Syntax as Graph Theory

Diego Gabriel Krivochen

University of Reading, CINN

diegokrivochen@hotmail.com

Contents

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 Synaxe 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 and strings in natural language are assigned structural descriptions. Perhaps the most popular way of representing hierarchical structure in language is by means of *tree diagrams* which correspond to derivations of sentences obtained by either top-down rewriting rules (as in Phrase Structure Grammars) or bottom-up recursive discrete combinatorics. These trees belong to the larger class of mathematical objects known as *graphs*, and are subjected to a set of conditions which determine their growth properties, axiomatically favouring binary-branching endocentric structures with no cycles or loops. The present monograph explores a radically different option. Rather than minimising connectivity between nodes which correspond to lexical items in order to get 'unambiguous paths' in command relations, we will investigate the empirical and theoretical consequences of pursuing the idea that syntactic structure grows by maximally connecting sub-graphs in local domains, whose nodes stand not for lexical items, but rather for intensional logic translations of expressions of the language. The result will be a model of grammar in which empirical aspects of natural language that have proven problematic for mainstream generative grammar (MGG) and accounting for which has required supplemental *ad hoc* assumptions -like discontinuity in constituent structure and crossing dependencies- have a natural place.

Keywords: graph theory; discontinuity; multidominance; syntax

1. Introduction

1.1 Phrase markers, strings, and trees: characterisation and limitations

Diagrams of structural descriptions for natural language strings adopt many forms. A structural description for a string is an annotated analysis of that string, in which local hierarchical relations are indicated by various means: for example, *phrase structure grammars* (PSGs), perhaps the best known of grammatical formalisms, appeal to a combination of phrasal labels (NP, VP, and the like) and configurational information (the primitives are two-place predicates: *dominates*, *precedes*). 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 (as McCawley, 1998: 47-48 warns) must not be confused with *diagrams* of L-trees, which are merely pictures composed of lines and symbols (in other words, they are typographical objects rather than mathematical objects). L-trees are *generated* (read: *recursively enumerated*) in the sense of Post (1944: 285-286) by a system of intrinsically ordered rules operating over a set of symbols (an *alphabet*); such a system is customarily referred to as a *grammar* (see also Post, 1943: 203, ff. for a presentation of a deterministic system based on rewriting rules which can generate Context-Free languages). The role of the grammar, in this view, is to recursively enumerate the grammatical strings (and *only* these) of a language (Chomsky & Miller, 1963: 283; see Langendoen & Postal, 1984 for critical discussion), where a *language* is

> *a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements.* (Chomsky, 1957: 13)

The set of non-isomorphic L-trees generated by a grammar *Gr* (i.e., the full set of *structural descriptions* recursively enumerated by *Gr*) defines the *strong generative capacity* of *Gr* (Chomsky, 1965: 60). Frameworks that operate with rules of inference applying from axioms and producing sets of strings as outputs (including all versions of Phrase Structure Grammars) are called 'proof-theoretic' or 'generative-enumerative' (Postal, 2010; Pullum & Scholtz, 2001 respectively). The application of rules given an input produces a sequence of intermediate steps, from the initial symbol to terminal strings. A *derivation* in *Gr* is the ordered sequence of strings obtained by applying the rules of *Gr*, of the form A → B, for A, B members of the language's alphabet, stepwise: these rules rewrite the symbols on the left-hand side of the rule as the symbols on the right-hand side sequentially (i.e., one symbol at a time, from left to right; this is usally referred to as a *traffic convention*). A derivation ends when a string of terminal symbols is produced. Originally, a *derivation* of a string S was defined as

> *a sequence* $D = (S_1,. . ,S_t)$ *of strings, where* $S_1 \in \Sigma$ [Σ being the set of initial –possibly unary-strings of the grammar] *and for each* $i < t$, S_{i+1} *follows from* S_i (Chomsky, 1956: 117)

For example:

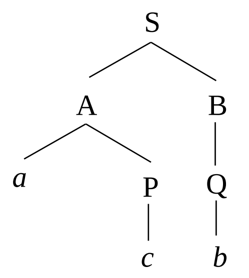
1) Alphabet: {S, A, B, P, Q, *a*, *b*, *c*}

   Rules: S → A, B
   A → *a*, P
   P → *c*
   B → Q
   Q → *b*

   Derivation: S
   AB
   *a*PB
   *ac*B
   *ac*Q
   acb

Now, given a derivation like the one above (which exemplifies 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 binary relation between strings *rewrites as* (or, conversely, *follows from*) becomes the graph-theoretic two-place relation *is the mother of*, or simply *dominates* (see Zwicky & Isard, 1963 and McCawley, 1968: 245 for discussion and additional details). The L-tree corresponding to the derivation in (1) can be diagrammed as in (2):

2)



The object in (2) satisfies the basic conditions to be considered a diagram of a mathematical construct known as a *graph*. 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 joining nodes in a structural description (with the additional assumption, at this point in the history of generative grammar covert, that any node in such a path must be visited *only once*) is a staple of the theory of phrase structure in transformational generative grammar; this requirement has been adopted in non-transformational frameworks almost by inertia. A crucial structural relation between nodes in a tree in transformational and non-transformational models alike –so long as they are based on Phrase Structure Grammars- is so-called *command*, originally defined by Langacker (1969) in the context of a discussion about the transformation *pronominalisation* (to which we will return below):

> *A node A commands a node B if (1) neither A nor B dominates the other; and (2) the S node that most immediately dominates A also dominates B* (Langacker, 1969: 167)

In (2), the set of *command* relations is {(A, B), (B, A), (*a*, P), (P, *a*), (*a*, *c*), (*c*, *a*), (A, Q), (Q, A), (A, *b*), (*b*, A)}. *Command* was later reformulated (and renamed) in Reinhart (1976) in terms which are still widely used today:

> *Node A c(onstituent)-commands node B if neither A nor B dominates the other and the first branching node which dominates A dominates B* (Reinhart, 1976: 32).

Note that the weaker condition imposed by Langacker pertaining to the presence of an *S* node has been strengthened to *the first branching node*: the determination of c-command relations can be now done more locally. Furthermore, in strictly binary-branching phrase markers where each generation is of the form {terminal, nonterminal} *c-command* relations are *asymmetric* in the sense that A *asymmetrically c-commands* B *iff* A c-commands B but B does not c-command A. The relation between *c-command* and linearisation, as well as the role of *c-command* in defining crucial syntactic and semantic relations (scope, binding) makes *c-command* a central relation in phrase structure grammars (in particular, MGG). However, conditions derived from *c-command* impose restrictions on the format of allowable structural descriptions, which we will argue are too restrictive.

Formal grammars of the type 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 pure phrase structure rules for generating structural descriptions for all and only the grammatical strings in a given natural language (Chomsky, 1956, 1959; Postal, 1964, among others). The representation of discontinuity was often referred to as one of these limitations (see, e.g., Brame, 1978: Chapter 1), but more generally the original argument for transformations pertained to rules whose formulation required reference to the 'derivational history' of a string (Chomsky, 1957) –thus more than just the last line of rewriting, which defines the relation *follows from*- and the conjunction of well-formed strings in a language L to form another, well-formed string ('generalised transformations', see Chomsky, 1955: 480-481). To address the shortcomings of pure context-free PSRs, Harris (1957) 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 (or strings onto strings; a 'symbol' is a unary string in this context), but rather *L-trees onto L-trees* (in what follows we will omit the *L*- qualification when talking about trees, presupposing it). The characterisation in Lees (1976) is quite explicit:

> *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 application of 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, …, X_n). Then,

> *If the structural index of a transformation has n terms, a_l, 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 string generated only by phrase structure rules (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 tree. In this specific sense (but see below), transformations can be thought of as *homomorphisms*: they are mappings which preserve (topological) relations within selected structure (see Hefferon, 2014: 181 for a more formal definition of *homomorphism*). However, we need to make further considerations, because there are second-order conditions over phrase structure building and mapping which will require a reinterpretation of what 'preserving relations' actually means in the context of transformational generative grammar. The generative power of transformations was greatly increased with the introduction and definition of *variables*, which can range over strings of arbitrary length and therefore capture 'unbounded' dependencies (Ross, 1967). Consider, for example, the formulation of the transformation that maps a declarative sentence into a polar interrogative one:

3) Interrogative:

S[tructural]D[escription]: $\begin{Bmatrix} \text{(a) } NP, C, VP \\ \text{(b) } NP, C + M, X \\ \text{(c) } NP, C + have, X \\ \text{(d) } NP, C\ +\ be, X \end{Bmatrix}$

S[tructural]C[hange]: $X_1 – X_2 – X_3 \rightarrow X_2 – X_1 – X_3$

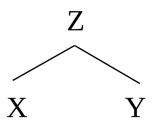
[where C = Pres/Past and agreement; and M = modal auxiliary]

Chomsky (1964: 227)

Derivations in transformational generative grammar instantiate a more general Immediate Constituent view of natural language structure (Wells, 1947; see Schmerling, 1983b for discussion) and the even more general view of a natural language as a recursively-enumerable set of strings (see Langendoen & Postal, 1984; Pullum & Scholtz, 2001; Pullum, 2007 for critical discussion). One of the roles of transformations (perhaps their most notable one) is precisely to impose underlying continuity on superficially 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 to at least two: one in which grammatical functions and thematic roles are assigned and which constitutes the *input* for the transformation (Deep Structure / D-Structure, depending on the model) and one enriched with displaced constituents and indexing mechanisms, the *output* of the transformation (Surface Structure / S-Structure). 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); later on, for intra-theoretical reasons, *traces* were replaced by *copies* (Chomsky, 1995: 202)[1].

