditions is satisfied: (i) In its new position in T, X is immediately

dominated by the highest S or by any S in turn immediately dominated by the highest S. (A transformation having such an effect is a root transformation.) (ii) The new position of X is a position in which a phrase structure rule, motivated independently of the transformation in question, can generate the category X. (A transformation having such an effect is a structure-preserving transformation) (Emonds, 1970: ii. Highlighted in the original)

Of course there is no movement, copy, or insertion, but a translation of Emonds' insights into the present framework seems possible. Let us take a first, informal stab at it:

A process P [read: a construction] may link a non-root node N in a single-rooted graph G to a node N' in a single-rooted graph G' in either of the following ways:

- I. *P creates an edge between N and N' where*
 - a. *$N \neq N'$ and N' is the root of G', **and***
 - b. *G' is the smallest cycle that properly contains G*
- II. *P creates an edge between N and N' where*
 - a. *$N \neq N'$ and N' is the root of G', **and***
 - b. *There is at least one single rooted graph G'' such that $G \subsetneq G'' \subsetneq G'$*
- III. *P creates an edge between N and N' where*
 - a. *$N = N'$*

Above, 'equal to' means 'have the same IL translation as'. It is easy to see that condition I refers to processes of the kind (154 a), in which *strong cyclicity* comes into play (each cycle counts); condition II refers to processes of the kind (154 b), in which only the *last cycle* is relevant; and condition III refers to processes of the kind (154 c), in which graphs are connected by means of identifying common components. Here, we are interested in I and II. Of these, only I instantiates the classic *cyclic principle* in its strongest version: a root phenomenon cannot 'jump' across a root. II, in contrast, pertains to a different class of transformations, *last-cyclic* or *higher-trigger cyclic* (Postal, 1972: 212). To say that Relative Clause Extraposition and Heavy NP Shift are transformations of the kind (154 a) amounts to establishing I as an admissibility condition over graphs corresponding to the structural descriptions of sentences displaying RCE and HNPS. On the other hand, we have Wh-movement and RNR as examples of constructions whose relevant admissibility condition is II; that gives them their unbounded character (without requiring *successive cyclicity* as an additional assumption). In a sense, I-III refine Emonds' original distinction, because it seems to be significant whether a process can only target the immediate root or whether it can link a node with a remote root.

It is an *ad hoc* stipulation that captures the behaviour of some English rightwards extractions, but what it follows from, if anything, is still unknown.

13.2 On deletion

Bach (1964: 70) lists the possible things that PSGs and transformations can do, in a very general (variable-free) format:

- 155) a. Delete: $a + b \rightarrow b$ (or $a \rightarrow \varepsilon$)
- b. Replace: $a \rightarrow b$
- c. Expand: $a \rightarrow b + c$
- d. Reduce: $a + b \rightarrow c$
- e. Add: $a \rightarrow a + b$
- f. Permute: $a + b \rightarrow b + a$

So far, we have dealt with additions, permutations, replacements, reductions, and expansions (in one way or another, sometimes denying these things exist at all). But we have said nothing about *deletion*. So, the question is: what do we do with *deletion* operations? First of all, let us give the reader an idea of the kinds of transformations we have in mind:

- 156) a. Gapping
- b. Equi NP deletion
- c. Ellipsis
- d. Sluicing
- e. Stripping
- f. Bare Argument Ellipsis
- ...

Here we need to distinguish two big kinds of deletion transformations:

- a) Deletion under (strict/sloppy) identity
- b) Deletion without identity

Only the former establishes relations between sub-graphs: *Equi* can be simply expressed as visiting an NP node twice (see **Section 5**) and Gapping, as visiting the same V node twice, etc. But deletion without identity seems to be a whole different matter. Consider, for instance, the cases in which there is no accessible syntactic antecedent as in BAE or instances of deletion without reconstruction: e.g., island repair via sluicing. In the latter case, specifically, if we do reconstruct we get an illicit structural description, as can be seen in (157 a), or cases in which the remnant of sluicing does not correspond to any plausible reconstructible syntactic object, as in (157 b-d; taken from Culicover & Jackendoff, 2005: 268):

- 157) a. They want to hire someone who speaks a Balkan language, but I don't know

{	<i>which</i> * <i>which they do</i>	}	(Wh-Island
	* <i>which Balkan language they want to hire someone who speaks t</i>)

 repair under sluicing; taken from Merchant, 2008: 138)
- b. Harriet is drinking scotch again, but I don't know what kind ~~Harriet is drinking [+scotch]~~ ~~again~~. (ungrammatical if reconstruction takes place; violation of the Left Branch Condition)
- c. Harriet drinks scotch that comes from a very special part of Scotland, but I don't know where ~~Harriet drinks scotch [that comes from t]~~. (ungrammatical if reconstruction takes place; Complex NP Constraint violation)
- d. Harriet either drinks scotch or smokes cigars, but I can't remember which ~~of drinks scotch~~ ~~or smokes cigars Harriet does~~. (ungrammatical if reconstruction takes place; word salad)

In (157 b) it seems that we need ρ (what kind, scotch) to hold, such that *scotch* can be the lexical restrictor for *what kind*, following the proposal sketched in **Section 7**, above. However, the same trick does not straightforwardly work in (157 c) unless the restrictor for *where* is *a very special part of Scotland*. There are some clues that this might not be completely wrong: note that the sluiced clause asks roughly *whereabouts in Scotland does the scotch that Harriet drinks come from?*; it is not an open locative variable that can be bound by, say, *Argentina*. But not all cases can be fixed in this manner, as the following examples illustrate:

- 158) a. *John bought an expensive car, and you won't imagine how ~~expensive~~ Mary did ~~buy a~~
ear!
- b. *Abby *did* say that a biography of Harpo is going to be published this year –guess which
Marx brother she *didn't* say that a biography of *t* is going to be published this year (from
Merchant, 2008: 143)
- c. *?John had three beers and two bottles of wine at the party, but I'm not sure how many
[beers? Bottles of wine?]; Mary did ~~have t_i at the party~~

It seems that we need a different account for these cases, which do not involve trails. A possibility as far as this kind of deletions is concerned is that we can simply *disconnect* the 'deleted' treelet from every other node of the graph. That is: we need not erase or eliminate the relevant object itself (a *vertex* or set thereof), but erase the *edges* connecting the relevant object with the rest of the graph, thus creating an *isolated vertex* or *sub-graph*. In that case, a node would not be ordered with respect to any of the other nodes, and by virtue of violating the condition on strict ordering for well-formed graphs, the resulting sentence is ungrammatical if the disconnected graph is somehow forced into the representation (as in the sub-strings marked with strikethrough in (158)).

Incidentally, the elimination of deletion operations constrains the strong generative capacity of the theory, such that we are safe from Peters and Ritchie's (1973) warning that a grammar with certain deletion rules could generate any recursively enumerable set (unlike some versions of Minimalism, in particular those that assign arbitrary features to heads which trigger deletion at PF, see Merchant, 2008: 133, ff. for such a view; also Berwick, 1984 for some discussion).

Conceptually, there is some appeal to this view. But, how can we implement 'deletion-as-disconnection'? We will briefly sketch a way, much research pending. Consider Tabor's (2015: 3) definition:

Lab_{x(0)} is said to process a symbol j at a point x iff $j \in \text{Lab}_{x(0)(x)}$ and the system moves from x to $f_j(x)$.