It is crucial to note that because transformational rules (as conceptualised in the Standard Theory, Government and Binding, and Minimalist frameworks; see Chomsky, 1973, 1981, 1995 respectively) map (possibly kernel, possibly derived) trees to (derived) trees, alter the configurational relations between nodes in these trees, changing the connections between nodes and generate new local configurations. In this sense, then, transformations are *not* homomorphic mappings: they do *not* preserve relations within selected structure; rather, they *create new relations* between syntactic objects (including the introduction of new objects), deleting old ones. Let us illustrate this point in some detail. Imagine that we have a phrase marker in which objects X and Y are in a local relation, as represented in (4):

4)



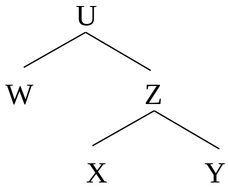
Now suppose that there is some relation *R* between X and Y: for instance, say X theta-marks Y in Z (where Z is the *label* of the object {X, Y}). That relation needs to be maintained throughout the

[1] Strictly speaking, *trace theory* and *copy theory* are not always very 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)* … NP_j … NP_i …
> | |
> *A* *B*
>
> *(31)* … NP_i … NP_i …
> | |
> *B* *e*
>
> *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*.

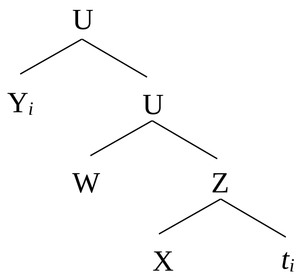
derivation, or reconstructed at the level of semantic interpretation if it is disrupted by a reordering or deletion rule. We have seen some problems with the latter option, so we would like to give some general strategies for exploring the former. Let us now introduce a further element in the derivation, call it W, which requires a local relation with Y in order to satisfy some formal requirement (which one in particular is not relevant for the present argument). W is external to {Z, {X, Y}}, following a monotonically cumulative approach to derivational dynamics which Chomsky (1995: 190) encodes in the so-called ‘Extension Condition’: the introduction of new elements always targets the *root* of the tree. In (4), the root is Z, which means that the operation that introduces W in the derivation targets Z and forms a new syntactic object {W, {Z}}, labelled U. This is diagrammed in (5):

5)

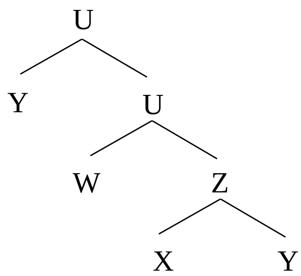


But, what happens if a local configuration is required between W and Y (as could happen if, for instance, Y satisfies a criterial feature on W), and the relevant relation *cannot* hold if X intervenes (i.e., if the scope domain of X includes Y but excludes W; in other words, if we cannot define a *c-command* relation between W and Y that excludes X)? The grammar must somehow create a configuration in which there is a syntactic object which contains W and Y, but not X. Transformational generative grammar deals with these situations by *reordering* syntactic objects, by means of the application of rules which map trees onto trees. This *reordering*, which yields non-adjacent dependencies between syntactic objects, is usually referred to as *displacement*, and it can be implemented in a theory of the grammar in different ways. A *displacement-as-movement* approach can either (a) *move* Y to a higher position in the (feature) checking domain of W (thus extending U), outside the scope of X, leaving a co-indexed trace behind (as is the case in so-called *trace theory*); or (b) *copy* Y and *re-introduce* the copy of Y in the derivation where appropriate (the so-called *Copy Theory of Movement*, or *Copy+Re-Merge theory*; Chomsky, 2000; Uriagereka, 2002; Nunes, 2004 and much related work). These two options are diagrammed below:

6) (a)



(b)



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} (see **Appendix A** for additional discussion). 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: this is a set of pairs (LI, *i*), where LI is a Lexical Item and *i* is an integer (known here as an *index*) indicating how many times LI is used in the computation. If movement is implemented by means of traces and co-indexing, does the numeration contain *traces*, then? No. Chomsky (1995, 2000) correctly claims that the inclusion of *traces* in a derivation violates the so-called *Inclusiveness 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 Inclusiveness 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 copies[2]. If an element of arbitrary complexity is Copied and then (internally) 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 ready beforehand, crucially including copies), it is therefore treated as a whole new element: after all, until we get to the interface levels (where phonological and semantic interpretation takes place), we have no possibility of establishing a referential connection with an object already introduced in the derivation, except under special stipulations which depart from the simplest scenario and thus require independent theoretical and –more importantly- empirical 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 for the interface levels would be already predicted (somehow) in the NUM (incidentally, making the syntactic component also superfluous and requiring massive amounts of *look ahead*: the computation needs to be aware of derivational steps that have not been developed yet). 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. Let us clarify this. 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.; Epstein & Seely, 2002). This discharging requires a local relation between at least two categories sharing the feature that is relevant to the case at hand (valued in one instance, unvalued in the other), 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 specifications if operations are indeed driven by the necessity to valuate/check features.

[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' are not addressed in syntactic works in MGG, although appeal to copies and re-merges is ubiquitous.

Identity between occurrences of syntactic objects in distinct contexts generated by transformations is not possible in a system that admits feature checking relations as the triggers for transformations, since the feature matrix of an element varies as the derivation unfolds and copies are (Internally) merged in places in which they can valuate 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) $\alpha = \{i\text{-F}_1, i\text{-F}_2, u\text{-F}_3, u\text{-F}_4\}$

$\beta = \{i\text{-F}_1, i\text{-F}_2, u\text{-F}_3, u\text{-F}_4\}$

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*).

All in all, transformational grammar followed closely on the structuralist conception of clause structure in terms of immediate constituency, phrase structure, and dominance configurations. 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*'. The present work essentially agrees with Perlmutter's insight, and explores an alternative to constituency-based frameworks (including all frameworks based on phrase-structure grammars), both transformational and non-transformational. Some empirical 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; 2016a; 2018; Krivochen & Schmerling, 2016a; Bravo et al., 2015, among others), and we will not repeat that case here. 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 minimising the primitives of the theory. We begin by considering the following passage 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 to the theory of grammar that we are going to develop in this work, and we can go even further if we give up the requirement of structures being *trees* while keeping strict *order* as a basic admissibility condition for structural descriptions (the specifics of which we will come back to below). In this sense, the approach sketched in the passage from McCawley's quotation is an excellent transition between traditional PSGs and the strongly representational, *derivation-less* approach that we will develop in this work.

1.2 Model-theoretic vs. Proof-theoretic syntax

Let us draw a direct comparison between the conception of syntax that we will develop in this work and PSG-based MGG to start making things explicit. In the view that we explore here, the whole point of 'syntax' is not to not 'form sets' (as Chomsky 1995: 243 suggests; see also Collins, 2017 and much related work; in the view expressed in these works, the 'syntax' is a discrete recursive combinatoric system) or recursively enumerate expressions (for discussion, see Langendoen & Postal, 1984: Chapter 1); rather, the 'syntax' establishes connections between nodes and makes the most out of the smallest amount of nodes by maximising connectivity between existing nodes rather than keeping the relations that each node establishes at a minimum (a single mother and two daugthers) by means of introducing new nodes and assuming an indexing mechanism. In this way, a number of second-order conditions over structure building and mapping which do not seem to arise from empirical necessity are given up (including projection-based endocentricity, binarity, the Single Mother Condition, among others we will analyse here). Similarly, the whole point of 'doing grammar' in the present approach is to fully specify the connections between nodes in a representation (what we will refer to as a 'map' of that representation) 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, and following the distinction made in Pullum & Scholz (2001) and subsequent works, the theory we will sketch here instantiates a *model theoretic* approach to syntax rather than a *proof theoretic* one, at least in general lines. We can unpack this in some detail:

- *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 *S satisfies I*). 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', in the sense of 'what an expression must look like in order to satisfy the grammar' (Pullum, 2007: 1-2).

- *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*.

Pullum (2007: 2) summarises things rather neatly:

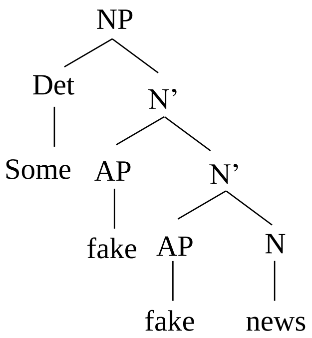
> *Grammar, on the MTS* [Model Theoretic Syntax] *view, is about what structure expressions have. It is not about devising a sequence of operations that would permits the construction of the entire set of all and only those structures that are grammatical.*

Our view incorporates elements from both meta-theories maintaining logical consistenct: we formulate rules as statements about the structure of 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 useful concepts from both metatheories. From proof-theoretic frameworks we will adopt some concepts, like the idea of 'syntax' being computational. This is an idea common to a number of theories, including the Standard Theory, (Combinatory and Classical) Categorial Grammar, and LFG (note that 'computational' need not be interpreted necessarily as 'discrete recursive combinatorics'; a more general definition of 'computational' as 'algorithmic' is closer to what we need). We will also build on insights from model-theoretic syntactic frameworks (Arc-Pair Grammar, Metagraph Grammar, Relational Grammar, and Dependency Grammar), like the idea of the 'grammar' being a set of admissibility conditions on graphs (which describe the structure of expressions of the language, in our case, English). 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 a model-theoretic framework (Rogers, 1998) reveals that we can indeed take advantage of the best of each metatheory without giving up consistency. 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: the operation Merge (in all of its incarnations) conserves terminal distinctness and unambiguous command paths throughout a derivation. 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 as 'recursively enumerative') to a specific level in the Chomsky Hierarchy assigns too much structure to substrings (as is the case with CFGs and simple iterative patterns, non-scopal adjunction, and certain kinds of coordination, see Lasnik, 2011; Lasnik & Uriagereka, 2012; Krivochen, 2015a, 2016b); at times, that restriction falls short (as is the case with CFGs and crossing dependencies, see Joshi, 1985 and much related work). Let us briefly illustrate the kind of problem that arises when a uniform *a priori* template for structural descriptions is adopted (such as X-bar theory). Consider the following string:

8) Some fake fake news

A context-free PSG of the kind we have been discussing can only assign a single structural description to (8), which goes along the lines of (9):

9)



Note that, if *c-command* relations are mapped to *scope* relations at the level of Logical Form, such that the *scope* of a node A is the set of nodes that A c-commands at LF (Ladusaw, 1980; May, 1985), then because the higher AP asymmetrically c-commands N' (which in turn contains a predicative structure {AP, N}) the only possible interpretation for (8) allowed by the theory is, roughly, (10):

10) Some news which is fake as fake news (i.e., truthful news)

But that is not the only interpretation for (8): a non-scopal interpretation, where there is no hierarchical relation between the adjectives, is also available:

11) Some news which sounds very fake (i.e., iteration intensifies the meaning of 'fake')

The meaning of intensive reduplication is reminiscent of the "rhetorical accent" identified by Stanley Newman in his classic work on English stress (Newman, 1946; see also Schmerling, 2018b). The same reading could have been obtained by means of vowel lengthening:

12) Some faaaaaake news ('Some fa·ke news', in Newman's notation)

It is clear that the structural representation in (9) cannot be an *adequate* structural description for (8), insofar as it is unable to account for both interpretations. There is no scope between both instances of 'fake' in the interpretation (11), which means that there cannot be a c-command relation between them. The structural description must then be *flat* (i.e., finite-state), but only *locally so*: we still want to keep a scope relation between the quantifier and the noun, which translates into the requirement that the quantifier c-commands the noun. The problem we are facing can be formulated as follows: a Context-Free structural description is locally adequate, but not globally so. It *necessarily* assigns too complex a structure for a substring of (8), the iteration of adjectives. Chomsky's solution to the 'too much structure' conundrum (also noted in Chomsky and Miller 1963; more recently revived by Lasnik, 2011; Lasnik & Uriagereka, 2012, forthcoming; Schmerling, 2018b; Krivochen, forthcoming) was of course to go beyond phrase structure rules and incorporate a transformational component into grammars. There has been nothing in the MGG metatheory that leads us to prefer a completely underspecified phrase structure building engine over a transformational model in which $\Sigma \rightarrow F$ rules are supplemented with mapping operations from trees to trees (Lees, 1976).

That empirical inadequacy of *a priori* uniform 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 uniformly monotonic function-based nature of syntactic computation that is at the heart of generative grammar, has been explored in past works (particularly Krivochen, 2015a, 2016a, b, 2018). 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 'passivisation', 'raising' or such terms: we will take *transformations* to have a *descriptive* rather than an *explanatory* value, our goal indeed being to *describe* the phenomena under discussion here. We use rules that make reference to a particular segmentation of an English sequence and a schematisation of its structure and how to operate over that structure to generate, in the formal sense, a sequence that is also grammatical in English, without making claims about psychological reality or universality. 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* (Ringen, 1972; Koutsoudas, 1972) or indeed of time (either 'real' –processing- or proof-theoretic –counted as steps in a proof-); there are also *no derivations*. This does not mean, however, that rules and constraints do not interact; as a matter of fact they do, but in a different way from the way they interact in classical transformational generative grammar:

> *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, which builds on the pioneering work of Stanley (1967), McCawley (1968), Zwicky & Isard (1967), *the grammar is a finite set of admissibility conditions over graphs*[3], which are

[3] It is interesting to note that our view also contrasts with that of Lasnik & Kuppin (1977), who define a *reduced phrase marker* by means of a set of admissibility conditions *over* (sets of) *strings* (rather than *graphs*, as in the

structural descriptions of natural language sentences (a.k.a., ‘expressions’). This view entails a departure from the classical approach to PSRs where these are mappings of strings onto strings, and an alignment with the core insight that PSRs can be interpreted…

> *as node admissibility conditions. A node labelled S in a tree is admitted by the rule* [S → NP, VP] *if and only if that node immediately and exhaustively dominates two nodes, the left one labelled NP and the right one labelled VP. A tree is analysed by the grammar if and only if every non-terminal node is admitted by a rule of the grammar. Under this interpretation, then, phrase structure rules are well-formedness conditions on trees* (Gazdar, 1982: 137)

We will see that there is much to be said about the properties of *trees*, and that even within the set of theories that may think of PSRs as node admissibility conditions there is considerable variation, inasmuch as there are several non-mutually reducible or intertranslatable theories whose only common feature is this non-traditional reading of PSRs. At this point, we need to introduce the mathematical notion of *graph* and how it interacts with earlier conceptions of phrase structure (of the kind we have reviewed so far), the empirical limitations of which motivate the present revision of the theory of the grammar.