Where

The labeling, $\text{Lab}_{x(0)}$, of [an Iterated Function System] IFS driven by [a one-sided infinite string language] L is a function from points in X to the power set of Σ , such that $j \in \text{Lab}_{x(0)(x)}$ iff there is a finite initial substring, S, of some string in L such that $\text{IFS}_S(x(0)) = x$ and the string formed by adding j to the end of S is also an initial substring of some string in L. (Op. Cit.)

It might be possible to say that deletion of nodes in a phrase marker, in the sense that we suggested above, amounts to $\text{Lab}_{x(0)}$ not being able to walk through some vertex or set thereof (the sub-graph that has been 'disconnected'), because the system (in this case, we can imagine a routine like 'walk trail') cannot transition there.

But this proposal is not free from problems. For this approach to work, we need to assume that phrase structure graphs are in some sense generated by IFSs, that is, by a finite set of functions that map a complete metric space to itself. And this is not contemplated in the vanilla version of the model, because what we generate are diagrams of structural descriptions. Strings have a peripheral role, emerging as a kind of 1-dimensional 'track' of how a graph has been walked (and in such a view Ojeda's FLO would play a significant role). It may be possible to simply encode linear order in terms of linear precedence statements, the full set of which constitute the range of π (its domain being in turn a function of edges \times vertices). Linearization would thus be a *second-order function*. This view is

more committed to processing than the model we have presented here (see also Kahane & Lareau, 2016), and for the time being we have nothing cogent to add about this.

14. (Some) conclusions

We began this exploration by asking a simple question: what if the grammar attempted to minimize the number of nodes and maximize connectivity between those nodes, instead of maximizing the number of nodes and establishing ‘unambiguous’ paths between them? One possible answer to that question was sketched here. There are many others: we could have assumed that the grammar *needs* to be the parser (as in Medeiros, forthcoming), or that the operation that we have half-jokingly called ‘Form Graph’ is a stepwise algorithm like its serious and literal counterpart Merge, and explore aspects of cognition and processing. But we have settled for grammatical description: the model presented here aims at describing the full set of relations between nodes (which correspond to the translation of basic expressions into IL) at a derivational point that the reader could identify –should he wanted- with REST’s Surface Structure (which was more of a derivational point than a level of representation), or MP’s Spell-Out (pre-Uriagereka, 2002). The nature of the dynamical process that we are taking a snapshot of is certainly a controversial matter: while we assume something like Krivochen (2016, 2017), Saddy’s (2017), and Saddy & Krivochen’s (2017) oscillatory dynamical computational system, that is *just one of several internally consistent options* (for instance, Gradient Symbolic Computation being an interesting, and potentially compatible, alternative). But the choices we have made in this paper –with which the reader might disagree to different extents- should not obscure the importance of asking the initial question and giving explicit answers.

15. References

- Abney, Steven Paul (1987) *The English Noun Phrase in its Sentential Aspect*. PhD Thesis, MIT.
- Aissen, Judith & David Perlmutter (1983) Clause Reduction in Spanish. In David Perlmutter (ed.) *Studies in Relational Grammar 1*. Chicago: University of Chicago Press. 360-403.
- Bach, Emmon (1970) Problominalization. *Linguistic Inquiry*, 1. 121-122.
- (1979) Control in Montague Grammar. *Linguistic Inquiry*, 10(4). 515-531.
- (1983) Generalized Categorical Grammars and the English Auxiliary. In Frank Heny & Barry Richards (eds.), *Linguistic categories: auxiliaries and related puzzles*, Vol 2. Dordrecht: Reidel. 101–120
- Barss, Andrew (1986) *Chains and Anaphoric Dependence*. PhD Thesis, MIT.
- Beim Graben, Peter & Sabrina Gerth (2012) Geometric Representations for Minimalist Grammars. *Journal of Logic, Language, and Information* 21(4). 393-432.
- Bell, Sarah (1983) Advancements and ascensions in Cebuano. In David Perlmutter (ed.) *Studies in Relational Grammar 1*. Chicago: University of Chicago Press. 143-218.
- Van Benthem, Johan (1988) The Lambek Calculus. In Richard Oherle, Emmon Bach, and Deirdre Wheeler (eds.) *Categorical Grammars and Natural Language Structures*. Dordrecht: Reidel. 35-68.
- Berwick, Robert (1984) Strong Generative Capacity, Weak Generative Capacity, and Modern Linguistic Theories. *Computational Linguistics* 10(3-4). 189-202.
- Bianchi, Valentina & Mara Frascarelli (2010) Is Topic a Root Phenomenon? *Iberia* 2(1). 43–48.
- Binder, Philipe (2008) Frustration in Complexity. *Science* 320. 322-323.

- Borer, Hagit (2003) Exo-Skeletal vs. Endo-Skeletal Explanations: Syntactic Projections and the Lexicon. In John Moore and Maria Polinsky (eds.) *The Nature of Explanation in Linguistic Theory*. Stanford: CSLI Publications. 31-67.
- Bouchard, Denis (1982) *On the Content of Empty Categories*. PhD Thesis, MIT.
- Bravo, Ana, Luis García Fernández, & Diego Krivochen (2015) On Auxiliary Chains: Auxiliaries at the Syntax-Semantics Interface. *Borealis*, 4(2). 71-101. <http://dx.doi.org/10.7557/1.4.2.3612>
- Bresnan, Joan (1983) The passive in lexical theory. In Joan Bresnan (ed.) *The Mental Representation of Grammatical Relations*. Cambridge, Mass.: MIT Press. 3-83.
- (2001) *Lexical Functional Syntax*. [1st Edition]. Oxford: Wiley Blackwell.
- Bresnan, Joan, Ronald M. Kaplan, Stanley Peters and Annie Zaenen (1982) Cross-Serial Dependencies in Dutch. *Linguistic Inquiry*, 13(4). 613-635.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler (2016) *Lexical Functional Syntax* [2nd Edition]. Oxford: Wiley Blackwell.
- Buszkowski, Wojciech (1988) Generative Power of Categorical Grammars. In Richard Oherle, Emmon Bach, and Deirdre Wheeler (eds.) *Categorical Grammars and Natural Language Structures*. Dordrecht: Reidel. 69-94.
- Carrasco Gutiérrez, Angeles & Luis García Fernández (1994) Sequence of Tenses in Spanish. *University of Venice Working Papers in Linguistics* 4(1). 45-70.
- Chomsky, Noam (1955) *The Logical Structure of Linguistic Theory*. Mimeographed, MIT.
http://alpha-leonis.lids.mit.edu/wordpress/wp-content/uploads/2014/07/chomsky_LSLT55.pdf
- (1956) Three models for the description of language. *IRE Transactions on Information Theory* (2): 113–124.
- (1957) *Syntactic Structures*. The Hague: Mouton de Gruyter.
- (1959) On Certain Formal Properties of Grammars. *Information and Control* 2. 137-167.
- (1964) A transformational approach to syntax. In Jerry Fodor & Jerrold Katz (eds.) *The Structure of Language: Readings in the Philosophy of Language*. New Jersey: Prentice Hall. 211-245.
- (1965) *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.
- (1977) Conditions on transformations. In *Essays on Form and Interpretation*. New York: North Holland.
- (1986) *Barriers*. Cambridge, Mass.: MIT Press.
- (1995) *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- (2000) Minimalist Inquiries: the framework. Minimalist Inquiries: The Framework. In R. Martin, D. Michaels and J. Uriagereka (Eds.) *Step by Step – Essays in Minimalist Syntax in Honor of Howard Lasnik*. Cambridge, Mass.: MIT Press. 89-155.
- (2001) Derivation by Phase. In Michael Kenstowicz (ed.) *Ken Hale: A Life in Language*. Cambridge, Mass.: MIT Press. 1-52.