1.3 On graphs and phrase markers: first- and second-order conditions on structural representations

First and foremost, we need to define *graph*[4]. A graph 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. For example, a vertex v with degree 2 and neighbourhood set 2(v) has two edges connected to it, and two vertices which are adjacent. There is no necessary correspondence between *degree* and *neighbourhood set*, since a vertex v_1 may be connected to v_2 by two distinct edges (in which case the neighbourhood set of v_1 would be 1 and its degree, 2). Here we will indifferently refer to vertices as ‘vertices’ or ‘nodes’. Now, let v_1 and v_2 be two (not necessarily distinct) vertices in G: a v_1-v_2 *walk* in G is a finite ordered alternating sequence of vertices and edges that begins in v_1 and ends in v_2. Further conditions can be imposed over walks. For example, we might require that

- each edge be walked on only once, or
- that each vertex be visited only once, or
- that both of these conditions hold, or

---

present view), thus the importance of the incorporation of *precedence* relations between monostrings alongside *domination* (Lasnik & Kuppin, 1977: 176-177; see also Lasnik & Uriagereka, forthcoming). Thus, they define a *reduced phrase marker* as follows:

> *℘ is an RPM if there exist A and z such that*
> *A ∈ ℘ and z ∈ ℘; and if {ψ, φ} ⊆ ℘,* [where φ = *x*Az; A a nonterminal, and ψ a string of terminal symbols]
> *either ψ* dominates *φ in ℘* [where *φ dominates* ψ in ℘ if ψ= *xχz, χ* ≠ Ø, *χ* ≠ A].
> *or φ* dominates *ψ in ℘*
> *or ψ* precedes *φ in ℘*
> *or φ* precedes *ψ in ℘* (Lasnik & Kuppin, 1977: 177)

[4] Introductions to Graph Theory in which some of the concepts we use here are defined in more detail, are van Steen (2010), Wilson (1996), Ore (1990), and Gould (1988).

- that neither of them hold

It is important to bear in mind that choosing between these alternatives has empirical consequences for the analysis of syntactic connections in natural language structural descriptions, so the choice must be made carefully.

We need to give some more definitions, which will be essential when considering relations between subgraphs in structural descriptions. A graph H is a *subgraph* of G iff V(H) ⊂ V(G) and E(H) ⊂ 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 Ore, 1990: 7; a related usage is that in Gould, 1998: 10, who uses the term *strongly connected* instead of *complete*). 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 are advancing 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 in a natural language. Locally, the graphs we will consider are *connected*: within single-rooted subgraphs, there is at least one walk from any vertex to any other vertex (we will see that these walks, in our model, need to be defined, technically, as *trails*. The motivation for this, as will become clear in the following sections, is eminently empirical).

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

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

(13b) is the option chosen by most works on generative grammar since GB[5], as well as Lexical Functional Grammar (LFG), Head-drive Phrase Structure Grammar (HPSG), and (locally) Arc Pair

[5] Although we do have to note that Chomsky & Miller (1963) and Katz & Postal (1964) (see also Chomsky, 1955) assume that Generalised 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 (Kayne, 1984, 1994) and more recently also labelling (Chomsky, 2013), are thus much older than X-bar theory (Chomsky, 1970b; Jackendoff, 1977; Stowell, 1981).

Grammar (APG). The base component of a generative grammar[6] 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 for discussion), 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. In such a system, the following conditions thus hold:

a) Every node has only one mother (the Single Mother Condition)
b) Every branching node has at most two daughters (the binarity axiom in X-bar theory)

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). In purely formal terms, the lack of a mother node is the definition of 'root'; alternatively, we can think of the root in terms of rewriting rules over strings, as a symbol that only occurs at the left-hand side of the transition function 'rewrite'.

In transformational generative grammar, labelled nodes that correspond to full clauses (*Sentence* in the Standard Theory and its developments, *Complementiser Phrase* CP after Chomsky, 1986) are sometimes also referred to as 'root' nodes, despite the fact that they may be dominated by some other node (as in the case of embedded clauses; but see Emonds, 1970: 5, ff. who claims that non-finite embedded clauses are VPs, not Ss), derivationally, these are indeed roots of their respective sub-trees. Emonds' (1970: 8) is a useful definition to bear in mind, quite representative of its time:

> *a root will mean either the highest S in a tree, an S immediately dominated by the highest S, or the reported S in direct discourse.*

Consider the requirements in (a) and (b) above: (a) the Single Mother Condition and (b) the binarity axiom. Given these *a priori* conditions (we call them *a priori* since they would be given by Universal Grammar, see Kayne, 1994; Chomsky, 1995 and much subsequent work), phrase markers grow introducing extra nodes in the form of non-terminals 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 in Kayne (1984), where explicit conditions pertaining to *unambiguous paths* in c-command relations between terminals and non-terminals 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 ($A_0 \ldots A_i$, A_{i+1}, $\ldots A_n$) 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 walk in T must be a sequence of *distinct* nodes; there are no loops and the walk does not double back on itself. In principle, this is consistent with both 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. But since Kayne uses the term

[6] 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.

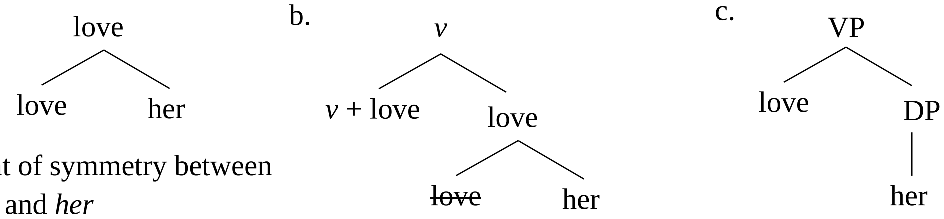
‘path’, we will use it as well when referring to his approach; we will see that there are indeed reasons to claim that walks in MGG trees must be *paths* rather than *trails*. Condition (b) means that a path in Kayne’s sense is a sequence of adjacent nodes. Let us now proceed to the definition of ‘unambiguous path’

> *An unambiguous path in T is a path* $P = (A_0, \ldots A_i, A_{i+1}, \ldots A_n)$ *such that:*
>
> $\forall i\ 0 \leq i < n$
>
> *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 (they have a mother, and no daughters), and nonterminals are nodes with degree 3 (having two daughters and a mother), except for the root which is of degree 2 (as seen above). Moreover, in the more recent (1994) version of the theory and under Bare Phrase Structure assumptions (Chomsky, 1994), one of the two most deeply embedded nodes in a tree T must be an empty category: either a trace product of *Move-α* or a base-generated empty node *e*; in any case, it must be a node that does not receive morphophonological exponent and thus is not linearised with respect to the rest of nodes in the tree under consideration. The alternative is to add structure; at least a non-terminal node dominating the most embedded terminal (as in (14c)) or a projection that can be the target of movement of one of the nodes involved in the point of symmetry (as in 14 b)), in that way the point of symmetry is dissolved (Kayne, 1994: 9-10; Johnson, forthcoming). We can exemplify these situations in (14):

14) a. b. c.



a. Point of symmetry between *love* and *her*

b. Asymmetric c-command: <*v*+ love, her>

c. Asymmetric c-command: <love, her>

These conditions over structural representations arise from the purported necessity of phrase markers to comply with the so-called Linear Correspondence Axiom (LCA): *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). The LCA is assumed to be the UG principle that establishes how to linearise syntactic structure for externalisation purposes (see also Moro, 2000 for a more radical version of the theory). Here, *d* is a syntactic relation known as asymmetric *c-command*, applying to the set of nonterminal nodes in a tree: *a* asymmetrically c-commands *b iff a* c-commands *b* but *b* does not c-command *a*.

As suggested by Uriagereka (2002, 2012), the LCA can be interpreted in two ways: a 'weak' interpretation and a 'strong' interpretation. 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*. Note that if the most deeply embedded node was not phonologically empty, because of the axioms of binarity and the SMC, there would be a structural configuration with two nodes in a *mutual c-command relation*, which cannot be linearised by the LCA: this is known as a *symmetry point*, and must be avoided at all costs (for example, in Moro's 2000 view, displacement is motivated in natural languages by the need to eliminate symmetry points by moving one of the nodes). The resulting model of phrase structure is one in which *each and every object* (read: each and every node in a tree) *is of one of the following three types*[7]*:*

a. A terminal node
b. A nonterminal node dominating two terminals (one of which must then be a *trace* in the sense of Fiengo, 1977[8])
c. A nonterminal node dominating a terminal and a nonterminal

Option (b) is explicitly rejected in Kayne (2018: 11): '*Merge never constructs a set consisting of two syntactic objects each of which is a phrase*': a head is merged separately with its complement and then with its specifier (i.e., specifiers would never merge with phrases), for entirely intra-theoretical reasons. It is worth noting that the LCA is not a privative feature of Move- or Copy-based Minimalism: for example, Johnson (2016, forthcoming) advances a Re-Merge approach which yields diagrams featuring multidominance where the LCA plays a prominent role: this is possible because the mechanism is still 'generative', based on recursive combinatorics and strictly binary-branching trees. It is thus possible to break down applications of the LCA to *pairs* of syntactic objects (a classical 'divide and conquer' computational approach): this move allows Johnson to keep the LCA as a linearisation mechanism and have syntactic objects remerge in the derivation (see Stroik & Putnam, 2013 for extensive discussion of potential problems with the Re-Merge view of displacement). The extent to which 'multidominance' in Minimalist approaches like Johnson's is simply an artefact of diagrams is not clear: these proposals are still not based on *graph theory*, but simply extend PSGs; the formalisations follow suit. It is interesting to note that Johnson's multidominance approach defines linearisation in terms of *paths* (his 2016: 14 proposal combines elements of Chomsky's 1995 and Kayne's 1984, 1994, 2018 versions of phrase structure: all usual X-bar constructs, like *heads* and *projections* are still present), despite the fact that graph-theoretically multidominance yields *trails*. It

[7] As David Medeiros (p.c.) has 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 related work even going back to conditions over generalised transformations in Chomsky, 1955; see also Collins & Stabler, 2016, *Definition 13*). Boeckx (2012: 56) mentions a 'plausible' Anti-Identity output condition *[XX], which yields the same results as forcing Merge to apply to two distinct terminals. However, see Lasnik & Uriagereka (forthcoming) for an LCA-inspired cyclic proposal in which categories can *self-Merge*.