- (2007) Approaching UG from Below. In Uli Sauerland and Hans-Martin Gärtner (eds.), *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics*. Berlin: Mouton de Gruyter. 1-29.
- (2009) Opening Remarks. In Massimo Piatelli-Palmarini, Juan Uriagereka & Pello Salaburu (eds.) *Of Minds and Language*. Oxford: OUP. 13-43.
- (2013) Problems of Projection. *Lingua* 130. 33-49.
- (2015) Problems of Projection: Extensions. In Elisa Di Domenico, Cornelia Hamann & Simona Matteini (eds.) *Structures, Strategies and Beyond: Studies in honour of Adriana Belletti*. Amsterdam: John Benjamins. 1-16.
- Chomsky, Noam & George Miller (1963) Introduction to the Formal Analysis of Natural Languages. In Duncan R. Luce, Robert R. Bush, & Eugene Galanter (eds.), *Handbook of Mathematical Psychology* 2. New York: John Wiley & Sons. 269–321.
- Cinque, Guglielmo (1999) *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford: OUP.
- Citko, Barbara (2005) On the Nature of Merge: External Merge, Internal Merge, and Parallel Merge. *Linguistic Inquiry*, 36(4). 475-496.
- Collins, Chris (2017) Merge(X, Y) = {X, Y}. In Leah Bauke & Andreas Blühmel (eds.) *Labels and Roots*. Berlin: Walter de Gruyter. 47-68.
- Collins, Chris & Edward Stabler (2016) A Formalization of Minimalist Syntax. *Syntax*, 19(1). 43-78.
- Comrie, Bernard (1986) Tense in indirect speech. *Folia Linguistica* 20. 265-296.
- Davies, William & Stanley Dubinsky (2004) *The grammar of raising and control*. Oxford: Blackwell Publishing.
- (2007) *New Horizons in the Analysis of Control and Raising*. Dordrecht: Springer.
- Dowty, David (1978) Governed Transformations as Lexical Rules in a Montague Grammar. *Linguistic Inquiry* 9. 393-426.
- (1985) On Recent Analyses of the Semantics of Control. *Linguistics and Philosophy* 8. 291-331.
- Emonds, Joseph (1970) *Root and structure preserving transformations*. PhD Dissertation, MIT.
- (1979) Appositive relatives have no properties. *Linguistic Inquiry* 10. 211-243.
- Epstein, Samuel, Hisatsugu Kitahara & T. Daniel Seely (1998) *A Derivational Approach to Syntactic Relations*. Oxford: OUP.
- (2012) Structure building that can't be. In Myriam Uribe-Etxebarria & Vidal Valmala (eds.) *Ways of structure building*. Oxford, Oxford University Press. 253-270
- (2015) Simplest merge generates set intersection: Implications for complementizer-trace explanation. In *Explorations in Maximizing Syntactic Minimization*. London: Routledge. 175-194.
- Epstein, Samuel D. (1999) Un-Principled Syntax and the Derivation of Syntactic Relations. In Samuel Epstein & Norbert Hornstein (ed.) *Working Minimalism*. Cambridge, Mass.: MIT Press. 317-345.
- Fiengo, Robert (1977) On Trace Theory. *Linguistic Inquiry*, 8(1). 35-61.

- Fillmore, Charles J. (1963) The Position of Embedding Transformations in a Grammar. *Word* 19(2). 208-231.
- Frank, Robert (1992) *Syntactic Locality and Tree Adjoining Grammar: Grammatical, Acquisition and Processing Perspectives*. Doctoral dissertation, University of Pennsylvania.
- García Fernández, Luis & Diego Krivochen (2017) *Perífrasis verbales y verum focus*. Ms. Universidad Complutense, Madrid & University of Reading.
- García Fernández, Luis, Diego Krivochen & Ana Bravo (2017) Aspectos de la semántica y sintaxis de las cadenas de verbos auxiliares en español. *Moenia* 23.
- Gould, Ronald (1988) *Graph Theory*. California: The Benjamins/Cummings Publishing Company.
- Greibach, Sheila (1965) A New Normal-Form Theorem for Context-Free Phrase Structure Grammars. *Journal of the ACM* 12 (1). 42-52.
- Grimshaw, Jane (2000) Locality and Extended Projection. In Peter Coopmans, Martin Everaert, & Jane Grimshaw (eds.) *Lexical Specification and Insertion*. John Benjamins. 115-133.
- Grindrod, Peter; Zhivko Stoyanov; Garry Smith; and James Douglas Saddy (2014) Resonant Behaviour of Irreducible Networks of Neurons with Delayed Coupling. Ms.
- Hockett, Charles (1954) Two models of grammatical description. *Word*, 10. 210–234.
- (1958) *A Course in Modern Linguistics*. The Macmillan Company: New York.
- Horn, Laurence (1989) *A Natural History of Negation*. Chicago: Chicago University Press.
- Hornstein, Norbert (1999) Movement and control. *Linguistic Inquiry* 30. 69-96.
- Huang, C. T. James (1982) *Logical Relations in Chinese and the theory of Grammar*. PhD Thesis, MIT.
- Jacobson, Pauline (1977) *The Syntax of Crossing Coreference Sentences*. PhD Thesis, University of California, Berkeley.
- (1987) Phrase Structure, Grammatical Relations, and Discontinuous Constituents. In Geoffrey Huck & Almerindo Ojeda (eds.), *Syntax and Semantics 20: Discontinuous Constituency*. New York: Academic Press. 27-69.
- (1990) Raising as function composition. *Linguistics and Philosophy* 13. 423-475.
- (1992) Raising without movement. In Richard K. Larson, Sabine Iatridou, Utpal Lahiri, and James Higginbotham (eds.) *Control and Grammar*. Dordrecht: Kluwer Academic Publishers. 149-194.
- Jackendoff, Ray (2011) Alternative minimalist visions of language. In Robert D. Borsley and Kersti Börjars (eds.) *Nontransformational Syntax*. Oxford: Wiley-Blackwell. 268-296.
- Jiménez Fernández, Angel (2015) Towards a typology of focus: Subject position and microvariation at the discourse–syntax interface. *Ampersand* 2. 49-60.
- Jespersen, Otto (1985) [1937] *Analytic Syntax*. Chicago: University of Chicago Press.
- Johnson, David & Paul M. Postal (1980) *Arc Pair Grammar*. Princeton, NJ.: Princeton University Press.

- Kahane, Sylvain & François Lareau (2016) Word ordering as a graph-rewriting process. In Annie Foret, Glyn Morrill, Reinhard Muskens, Rainer Osswald & Sylvain Pogodalla (eds.) *Formal Grammar*. Berlin: Springer. 216–239. DOI: 10.1007/978-3-662-53042-9_13
- Karttunen, Lauri (1971) Definite Descriptions with Crossing Reference. *Foundations of Language* 7. 157-182.
- Katz, Jerry & Paul Postal (1964) *An Integrated Theory of Linguistic Descriptions*. Cambridge, Mass.: MIT Press.
- Kayne, Richard (1983) Connectedness. *Linguistic Inquiry*, 14(2). 223-249.
- (1984) Connectedness and Binary Branching. Dordrecht: Foris.
- (1994) *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.
- Koopman, Hilda & Dominique Sportiche (2000) Variables and the Bijection Principle. En Koopman, Hilda: *The Syntax of Specifiers and Heads*. London: Routledge. 16-36.
- Koyama, Masanori, Michael Orrison and David Neel (2007) Irreducible graphs. *Journal of Combinatorial Mathematics and Combinatorial Computing* 62. 35-43.
- Kracht, Marcus (2001) Syntax in chains. *Linguistics and Philosophy* 24. 467–529.
- (2008) On the logic of LGB type structures. Part I: Multidominance structures. In Fritz Hamm & Stephan Kepser (eds.) *Logics for Linguistic Structures*. Berlin: Mouton. 105–142.
- Krivine, Jean-Louis (1971) *Introduction to Axiomatic Set Theory*. Dordrecht: Reidel.
- Krivochen, Diego (2013) Los verbos de ascenso como expresiones modales: el caso del español. *Anu.Filol.Est.Lingüíst.*, 3. 33-56.
- (2015a) On Phrase Structure Building and Labeling Algorithms: Towards a Non-Uniform Theory of Syntactic Structures. *The Linguistic Review* 32(3). 515-572.
- (2015b) Types vs. Tokens: Displacement Revisited. *Studia Linguistica*, 70(3). 250-296. DOI: 10.1111/stul.12044
- (2016) Divide and...conquer? On the limits of algorithmic approaches to syntactic structure. *Czech and Slovak Linguistic Review* 1(2016). 15-38.
- (2017a) Aspects of Emergent Cyclicity in Language and Computation. PhD Thesis. University of Reading.
- (2017b) On intensive endophoric devices in English. Ms. University of Reading.
- (under review) A trans-derivational expansion of Generative Semantics: constraining dependencies in secondary predication. Ms. University of Reading. [Handout version available at: https://www.academia.edu/33485672/2017_A_trans-derivational_expansion_of_Generative_Semantics_Handout_]
- Krivochen, Diego & Luis García Fernández (2018) Perífrasis verbales y *verum focus*. Ms. University of Reading and Universidad Complutense de Madrid.
- Krivochen, Diego, Luis García Fernández & Ana Bravo (in preparation) Auxiliary chains and Clitic Climbing in Spanish. Ms.

- Krivochen, Diego & Susan Schmerling (2016a) Two kinds of coordination and their theoretical implications: An Aspect-Based Approach. Ms. Under review.
- (2016b) Mirage coordination in English. Ms.
[https://www.academia.edu/25110015/ Squib Mirage Coordinations in English](https://www.academia.edu/25110015/Squib_Mirage_Coordinations_in_English)
- Kroch, Tony & Aravind K. Joshi (1987) Analyzing Extraposition in Tree Adjoining Grammar. In Geoffrey Huck & Almerindo Ojeda (eds.) *Syntax and Semantics 20: Discontinuous Constituency*. New York: Academic Press. 107-151.
- Ladusaw, William (1980) *Polarity sensitivity as inherent scope relations*. Bloomington, Indiana: University of Iowa, Indiana University Linguistics Club.
- Lakoff, George & John Robert Ross (1967) [1976] Is Deep Structure Necessary? Letter to Arnold Zwicky, printed in James D. McCawley (1976) (ed.). *Notes from the linguistic underground* (Syntax and Semantics 7). New York: Academic Press. 159-164.
- Langacker, Ronald (1969) On pronominalization and the chain of command. En D. Reibel & S. Schane (eds.) *Modern Studies in English*. Engelwood Cliffs, NJ: Prentice-Hall. 160-186.
- Langendoen, D. Terrence & Paul M. Postal (1984) *The Vastness of Natural Languages*. Oxford: Blackwell.
- Lasnik, Howard (2011) What Kind of Computing Device is the Human Language Faculty?. In Di Sciullo, A-M. & C. Boeckx (Eds.) *The Bilingual Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*. Oxford: OUP. 354-65.
- Lasnik, Howard and Joseph Kuppin (1977) A restrictive theory of transformational grammar. *Theoretical Linguistics* 4. 173–196.
- Lasnik, Howard & Juan Uriagereka (2012) Structure. In Ruth Kempson, Tim Fernando and Nicholas Asher (eds.) *Handbook of philosophy of science, volume 14: Philosophy of linguistics*. London: Elsevier. 33–61.
- Lebeaux, David (1994) Where does Binding Theory apply? Ms. MIT.
- (2009) *Where does Binding Theory Apply?* Cambridge, Mass.: MIT Press.
- Lewis, David (1972) General Semantics. In Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language*. Reidel, Dordrecht. 169-218.
- Lyons, John (1968) *Introduction to theoretical linguistics*. Cambridge: Academic Press.
- McCawley, James D. (1968) Concerning the base component of a transformational grammar. *Foundations of Language* 4. 243-269.
- (1970) Where do Noun Phrases come from? In Roderick Jacobs & Peter Rosenbaum (eds.) *Readings in English Transformational Grammar*. Waltham: Ginn & Co. 166-183.
- (1982) Parentheticals and Discontinuous Constituent Structure. *Linguistic Inquiry* 13(1). 91-106.
- (1998) *The Syntactic Phenomena of English*. [2nd Edition]. Chicago: University of Chicago Press.
- Medeiros, David P. (forthcoming) ULTRA: Universal Grammar as a Universal Parser. Invited paper for *Frontiers In Psychology: Language Sciences*. Research Topic Approaches to Language: Data, Theory, and Explanation.

- Mel'čuk, Igor (1988) *Dependency Syntax: Theory and Practice*. Albany: State University of New York Press.
- Montague, Richard (1973) The proper treatment of quantification in ordinary English. In Hintikka, Jaakko & Moravcsik, Julius & Suppes, Patrick (eds.), *Approaches to natural language*. Dordrecht: Reidel. 221–242.
- (1974) English as a Formal Language. In Richmond Thomason (ed.) *Formal Philosophy: Selected Papers of Richard Montague*. New Haven: Yale University Press. 188-221.
- Morin, Yves-Charles and Michael O'Malley (1969) Multi-rooted vines in semantic representation. In Robert Binnick et al. (ed.) *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*. University of Chicago. 178-185.
- Moro, Andrea (2000) *Dynamic Antisymmetry*. Cambridge, Mass.: MIT Press.
- Müller, Stefan (2000) The passive as a lexical rule. In Dan Flickinger and Andreas Kathol (eds.) *Proceedings of the 7th International HPSG Conference*. Stanford, CLSI. 247-266.
- (2013) Unifying everything: Some remarks on simpler syntax, construction grammar, minimalism, and HPSG. *Language* 89(4). 920-950.
- Nunes, Jairo (2004) *Linearization of Chains and Sideways Movement*. Cambridge, Mass.: MIT Press.
- Ojeda, Almerindo (1987) Discontinuity, Multidominance, and Unbounded Dependency in GPSG. In Geoffrey Huck & Almerindo Ojeda (eds.) *Syntax and Semantics 20: Discontinuous Constituency*. New York: Academic Press. 257-282.
- (2005) Discontinuous Constituents. In Brown, Keith (Ed.) *Encyclopedia of Language and Linguistics*. [2nd Edition]. Elsevier. Volume 3. 624-630.
- Ordóñez, Francisco (2012) Clitics in Spanish. In José Ignacio Hualde, Antxon Olarrea and Erin O'Rourke (eds.) *The Handbook of Hispanic Linguistics*. Oxford: Blackwell. 423-451.
- Osborne, Timothy (2005) Beyond the Constituent: A Dependency Grammar Analysis of Chains. *Folia Linguistica XXXIX/3-4*. 251-297.
- Osborne, Timothy & Ruochen Niu (2017) The Component Unit: Introducing a Novel Unit of Syntactic Analysis. Paper presented at Fourth International Conference on Dependency Linguistics, Università di Pisa, Italy.
https://www.researchgate.net/publication/320010087_The_Component_Unit_Introducing_a_Novel_Unit_of_Syntactic_Analysis
- Osborne, Timothy, Michael Putnam, and Thomas Groß (2011) Bare phrase structure, label-less trees, and specifier-less syntax: Is Minimalism becoming a dependency grammar? *The Linguistic Review* 28. 315-364.
- Partee, Barbara (1974) Some transformational extensions of Montague grammar. Ms, MIT.
- Perlmutter, David (1968) Deep and Surface Structure Constraints in Syntax. PhD Thesis, MIT.
- (1983) Introduction. In David Perlmutter (ed.) *Studies in Relational Grammar I*. Chicago: University of Chicago Press. ix-xv.