[8] 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)

is thus difficult to evaluate such a proposal as different in some significant way from one based on indexing (*trace theory*) or identification (*copy theory*).

The theory of phrase structure that follows from the requirement that (non-root) nonterminal nodes be vertices of degree 3 and terminals be nodes of degree 1, plus the additional axiom that walks in phrase markers define *paths* (in the graph-theoretical sense) yields an *a priori* uniform template for structural descriptions. We have provided empirical arguments against MGG's rigid conception of phrase structure in previous works, based on *mixed computational dependencies* being established within local derivational units in natural languages in previous works (Krivochen, 2015a, 2016a, 2018; Bravo et al., 2015; Krivochen & Schmerling, 2016a, 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 a substring the simplest structural description that captures and represents the formal and semantic relations between syntactic objects.* (Krivochen & Schmerling, 2016a: 36)

The mixed computation proposal comes as an answer to a computational conundrum: structural descriptions must be *minimally appropriate*, that is, they must assign no more structure to substrings in L than strictly needed. 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: that allowable in Context-Free languages, nothing more and –crucially- nothing less. Because of this, it has long been recognised 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 (as pointed out in Joshi, 1985; Kroch & Joshi, 1985), for there are strings in natural languages the structural descriptions for which display crossing dependencies that cannot be captured by means of a context-free PSG (we will see some examples in **Section 4**). 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, 2018; Saddy & Krivochen, 2017; Krivochen, 2016a, 2018):

> […] *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 (e.g., McCawley, 1968, 1982; Levine, 1985; Lasnik & Uriagereka, 2012; *inter alios*). To begin with, we are not interested in psychological aspects (processing, memory, parsing, etc.) or 'biolinguistic' concerns (language as an adaptive 'organism', innateness, etc.): the present work is a purely abstract study of grammatical structure. More technically, and in contrast to transformationally-enriched phrase structure grammars, we eliminate empty nodes (including traces), indices, restrictions on the degree of vertices, and the requirement of graphs *minimising* connectivity (*as per* the SMC and strict binarity requirements), thus proposing a theory aligned with tendency (13a), *maximise connectivity*. We formulated (13a) and (13b) 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 (and it is likely that different languages exploit these tendencies differently; we will not deal with these issues in this work but they are definitely worth pursuing). This view of grammar as a dynamical oscillatory process yields graphs that can be evaluated as maximising or minimising connectivity *only within local domains*. That means that a model that embraces *n*-ary branching, discontinuity and multidominance does *not* necessarily need to reject binarity and monotonicity altogether: these are properties of phrase markers *at a local level*, whereas mixed computation is a *global* property of structural descriptions. We will explore the consequences of adopting a mixed computational view in this work, particularly in **Section 11**.

Formalisms that position themselves at the (13b) end of the spectrum (*maximise unambiguous paths*), which include transformational generative grammar as well as LFG, HPSG and (C)CG, assume -explicitly or implicitly- what Sampson (1975: 1) calls the 'Single Mother Condition', already familiar to us:

> *D is a set of NODES, α is a function from D into a vocabulary of symbols (if α (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* $\delta(d) = e_1, e_2, \ldots, e_n$, $\delta(d') = e'_1, e'_2, \ldots e'_n$, *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).* (our highlighting)

We saw how the requirement of unambiguous paths leads to the SMC in transformational generative grammar. For LFG, Bresnan et al. (2016: 46) claim that c-structures (which are generated by means of a CFG) also obey the SMC; GPSG and HPSG are also modelled upon CFGs, thus the SMC also applies to their constituent structure representations (see Gazdar, 1982; Pollard & Sag, 1987: 55, ff.). As for CGs, Partee (1974), among others, argues that traditional (i.e., non-combinatory) CGs have the strong generative power of CFGs: it is thus 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). Let us briefly exemplify this point. In a CG, if an expression of the language is assigned to the category X/Y, this means that X/Y must be concatenated with an expression of category Y to get an expression of category X. For instance, we can define the category of 'verb phrases' as the set of expressions that must combine with expressions of category 'noun phrase' to form expressions of category 'sentence': this can be translated into CG notation as the assignment of expressions that we call 'verb phrases' to the category S/NP. The general *context-free* format of Ajdukiewicz-style CG rules is translated into Chomsky Normal Form as follows:

> $c \rightarrow (c/c_1 \ldots c_n) + c_1 + \ldots + c_n$ (Lewis, 1972)
> (e.g., S → NP + S/NP; Partee, 1975: 214)

Things may not be all that simple, though: not all variants of CG have the same generative power; crucially, Montague grammar goes beyond strict CF power if we consider rules of quantification (see Hamblin, 1973: 43; also Van Benthem, 1988 and Buszkowski, 1988 for detailed discussion about the generative power of different variants of CG, including Ajdukiewicz-style and Lambek calculi). In CGs the SMC is respected (because trees are, strictly speaking, proofs that an expression belongs to a category in the algebra, thus, the root node in a CG analysis tree defines the category to which a complete expression of the language belongs; see Schmerling, 2018a for some details), although diagrams ('analysis trees', which diagram the proof that an expression belongs to a certain category) have a much less significant role in CGs than in PSGs, if any at all. There is nothing in Ajdukiewicz-style CG that is comparable to a PS tree.

A condition similar to the SMC was initially assumed by McCawley (1968) in his formalisation of the base component of a transformational generative grammar. The interpretation of PSGs in McCawley's work, which was hugely developed in HPSG (e.g., Gazdar, 1981; see also Pullum, 2019 for recent discussion) is that of well-formedness conditions over graphs, such that a rewrite rule like A → Bc is not seen as a mapping between strings #A# and #B⁀c# (for this latter interpretation, see Chomsky, 1956, 1959) but rather as a constraint over what a well-formed graph containing nodes A, B, and c looks like: roughly, *a tree T containing node A is a well-formed tree iff A immediately dominates B and c in T*. Let *x* and *y* be nodes, *ρ* be a dominance relation between nodes, and *λ* be a precedence relation. Then:

> *if xρy and x'ρy, then x = x'*
>
> *for any two nodes x and y, if x ≠ y, then either xρ*y or yρ*x or xλy or yλx* (see also Zwicky & Isard, 1963: 1 A.4; Wall, 1972: 149 for equivalent conditions)

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 (13a) type. The expansion of phrase structure grammars in Peters and Ritchie (1981), called 'Phrase Linking Grammar', also allows for multidominance *locally*, and is thus closer to (13a) than MGG and LFG. Other important grammatical frameworks, like Arc-Pair Grammar (Johnson & Postal, 1980) and Metagraph Grammar (Postal, 2010: 20) also reject the SMC; these two are particularly important because they constitute fully explicit models with a strong basis on graph theory and are based on model-theory in the sense of Pullum & Scholz (2001) (see, e.g., Postal, 2010: 4). We share some of their arguments and conclusions, but not all; importantly, our proposal is not equivalent to theirs.