- Perlmutter, David and Postal, Paul M. (1983a) Toward a universal characterization of passivization. In David Perlmutter (ed.) *Studies in Relational Grammar I*. Chicago: University of Chicago Press. 3-29.
- (1983b) Some Proposed Laws of Basic Clause Structure. In David Perlmutter (ed.) *Studies in Relational Grammar I*. Chicago: University of Chicago Press. 81-128.
- Peters, Stanley and R. W. Ritchie (1973) On the generative power of transformational grammars. *Information Sciences* 6. 49-83.
- (1981) Phrase Linking Grammars. Ms. Stanford University and Palo Alto, California.
- Pollard, Carl & Ivan Sag (1984) *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press and Stanford: CSLI Publications.
- Postal, Paul M. (1964) *Constituent Structure*. Bloomington, Indiana: University of Bloomington.
- (1969) On so-called 'pronouns' in English. In David Reidel & Sanford Schane (eds.) *Modern Studies in English: Readings in Transformational Grammar*. New Jersey: Prentice Hall. 201-224.
- (1972) On some rules that are not successive-cyclic. *Linguistic Inquiry*, 3(2). 211-222.
- (1974) *On Rising*. Cambridge, Mass.: MIT Press.
- (1986) *Studies of Passive Clauses*. New York: SUNY Press.
- (1997) *Three Investigations on Extraction*. Cambridge, Mass.: MIT Press.
- (2004) *Skeptical Linguistic Essays*. Oxford: OUP.
- (2010) *Edge-Based Clausal Syntax*. Cambridge, Mass.: MIT Press.
- Progovic, Ljiljana (1998) Structure for coordination. *Glott International* 3(7). 3-6.
- Pullum, Geoffrey K. (2007) The evolution of model-theoretic frameworks in linguistics. In James Rogers and Stephan Kepser (eds.), *Model-Theoretic Syntax at 10*. Dublin: Trinity College Dublin. 1-10.
- Pullum, Geoffrey K. and Barbara C. Scholz. (2001) On the distinction between model-theoretic and generative-enumerative syntactic frameworks. In Philippe de Groote, Glyn Morrill, and Christian Retoré (eds.) *Logical Aspects of Computational Linguistics: 4th International Conference (Lecture Notes in Artificial Intelligence, 2009)*. Berlin: Springer Verlag. 17-43.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik (1985) *A Comprehensive Grammar of the English Language*. London: Longman.
- Richards, Marc (2011) Deriving the Edge: What's in a Phase? *Syntax* 14(1). 74-95.
- Ringen, Catherine (1972) On Arguments for Rule Ordering. *Foundations of Language*, 8(2). 266-273.
- Rizzi, Luigi (1990) *Relativized minimality*. Cambridge, MA: MIT Press.
- (1997) The Fine Structure of the Left Periphery. In Liliane Haegeman (ed.) *Elements of Grammar: A Handbook of Generative Syntax*. Dordrecht: Kluwer. 281-337.
- (2004) Locality and left periphery. In Adriana Belletti (ed.) *Structures and Beyond: The cartography of syntactic structures*. New York: Oxford University Press. 223-251.

- Rogers, Andy (1974) A Transderivational Constraint on Richard? *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*. 551-558.
- Rooryck, Johan (1992) Negative and Factive Islands. *Journal of Linguistics*, 28. 343-374.
- Rosenbaum, Peter (1965) The Grammar of English Predicate Complement Constructions. PhD Thesis, MIT.
- Ross, John Robert (1967) *Constraints on Variables in Syntax*. PhD Thesis, MIT.
- (1969a) On the Cyclic nature of English Pronominalization. In David Reidel & Sanford Schane (eds.) *Modern Studies in English: Readings in Transformational Grammar*. New Jersey: Prentice Hall. 187-200.
- (1969b) A proposed rule of tree pruning. In David Reidel & Sanford Schane (eds.) *Modern Studies in English: Readings in Transformational Grammar*. New Jersey: Prentice Hall. 288-299.
- (1970) On Declarative Sentences. In Roderick Jacobs & Peter Rosenbaum (eds.) *Readings in English Transformational Grammar*. Waltham: Ginn & Co. 222-272.
- (1986) Inner islands. In Brugman, Claudia, Monica Macaulay et al. (eds.), *Proceedings of the 10th Annual Meeting of the Berkeley Linguistics Society*. Berkeley, 258-265.
- (2011) An automodular perspective on the frozenness of pseudoclefts, and *vice versa*. 2011. In Tista Bagchi, Katharine P. Beals and Etsuyo Yuasa (eds.), *Pragmatics and Autolexical Grammar: In Honor of Jerry Sadock*. Amsterdam and Philadelphia: John Benjamins. 243-260.
- (2012) A preliminary –but fattened– list of transformations. Ms. <http://www-personal.umich.edu/~jlawler/haj/Preliminarybufattenedlistoftransformations.pdf>
- Saddy, Douglas (2017) Syntax and Uncertainty. In Roger Martin (ed.) *Language, Syntax, and the Natural Sciences*. Cambridge: CUP.
- Saddy, Douglas & Diego Gabriel Krivochen (2017) Classical Problems, non-Classical proposals. Ms., University of Reading.
- Saddy, Douglas and Peter Grindrod (2016) Irreducible Subnets and Random Delays: Limiting the Vocabulary of Neural Information. Presented at First International Symposium on the Physics of Language. Sophia University, Tokyo. March 4-5, 2016.
- Sag, Ivan (2010) English Filler-Gap constructions. *Language* 86(3). 486-545.
- Sampson, Geoffrey (1975) The Single Mother Condition. *Journal of Linguistics* 11(1), 1-11.
- Schmerling, Susan F. (1975) Asymmetric Conjunction and rules of Conversation. In Peter Cole and Jerry Morgan (eds.) *Syntax and Semantics, Vol. 3: Speech Acts*. New York: Academic Press. 211-231.
- (1983a) Two theories of syntactic categories. *Linguistics and Philosophy* 6. 393–421.
- (1983b) A new theory of English auxiliaries. In Frank Heny & Barry Richards (eds.), *Linguistic categories: auxiliaries and related puzzles*. Vol 2. Dordrecht: Reidel. 1–53.
- (1988) On the definite article in German. November, 1988 revision of a paper presented at the *Symposium on Determiners and A' Binding*. Austin, TX. March 22-24, 1985.

- (2016) *Towards a Neo-Sapirian Theory of Language*. Ms. [https://www.academia.edu/19277408/Towards a Neo-Sapirian Theory of Language revised 12 2015](https://www.academia.edu/19277408/Towards_a_Neo-Sapirian_Theory_of_Language_revised_12_2015)
- Schmerling, Susan F. & Diego Krivochen (2017) On Non-Progressive 'being'. *Canadian Journal of Linguistics*. DOI: <https://doi.org/10.1017/cnj.2017.38>
- Shieber, Stuart (1986) *An Introduction to Unification-Based Approaches to Grammar*. Brookline, Mass.: Microtome Publishing.
- Sloan, Kelly & Juan Uriagereka (1988) What does 'everyone' have scope over? GLOW. Budapest.
- Sloat, Clarence (1969) Proper nouns in English. *Language* 45(1). 26–30.
- Sternefeld, Wolfgang (1998a) Connectivity effects in pseudo-cleft sentences. In Artemis Alexiadou, Nanna Fuhrhop, Paul Law & Ursula Kleinhenz (eds.) *ZAS Papers in Linguistics* 10. 146-162.
- (1998b) The semantics of reconstruction and connectivity. *Arbeitspapiere des SFB 340*. Universität Stuttgart und Tübingen. 1-58.
- Stroik, Tom & Michael T. Putnam (2013) *The Structural Design of Language*. Oxford: OUP.
- (2015) Is Simplest Merge too simple? Ms. University of Missouri-Kansas City & Penn State University.
- Tabor, Whitney (2015) Fractal grammars which recover from perturbations. Proceedings of the 2015 International Conference on Cognitive Computation: Integrating Neural and Symbolic Approaches - Volume 1583. 120-128.
- Tesnière, Lucien (1959) *Éléments de syntaxe structural*. Paris: Klincksieck.
- Torrego, Esther (1984) On Inversion in Spanish and Some of Its Effects. *Linguistic Inquiry* 15(1). 103-129.
- (2002) Arguments for a derivational approach to syntactic relations based on clitics. In Samuel Epstein & Daniel T. Seely (eds.) *Derivation and explanation in the Minimalist Program*. Malden, MA: Blackwell. 249–268.
- Uriagereka, Juan (2002) Multiple Spell-Out. In Uriagereka, J. *Derivations: Exploring the Dynamics of Syntax*. London: Routledge. 45-65.
- (2008) *Syntactic Anchors: On Semantic Restructuring*. Cambridge: CUP.
- (2012) *Spell-Out and the Minimalist Program*. Oxford: OUP.
- Van Steen, Maarten (2010) *Graph Theory and Complex Networks: An Introduction*. Available for free at <https://www.distributed-systems.net/index.php/books/gtcn/gtcn/>
- Villata, Sandra, Luigi Rizzi, & Julie Franck (2016) Intervention effects and Relativized Minimality: New experimental evidence from graded judgments. *Lingua* 179. 76-96. <http://dx.doi.org/10.1016/j.lingua.2016.03.004>
- Wall, Robert (1972) *Introduction to Mathematical Linguistics*. New York: Prentice Hall.
- Wasow, Thomas (1972) *Anaphoric Relations in English*. PhD Thesis, MIT.

Wiggins, Stephen (2003) *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. [2nd Edition] New York: Springer.

Williams, Edwin (1982) Another argument that the passive is transformational. *Linguistic Inquiry*, 13(1). 160-163.

(2003) *Representation Theory*. Cambridge, Mass: MIT Press.

Wilson, Robin (1996) *Introduction to Graph Theory*. [4th edition]. London: Addison Wesley.

Wells, Rulon S. (1947) Immediate Constituents. *Language* 23. 81-117.

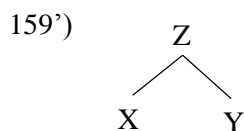
Zwart, Jan-Wouter (2009) Prospects for top-down derivations. *Catalan Journal of Linguistics* 9. 161-187.

Appendix A: On some inadequacies of trees to represent Merge

A most surprising decision of MGG is to use trees to represent bottom-up derivations built via Merge. The reason why that is surprising is that in the tree corresponding to the result of

159) Merge(X, Y)

Which is



There is no direct connection between X and Y: the only path from X to Y is the one that goes through Z. This is, we think, a consequence of history rather than theory: for PSR, the format in (159') is indeed suitable. Thus,

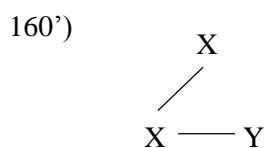
159'') $Z \rightarrow X, Y$

Is strongly equivalent to (159'): the PSR goes *from* higher nodes *to* lower nodes, and no direct connection between X and Y is ever suggested. But Merge is different: derivations are built from the bottom upwards, and the rule only makes direct reference to X and Y (Chomsky, 2015; Collins, 2017). Thus, we should have (160):

160)

```
graph LR; X --- Y;
```

The structure generated by Merge is then labelled (either by 'Minimal Search', whatever that is, or via more classical procedures), with the label being either X or Y (Chomsky, 1994): Merge can *only* generate endocentric binary branching structures (Kayne, 1994 and much subsequent work). Take (160) as the input for a further application of Merge: how does the system proceed? If (160) needs to be labelled before entering further syntactic relations, the intermediate step between 'first-Merge' and 'second-Merge' (see, e.g., Zwarts, 2009) is (160')

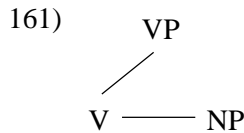


Concretely, assume that X = V and Y = NP (given that Merge of {H, XP} is 'virtually everything'³⁷, according to Chomsky, 2009). Then,

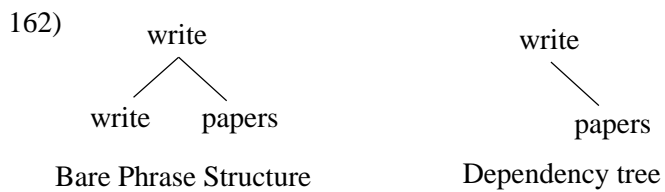
³⁷ As Postal (2004: 323) correctly observes,

given the meaning of virtually [...], a claim that P is 'virtually conceptually necessary' admits that it is not conceptually necessary.

The same holds for 'virtually everything'. In my opinion (and that of Postal, Behme, and others), these rhetorical resources are a way of refusing to acknowledge alternatives and respond to critics, because, what kind of person argues against something that is 'necessary' or 'everything' (regardless of the 'virtual' character of that necessity)?