As we have already observed, since LSLT (Chomsky, 1955), generative 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. The transformational component was incorporated as a way to deal with certain inadequacies of phrase structure grammars, including processes that make reference to relations between separate well-formed sequences (as in the case of conjunction; see Chomsky, 1957: 36, ff.) and the attachment of inflectional affixes to verbal stems, among many others. Transformations operate in different ways and are subjected to different kinds of constraints depending on what they do: given a sequence, they can reorder elements, add elements, delete elements, etc. (see **Section 13.2** for further discussion). It is important to note, in the consideration of transformations, that they were formulated within a framework in which *strings*, rather than *graphs*, played a fundamental role due to the background that many early generative grammarians had on *automata theory* and *formal language theory*, but not on *graph theory* (this point is made in McCawley, 1982: 92; we will deal with exceptions like Zwicky & Isard, 1963 and Morin & O'Malley, 1969 in some detail here). Here, as will become obvious below, we will take *transformations* to have an exclusively *descriptive* value. In this sense, we propose that *transformations*, i.e., mappings from *structural descriptions* to *structural changes* using traditional terms, can be divided in two types:

15) 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 (15a) also change *constituency* ('relation-changing transformations' in McCawley's terms, RCT henceforth); whereas transformations of type (15b) do not (which we will refer to as 'relation-preserving transformations', RPT henceforth, an obviously derivative term). McCawley (1982: 94) argues that consideration of empirical phenomena provides grounds for distinguishing

> *two essentially different types of transformation that hitherto have been classed together under the single name of movement transformations:* ***transformations that change syntactic relations*** *(not only "grammatical relations" such as "subject of' and "object of', but also relations such as what Pullum (1980)* [Syntactic Relations and Linguistic Universals. *Transactions of the Philological Society,* 78(1). 1-39] *calls "query of', which an item can manifest through its occurrence in some position of syntactic focus), and* ***transformations whose sole syntactic function is to change constituent order.*** (our highlighting)

The present investigation stems from the idea that these two kinds of processes need to be incorporated to the core of the theory of grammar and be appropriately distinguished formally.

In this work, we are *not* concerned with the linearisation of graphs[9], because 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 (15b)). Like Tesnière's *stemmas*, our graphs are intended to represent hierarchical relations only, and while we will introduce a notion of *strict ordering* as a condition over well-formed graphs, that is not to be understood in *left-to-right* linear terms. By 'constituency' –above- we simply mean 'grammatical relations' (but see **Section 4.2** for discussion about the role of grammatical relations in the present model): if grammatical relations are indeed defined in a pre-transformational structure, as in Chomsky (1965: 71), Lees & Klima's 'be in construction with' is equivalent to 'be GF of', where GF is a grammatical function or relation ('subject of', 'object of', etc.). However, there is no need to resort to base structures (which need to be reconstructed when linearity is disrupted by reordering transformations) if grammatical relations are in fact primitives and not derived from a primitive structural template. Let us clarify this point: 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. In contrast, frameworks like Relational Grammar (and its descendants Arc Pair Grammar and Metagraph Grammar; see Perlmutter & Postal, 1983a, b; Johnson & Postal, 1980; Postal, 2010 respectively) takes grammatical relations as primitives, and eliminates derivations as understood in [Σ, F] grammars from the theory (and with them, the need for triggers for derivational operations; that means no formal features, no Agree, no Merge, etc.). Relational Grammar and related formalisms do, however, allow for *strata* in its representations where elements can get *promoted* or *demoted* (changing their grammatical function), a mechanism we will not make use of. There is still the question of how to link roles with lexical items, and that is an essential question that we will return to below. For the time being, we need to provide some basic definitions, as a way to make our proposal formally explicit. Comparisons with alternative

---

[9] Strictly speaking, the translation of hierarchy into linear order can be taken care of by means of independent linear order statements (for instance, McCawley, 1968 introduces two independent basic relations, ρ 'dominates' and λ 'is to the left of'), which would not be any more or less stipulative than Kayne's (1994) LCA. We will see some examples of such statements in the discussion of Dependency Grammars, below.

conceptions of syntactic structure will be the topic of **Section 4**, although strictly speaking we will provide comparisons and contrasts at almost every empirical and theoretical turn.

2. Definitions and general considerations

In the context of this work, a *grammar* is a finite set of conditions specifying what a well-formed structural description for an expression of the language looks like: we are concerned with defining *models* for the language (Pullum & Scholz, 2001; Pullum, 2007). Thus, the conditions that the structural descriptions 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, …, *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, …, v_n). The subscripts v_x, v_y, v_z, … will be used to denote variables over vertices. We will use 'vertices' and 'nodes' indistinctly.
2. Edges specify the *directionality* of the connection: $e <v_1, v_2>$. Thus, $e <v_1, v_2> \neq e <v_2, v_1>$. If a graph contains a function *i* from E → V × 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 (see Saddy, 2018; Krivochen, 2016b for additional discussion about language as a dynamical system).

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 $e <v_1, v_2>$ (i.e., there is a directed edge *from* v_1 *to* v_2).

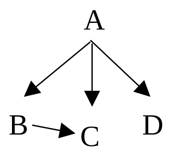
The reason why we restrict ρ to be a *binary* relation is strictly because it is defined in terms of the existence of an edge between nodes, and edges, by definition, link *two* nodes.

4. The ρ-domain of a vertex *v* is the set {v_1, v_2, …, 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:

16) $\tau(v_1, v_2, \ldots v_n)$ *iff* $\exists(v_x) \in X \mid v_x\rho(v_1, v_2, \ldots v_n)$

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 graph (where directed edges have been marked with arrows)

17)

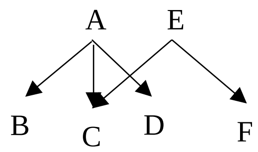


We have the following set of dominance relations (*ρ-set* henceforth)

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

And, since ρ(A, B), ρ(A, C), and ρ(A, D) hold, we say that B, C, and D are sisters because they share the mother node A, regardless of the fact that (B, C) ∈ ρ. In a slightly different case, illustrated in (18):

18)



Nodes 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 τ_A(B, C, D) is the set of nodes that are sisters because they share the mother node A.

Let X and Y be sub-graphs. Then,

6. $X \subset Y$ *iff* $\exists(v) \in Y \mid y\rho{*}X$
7. $\exists(v_0) \in X \mid \forall(v_x) \in X\ v_0\rho{*}v_x$ (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 (cf. Zwicky & Isard, 1963: 1), as a matter of fact, the second sample graph above has two roots (A and E), two nodes that dominate others but are not dominated themselves. (18) is an example of what Morin & O'Malley (1969: 182) call a *vine*: a directed, acyclic, labelled, rooted graph.[10] 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: in this work a *cycle* is, in purely configurational terms, a *single-rooted graph*. Complex structures, including multi-rooted graphs, are obtained by relating cycles: concretely, by conjoining transformations (in the sense of Fillmore, 1963) and identification of common nodes. We weaken a condition that has been around in syntactic theory for some decades now (see, e.g., Joshi, 1985; Rogers, 2003): it is not necessary that the linking vertex between graphs X and Y be the root of either, all we require is node identity between cycles. Some versions of Tree Adjoining Grammar (e.g., Sarkar & Joshi, 1997) assume that nodes in an elementary tree are assigned Gorn addresses (Gorn, 1967), which serve to identify corresponding expressions: if nodes on distinct

[10] We need to note that Morin and O'Malley'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).

elementary trees T and T' are assigned the same address, then a derived tree that contains T and T' will collapse those nodes into one. Our perspective is very close to Sarkar & Joshi's, although we will define our nodes in a slightly different way.

In this work, we will use *cycle* in the sense that it has in generative linguistics since the 1960s (as a local domain for purposes of syntactic operations); 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. A relation $\mathbb{R}$ between sub-graphs X, Y, …, *n* is *total* and *transitive*, but *not* antisymmetric (see also Ojeda, 1987: 258).
9. X is 'in construction with' Y if *either*
   (a) $\exists(Z) \mid Z\rho X \wedge Z\rho Y$, *or*
   (b) $X\tau Y$.
10. Let π be an ordering of G, such that π(G) is the (possibly unary) set of *trails* within G starting at the root and following the direction of edges.