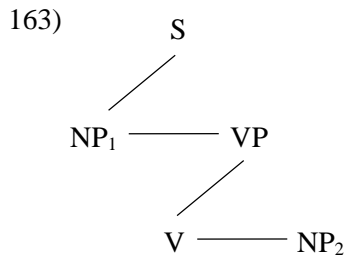


Intuitively, this makes sense: the NP is selected by the verb, not by the verb phrase. And it is the V-NP complex that determines the Aktionsart of VP. There is nothing in the definition of Merge (at least in those that are explicit enough to be evaluated) that yields constituency, the ‘is a’ relation (which was a given in PSR) is not part of Merge, but of Label. And if a label is nothing but a syntactic diacritic (because it’s the projection of the features of a head, basically identical to that head but with the added information ‘I’m phrasal’), there is no constituency there either. Crucially, there are recent proposals to *eliminate* labels altogether (Collins, 2017; Epstein et al., 2012, 2015, among others; see Stroik & Putnam, 2015 for a critical view): this pushes Minimalism away from Phrase Structure Grammars and, as Osborne et al. (2011) correctly point out, moves it towards Dependency Grammars. As a consequence, any reference to ‘constituency’ and most references to labelling (particularly, finite VP/vPs) becomes void, because Dependency Grammars have no notion of ‘constituency’ (in the strict sense in which this is understood in IA theories structures around PSRs). The point that Osborne et al. (2011) make can be summarized in the following equivalence in both strong and weak generative capacity:



In what follows, we will discuss a version of Minimalism which assumes some form of constituency and labels, the proposals in the previous paragraph notwithstanding. We will do that because, in practice, most Minimalists still make use of formalisms that resemble more or less ‘traditional’ IA phrase structure grammars.

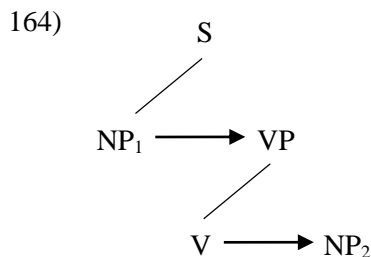
In (161) we assembled a verb phrase, containing a V and its object following the idea in Chomsky (1995: 246) that Merge is asymmetrical (see also Epstein, et al.’s 1998 *Introduction*; also Collins & Stabler, 2016). Now, we introduce the subject, which Merges *to* the VP in the initial version of the theory (and *with* VP in the ‘unordered set’, more recent version). The problem is that sentences are not endocentric (what it means for a sentence to be a ‘TP/IP’ or a ‘CP’ is not clear and where exactly it matters in the architecture of the grammar, and it doesn’t seem to translate into any non-IA grammar), which was indeed acknowledged in pre-1986 MGG, not to mention (European and American) structuralism and IP grammars, with no loss of empirical coverage:



Note that walking the path from NP₂ to S gives us all the Merge relations: horizontal lines translate as ‘merges to’, whereas diagonal lines translate into ‘projects as’. Regardless of how we diagram it, be it horizontal and diagonal lines, or solid and dotted lines, or whatever else, there is a fundamental difference between ‘merges with’ and ‘projects as’. Thus, (163) reads as:

163’) V merges to NP₂ and they project a VP. NP₁ merges to VP and they project an S.

But the graphs in (143 - 147) are not particularly good at representing the *asymmetry* of Merge, which MGG has argued is an essential property of the operation, deriving crucial notions like c-command and labelling (see e.g., Epstein, 1999 and much subsequent work; also Collins & Stabler, 2016). We can do that with arrows: X → Y means ‘X merges to Y’, such that (163) becomes (164):

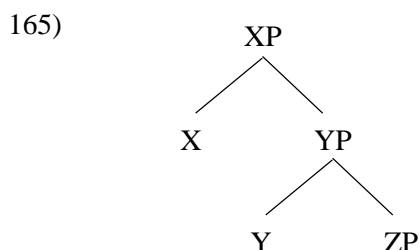


Walking from S to NP₂ follows a unique, unambiguous path in which the relevant relations are indeed preserved. In this context, some more graph theory is useful. A connected graph G is said to contain a *Hamilton* (or *Hamiltonian*) *path* if there is a path that visits every vertex of G exactly once (Van Steen, 2010: 92). In this sense, (164) is a Hamilton path, as opposed to regular PS trees, which only *locally* contain Hamilton paths. Note that a graph may contain a Hamilton path and *not* be Hamiltonian: for a graph G to be Hamiltonian it needs to be *cyclic*, and contain a cyclic path visiting every vertex of G exactly once.

This last claim requires some unpacking.

Tree building structure must always manipulate elements of the form {X, YP}, where X is a head and thus a terminal node (Chomsky, 2009; 2013); this is so both for the purposes of the Linear Correspondence Axiom (Kayne, 1994) as well as labelling: the LCA is a function from terminals to non-terminals, and the labelling algorithm proposed by Chomsky is sensitive to prominent features of X, which project a label in {X, YP} yielding {XP {X, YP}} or, equivalently in *bare phrase structure*, {X, {X, Y}} (in turn, by the pairing axiom in Zermelo/Fraenkel set theory, equivalent to <X, Y>; Krivine, 1971: 3).

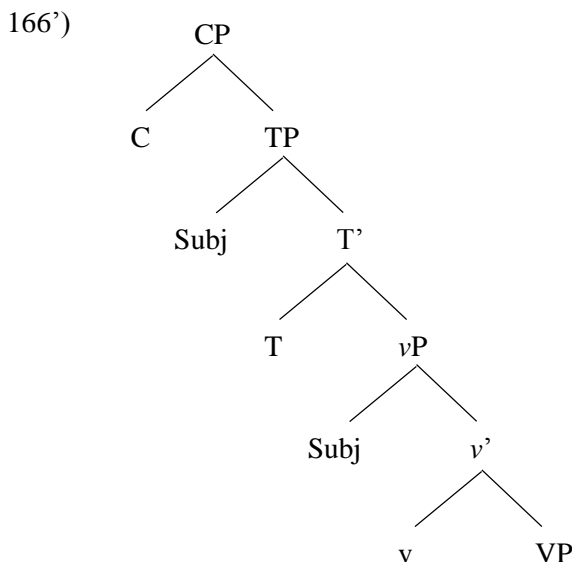
Consider now the abstract structure in (165):



As it is, (165) contains *two* local Hamilton paths:

- 166) i) X → XP → YP → Y
 ii) Y → YP → ZP → ...

And that's it. That means that there is no way to connect X and anything within ZP via a Hamilton path (because in order to do so we would need to skip Y, and that would leave a node 'unwalked'). That yields a strong locality condition for operations triggered by X, which is precisely what strong phase-based theories claim (if phase heads follow a certain derivational rhythm Phase-Non phase-Phase... see e.g., Chomsky, 2001; Richards, 2011). The cycle would be, from the bottom upwards, [YP [ZP]], with Z a phase head, and Y a non-phase head. In this state of affairs, X can affect Y and its 'edge', but not its complement. A configuration like (166) arises, with irrelevant differences, in the C-T-v skeleton assumed in MGG (but *not* in its label-less variants, we hasten to add):



Where C and *v* are phase heads, and T and V are not (Chomsky, 2001 and much subsequent work). C can probe into TP up to *v*P's edge (e.g., to find a Wh-phrase) *only* if there is no Subj in Spec-T, for Minimality reasons. The way in which Chomsky deals with these issues is stipulative, defining 'equidistance' in checking domains (Chomsky, 1995). Intuitively, we would like to say that if a head H looks for an NP –say-, it stops looking at the first NP it sees: this is walking a path. We are, though, forced to skip nodes in doing so if we try to keep things as short as possible.

But taking trees too seriously generates some problems and inconsistencies within MGG. For starters, probing requires the computation to indeed go and visit the same node at least twice. Say C is looking for an NP with a Wh-feature to satisfy the Wh-criterion (Rizzi, 2004). It cannot know that Subj in Spec-TP will not have such a feature (unless we are shamelessly accepting stipulations): if probing is some kind of searching mechanism, it will follow the walk:

167) C → CP → TP → Subj

And find no Wh-feature there. Now, two things can happen: one, the system starts walking the path again, jumping over Subj

167') C → CP → TP → T' → T

And find that there is no NP there...and so on. In this view, for each failed probing, there's a new walk which needs to go back to the root C to start over (each of which is an *Eulerian tour* of the graph; Wilson, 1996: 4; Van Steen, 2010: 82-83), and which 'jumps over' the nodes that have failed to satisfy the requirements of the probe. This requires the system to have some memory (more than a *last-in-first-out* stack) as well as access to sub-lexical features. It will need to keep track of what it has already visited and what has failed to satisfy the requirements of the probing head.

Or, as an alternative option, the system can go back a node, to the immediately dominating branching node, and keep looking:

168) $C \rightarrow CP \rightarrow TP \rightarrow \text{Subj}_{[\text{NO}]} \rightarrow TP \rightarrow T' \rightarrow T_{[\text{NO}]} \rightarrow \dots$

We have indicated the nodes at which failed attempts are recognized as such with [NO]. In our view, both options are not only problematic, but also unjustifiedly so: these complications arise because the *single mother condition* and *binary branching* are taken to be hard constraints on phrase structure. Not to mention operations over features. Neither of these are required by the data, we argue.

There are further problems. Take the translation from structural order to linear precedence for example: linearization must apply only to terminals, and it must not be sensitive to non-terminals (because only terminal nodes are targets for lexical insertion). The LCA is a total order from terminals to nonterminals (since in a structure of the kind $\{H, XP\}$, H asymmetrically c-commands the Spec- of XP, X, the head of XP, and of course the complement of X; thus, H precedes all elements of XP), but strings are concatenations of terminal nodes. Thus, what the grammar must generate in a Kaynean/Chomsky fashion is a representation in which

169) $X \rightarrow Y \rightarrow Z$

Holds, and *not any other walk*: linearization can proceed only if the sequence is *unambiguous*.

Under a Kaynean/Chomskyan view of phrase structure, in a phrase marker like (149), ZP must be an empty category (i.e., a category with no phonological content) for linearization purposes: The Linear Correspondence Axiom maps asymmetric c-command into precedence. Wherever two nodes c-command each other, that is called a symmetry point, and must be broken via movement of either (in principle; the rightmost in practice, as in Moro, 2000). Thus, at the relevant level of representation, ZP is not there to be 'walked on'. In Phase Theory, given a rhythm Phase-Non-Phase-Phase..., ZP is not accessible by operations at X, which -relevantly- we take to mean that there is no path communicating X and ZP. How exactly non-terminals are 'ignored' is a problem in and of itself (to which MGG does not offer explicit solutions), and it is not clear how that could be actually implemented (but see Osborne et al., 2011 for some clues).

Appendix B: What remains of the generative barrel

Postal (2010: 1) identifies a set of concepts which constitute the received wisdom for anyone attempting to do NL syntax within a generative framework. He calls that set *Barrel A*. Here, we present an adapted, slightly annotated, and –sometimes- updated version of *Barrel A*, specifying which elements of Barrel A remain in our own barrel. This is important for various reasons: perhaps the most practical of them is that, since it does not make sense to refute a theory because it does not appeal to or is incompatible with some orthogonal theoretical principle, we do need to specify what it is that we don't assume or what we explicitly reject.

What we keep	What we don't
abstract case	atomic node labels
lexical entries	atomic traces
lexical rules	binding principles based on c-command (principles A, B, C)
Clause Reduction / restructuring	c-command
theta roles	complex node labels composed of sets of feature(s) (specifications)
the case filter	configurational definitions of grammatical relations
the principle of full interpretation	constituent structure trees
the structure preserving hypothesis	copy traces
the theta-criterion	derivations
the Wh-Island constraint	economy principles (Minimal Link Condition, Greed, Enlightened Self-Interest, Suicidal Greed, Procrastinate, Earliness, Last Resort)
the Coordinate Structure Constraint	empty nodes
transformations	Agree (feature checking / valuation / donation / sharing...)
the Complex NP Constraint	phrase structure rules
	(Relativized) minimality
	the A-over-A principle
	the chain condition
	Empty Category Principle
	the Extension condition
	the No-Tampering Condition
	the head movement constraint
	the principle of recovery of deletion
	the projection principle
	the superiority condition
	the visibility condition
	X-bar theory
	Merge
	Subjacency

Note that we have no quarrel with either grammatical functions or Case; the only caveat is that grammatical functions are primitives in our proposal (as in RG) rather than being defined over structural configurations (daughter-of-S for Subject, sister-of-V for Object). Because we maintain transformations (as purely *descriptive* tools) and constructions (in the same manner), lexically governed processes (say, Wrap or Dative Shift) do have a place in our grammar; and so do construction-specific constraints (Wh-islands, CNPC, CSC). However, phrase structure and, more

specifically, X-bar theory and Merge, do not (nor do specific conditions over these; the NTC and the Extension Condition are thus eliminated as unnecessary). Structure is not obtained by means of discrete recursive combinatorics, and it is not the aim of the theory presented here to provide a recursive enumeration of the strings generated by a grammar: in contrast, we aim at obtaining a ‘finite, rigorous characterization’ of English sentences (the expression is Langendoen & Postal’s) by means of descriptive connectivity graphs which exhaust the sets of relations in specific expressions. More generally, there are no second-order principles which quantify over rules of the grammar (Postal, 2010: 6; Pullum & Scholz 2001); only first-order conditions over admissible expressions. The rejection of IA formalisms has been explicit throughout the paper (we hope), due to its ambivalence: they are too restrictive in some cases (more often than not, they do not allow for discontinuity and multidominance; they are also committed to uniformity desiderata), and too permissive in others (requiring gratuitous *ad hoc* assumptions, for instance, to restrict structure building or mapping; see Peters & Ritchie, 1973 for an early analysis of the unbounded generative power of some generative grammars). Phrase structure rules and derivations are replaced by graph admissibility conditions, and reordering transformations are replaced by...well, also graph admissibility conditions. It is crucial to note that (in line with other model-theoretic approaches) at no point are we dealing with infinite sets of (weakly generated) sentences. There is no *structure building vs. structure mapping* (or *construal vs. movement*) debate. No movement also means no chains, copies, traces, and conditions over their distribution (thus, no ECP). *Qua* second-order condition (over Move- α), the Head Movement Constraint has no place in this theory; nor does Relativized Minimality if conceptualized as a condition over rules (a constraint over Move / Agree), Superiority, or A-over-A (among others). Any principle of the form *No rule can relate X, Y in configuration C* (e.g., Chomsky’s 1977: 101 formulation of Superiority; Rizzi’s 2004: 223 formulation of Relativized Minimality) quantifies over a rule, and thus belongs in a proof-theoretic –but *not* a model-theoretic- framework.

Our adoption of a radical version of *economy of expression* further restricts the possibility of having base-generated empty nodes (essentially, anything dominating only \emptyset can, and must, be pruned). There is also no universal hierarchy of functional nodes and abstract morphemes, as the presence of any node must be empirically motivated *for each expression of the language*: only individual expressions satisfy or not the model.