There is a key word in **Definition 10**: *trail*. Above we defined a *walk* as an ordered alternating sequence of adjacent vertices and edges. We also said that the concept of *walk* is neutral with respect to whether vertices or edges can be visited more than once (in other words, whether a given vertex *v* can be ordered with respect to itself so that it transitively dominates itself): now we need to specify what kinds of relations we allow in our walks. Walks can be either *trails* or *paths*. 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, but *vertices* can be (Gould, 1988: 9; Wilson, 1996: 26; Van Steen, 2010: 37, 61). In Van Steen's words,

> *In a* ***closed walk****, $v_0 = v_k$, a* ***trail*** *is a walk in which all edges are distinct; a* ***path*** *is a trail in which also all vertices are distinct.* […] *A* ***directed trail*** *is a directed walk in which all arcs are distinct; a* ***directed path*** *is a directed trail in which all vertices are also distinct.*

Because we will make use of repeated visits to certain vertices, as a crucial aspect of our model which is empirically motivated, 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; Kural, 2005) 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 pertaining not only to the nature of syntactic structure (monotonicity, the SMC, movement, etc.), but also the architecture of the grammar (including properties like an autonomous generative syntax, interpretative semantics-morphophonology, the relations between these components proposed in the so-called *Y-model*; Chomsky, 1995; Hauser et al., 2002; Siloussar, 2014). 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. In the typology presented in Hockett (1954) and Schmerling (1983b), 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 projection or endocentricity (against a generalisation 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, multi-rooted representations, *and* indirect loops, both the NTC and the EC are violated (but not in an unrestricted manner; the notion of *cycle* we propose here restricts 'tampering' in an adequate manner, as we will show); the theory advanced here contrasts heavily with GB/MP-inspired MGG, including Minimalist versions of 'multidominance' theories of movement based on recursive binary combinatorics where only the SMC is rejected (e.g., Johnson's, 2016 *Re-Merge*; Citko's 2005 *Parallel Merge*, among others). We will see that there are some simplifications on representations assigned to natural language strings that follow directly from these violations.

2.1 Preliminary (syntactic and semantic) notes on other graph-based approaches

We need to mention quantitative approaches which make use of networks of several sorts (in general, inspired by Dependency Grammar) as an emergent group of theories that need to be distinguished from ours (e.g., Čech et al., 2011; Liu et al., 2017). These are sometimes heavily influenced by corpora analysis and aim at clustering languages: there is a measure of *distance* which is used to define some notion of 'language family' or equivalent and produce reliable phylogenetic linguistic trees (Marcolli, 2014, 2016; Siva et al., 2017; Shu et al., 2018). In some cases, these quantitative approaches use dependencies (in the technical sense) and valency analysis as a way to measure linguistic complexity in a way that is cognitively relevant (Fang & Liu, 2018; Liang et al., 2017). Despite the differences between the theories handled in the references that we somewhat unfairly have grouped here, it is worth pointing out that none of these approaches adopt PSGs or can be said to involve *derivations* in any meaningful sense, although because their focus is not set on providing definitions or characterisations of *well-formedness* for either expressions of the language or formulating second-order conditions over rules, they are orthogonal to the IP/IA distinction. Strictly speaking, they are neither *proof-theoretic* nor *model-theoretic*: their focus is not grammar *per se*, but rather *grammar modelling*[11]. It will become clear in the sections devoted to grammatical analysis that our focus on a single language, the centrality of syntax as an abstract system of relations, and the interplay that we assume between syntax and semantics separate the quantitatively-inspired network-approaches (and theories building on signal analysis) from our narrowly grammatical inquiry. We are not concerned with typological or psychological aspects of language, and our methodology obeys the requirements of explicit grammatical analysis; a graph in the present theory is *not* designed to be a representation of what humans 'do' (or even what a computarised 'parser' would do) when interpreting linguistic stimuli, whatever *interpreting* is taken to mean.

We have hailed the fact that our grammar allows for a restricted range of multidominance and discontinuity as a major advantage of our graph-theoretic model over (transformational or not) PSGs when it comes to empirical analysis (a promise we will do our best to deliver on in **Sections 5** to **13**). 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, 1998), 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 or of Minimalist grammars in Johnson, 2016). In this work, we allow for *both* discontinuity and multidominance; this separates the proposal made here from MGG-inspired graph-theoretic works[12], and puts us closer to APG and its descendant 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 mothers in different sub-graphs (each mother can be the root of said sub-graph, 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.

We have defined some basic properties of edges and graphs, now we can turn to characterise the nodes. In our graphs, unlike in the Dependency Grammar and MGG traditions, nodes 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). This choice is not entirely novel; pure PSGs of the kind explored in Gazdar (1982) and much subsequent work adopt a very similar viewpoint. This is where our view

---

[11] This is rather evident in papers like Ferrer-i-Cancho's (2014) comment to a paper by Liu, and touches on the problem of what the object of 'linguistics' is, if there even is something like a unified discipline that can be called that. The centrality of statistical methods in quantitative approaches contrasts with branches of linguistics focused on grammatical description: statistics are nearly meaningless in these (it is legitimate to ask whether they say anything about grammatical description as opposed to humans' usage of the mechanisms made available in a grammar).

[12] The present work 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, for instance, to compare the presentation here with the rigorous axiomatisation of Copy-Chain and Multidominance Structures in Kracht (2001).

differs from Sarkar & Joshi's (1997): instead of assigning nodes in a graph a Gorn address, we assign them a translation into IL. The advantages of such a move include (but are not limited to) the elimination of any need to have a distinct level of 'Logical Form' in the grammar where indexing takes place; the grammar is thus 'single-level' (Perlmutter, 1982). In this context, then, **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 λPP{^d_x} ∈ λPP{^X} or, abridged, d_x' ∈ X'

λPP{^d_x} is in turn an abbreviation (omitting parentheses) of λP[[$^\vee$P](^x)]. Some unpacking is required here. The lambda formalism allows us to form a function from an expression containing a variable (Church, 1936; Montague, 1970). Thus, if x is a variable of type e and P is a variable of type t, then λ(x)[P] is of the type ⟨e, t⟩: P is a first-order monovalent predicate. Let us consider, as an example, the expression λ(x)tall'(x), where x is a variable over individuals: this denotes tall'(x) viewed as a function of x. But the bound variable need not be an individual variable: consider now λP[[$^\vee$P](^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* graphs[13]) corresponding to (19):

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

The two corresponding interpretations, namely:

19') 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 the same connections between nodes:

20) ρ = {(seek, *John*); (seek, *unicorn*)}

Where *John* and *unicorn* stand for λPP{^*John*}and λPP{^*unicorn*} respectively. We will see in **Section 9**, however, that things are not always so simple: we will require of our grammar that it represents *scope* relations between nodes, and we will build on Montagovian analysis trees to model

[13] Where *isomorphism* is defined as follows:

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* $e = <u, v>$*, there is a unique edge* $e^* \in E^*$ *with* $e^* = <\phi(u), \phi(v)>$ (van Steen, 2010: 33; see also Ore, 1990: 10, ff.)

scope. What is not entirely clear at this point is whether the *de re* / *de dicto* distinction modifies the graph-theoretic properties of structural descriptions; this is an issue we will come back to in due time. We will make use of some analytical tools that Montague developed for these cases applied to different syntactic phenomena.

Our theory of 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; Bresnan et al., 2016); 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 proposed in Williams' 2003 *Representation Theory*), etc. A multi-layered syntax entails having mappings from trees to trees, levels of representation related by means of the application of specific operations (i.e., transformations). 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*. In the theory exposed here, because there is a direct translation between syntactic structure and semantic interpretation, neither the syntactic component nor the grammar require the multiplication of levels of representation: importantly, there is no 'generative vs. interpretative' distinction separating syntax from semantics (see also Gazdar, 1982).

Similarly, we are not concerned with the phenomenological existence of unicorns or the truth value of sentences containing the noun 'unicorn'. We only care, for purposes of the formalisation in this work, about the connectivity patterns that we can find and model in natural language sentences. Issues pertaining to what those sentences are used for fall outside the scope of this work.

The fact that nodes in a graph 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 our model, as already stated, nodes in a graph correspond to the Intensional Logic translation (Montague, 1970, 1973, 1974) of *basic expressions* of the language, 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.

Ultimately, the definitions we have provided need to be justified in terms of their usefulness for grammatical analysis. We emphasised above the significance of McCawley's distinction between RCT and RPT; we can now come back to that in somewhat different terms. The core idea that we want to put forth is that most 'transformations', understood here as *purely descriptive devices* (see fn. 14), actually *do not change relations*: at most, they *create* new relations, but *without disrupting those already existing.* In order to describe and characterise 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.

The grammar, in the present conception, aims at maximising connectivity and minimising 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 (there are no *derivations* in the present view, which in and of itself defeats the whole purpose of transformations). This view on what transformations are and what they can do was actually rather widespread in the early days of Transformational Generative Grammar[14], but the line between description and purported explanation was blurred in more recent times, and in orthodox circles (but see, e.g., Postal, 1998: ix-x, 1, for a take on 'extraction' as a purely descriptive term).

Conceiving of phrase markers as *connected cycles* (cf. Rogers', 2003 *composite trees*; his composite *n-dimensional* structures are strictly single-rooted, however), where a *cycle* is an irreducible *single-rooted directed sub-graph*, has the advantages already noted by Grindrod et al. (2014), Saddy & Grindrod (2016) in their formalisation of neural networks as sets of irreducible sub-graphs. 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):

21)

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

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

22) $v_{ij} = 0$ *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 vertex *directly* dominates itself:

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

This is rather conventional, see, e.g., APG's *No Loop Condition* (Johnson & Postal, 1980; 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; also Osborne et al., 2011;

---

[14] For instance, Rogers (1974: 556) lucidly says that:

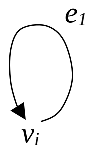
> *Transderivational constraints, global and interpretive rules, and transformations, it seems to me, don't explain anything: they describe* (underline in the original).

Kahane & Mazziotta, 2015), with *mother of* being a dyadic relation between distinct nodes in the tree (such that a node cannot be its own mother). However, we introduce the following condition, pertaining to *transitive domination*:

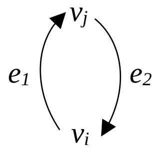
24) $(v_n, v_n) \in \rho^*$

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 (although *not equivalent*) 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's account of 'displacement' phenomena (see also Kracht, 2001). In transformational grammar, *chains* play a fundamental role in the analysis of dependencies that MGG models in terms of *co-indexing* (anaphora, pronominal reference, filler-gap dependencies, etc.), which means that allowing multidominance or not will impact several areas of the grammar.

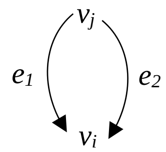
We need to characterise the allowed relations in a graph in some more detail. The reader familiar with APG will note that the *Bicircuit* relation (decomposable into Branch(A, B) ^ Branch(B, A) for any A, B) does hold in the theory exposed here (neither does the relation *Circuit* in Postal's more recent Metagraph grammar; Postal, 2010: 12-13). The relation *Parallel* (where two arcs share both 'heads' and 'tails') (Johnson & Postal, 1980: 41; Postal, 2010: 12-13), however, does. Let us illustrate some of the relations we have been mentioning, borrowing some graphical tools from APG and MG (see also Harary, Norman & Cartwright, 1965):



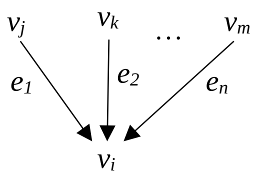
<u>*Loop</u>: v_i cannot directly dominate itself. e_1 must connect two and only two nodes.



<u>Bicircuit</u>: edges e_1 and e_2 form a bicircuit iff $e_1\langle v_i, v_j\rangle \wedge e_2\langle v_j, v_i\rangle$. In this configuration, v_i dominates v_j and v_j dominates v_i. Note that this entails that v_i and v_j dominate themselves *transitively*



<u>Parallel</u>: edges e_1 and e_2 are parallel iff $e_1\langle v_i, v_j\rangle \wedge e_2\langle v_i, v_j\rangle$. In this configuration, v_i dominates v_j establishing two different relations with it (see **Section 4** for details about these relations).



<u>Kiss</u> (Postal, 2010: 13): edges e_1 e_2, …e_n *kiss* iff $e_1\langle v_j, v_i\rangle \wedge e_2\langle v_k, v_i\rangle \wedge e_n\langle v_m, v_i\rangle$. In this configuration, we say that v_i has more than one *mother node*, or that it is *multidominated*

We must note that in APG and MG these are structural relations holding *between edges* (*arcs* in their terms), not nodes: this makes a great difference when defining conditions over 'long-distance dependencies' and indexing given the fact that we are interested not in *paths* but in *trails* in our digraphs. To this effect, we have included the kind of relation that holds between nodes in the annotations in addition to the relations between edges (which are crucial in APG and MG, but are secondary here). We allow for less primitive relations than vanilla APG, but those we allow are *n*-ary and defined over a *single* level of 'representation': there are no *deep* and *surface structures* or (more generally) *pre*- and *post-transformational* representations, which are still very much strong in Minimalism. 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 linearisation of Dependency trees, like Kahane & Lareau (2016); and the axiomatisation of precedence relations in Postal (2010: 26); for a Minimalist perspective, see the linearisation mechanism in Kural (2005).

The representations whose connectivity patterns we try to describe exhaustively are subjected to the usual constraints over dependencies across variables (*à la* Ross, 1967), which also apply to TAGs (see Kroch & Joshi, 1987 for an analysis of Extraposition in TAG terms) and other non-transformational models; this is desirable, given the remarkable empirical robustness of some of Ross' constraints (see Postal, 1998 for discussion). 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) (see also Langendoen, 1975 for extensive linguistically-relevant discussion):

> *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 *maximise irreducible Markov chains (IMC) in derivations*. The connection between these IMCs *can* (but need not) be of a higher order (e.g., a PSG).

We have now enough information to situate this work in the wider context of linguistic theories in terms of what the grammar *is* and what it *does*. In the present 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 considerations about discontinuous constituents

In general terms, '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; Johnson, 2016: Lecture 2), 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 linearisation 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. Postal (2010: 25-26) incorporates a relation *Linearly Precedes* as part of the definition of a Metagraph, as a relation Edge × Edge; such that *1 arcs* precede *2 arcs* and these precede *3 arcs* if we are dealing with neighbouring arcs (which leaves open the possibility of discontinuity, crucially). Precedence between *nodes* is recursively defined upon the relation proposed for edges (or *arcs*):

> **Definition:** *Node Precedes*
>
> *Node a* ***node precedes*** *b if and only if a is the head of an arc A and b is the head of an arc B such that A linearly precedes B.* (Postal, 2010: 26. Highlighting in the original)

Overall, our view on linear precedence may be seen as close to Postal's, although we are careful in not defining the conditions over which linear precedence relations are established in this work because they fall outside our scope. We will see in detail how our proposal differs from DG, APG and Metagraph Grammar, 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:

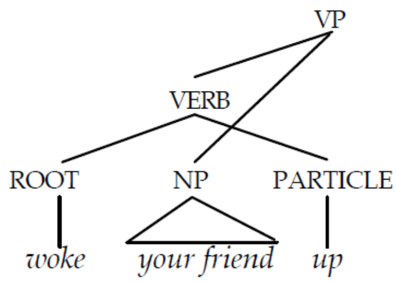
$\rho^{(*)} = \{(A, B); (A, C)\}$

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

e<A, B>; *e*<B, C>; *e*<B, D>; *e*<D, C>

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):

25)



Here, we have the following relations:

26) ρ = {(VP, VERB); (VP, NP); (VERB, ROOT); (VERB, PARTICLE)}

τ = {(VERB, NP); (ROOT, PARTICLE)}

e = <VP, VERB>; <VP, NP>; <VERB, ROOT>; <VERB, PARTICLE>

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:

26') ρ = {(VP, NP); (VP, ROOT); (VP, PARTICLE); (ROOT, NP)}

τ = {(ROOT, NP); (ROOT, PARTICLE); (NP, PARTICLE)}

e = <VP, ROOT>; <VP, NP>; <VERB, NP>; <VERB, PARTICLE>

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} and the edges connecting those nodes), 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 is[15]:

[15] The rationale behind the 'tree pruning' undertaken in the paragraphs above, and which we will assume for the rest of this work, 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

26”) ρ = {(ROOT, NP); (ROOT, PARTICLE)}

τ = {(NP, PARTICLE)}

e = <ROOT, NP>; <NP, ROOT>; <ROOT, PARTICLE>; <PARTICLE, ROOT>

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 subcategorises 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* <NP, ROOT> 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*).

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

27) 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 in our opinion- identifies discontinuity in RCE cases, which ensues the CNPC violation that renders (27b) 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 (28), which contains a deleted copy of [a man]:

28) 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:

29) Cycle 1: [a man$_1$ entered]
Cycle 2: [a man$_2$ was wearing a black suit]
ρ_1 = {(entered, a man$_1$)}
ρ_2 = {(wearing, a man$_2$); (was, wearing); (wearing, suit); (black, suit)}
e = <enter, a man$_1$>; <a man$_2$, a man$_1$>; <wearing, a man$_2$>; <was, wearing>; <wearing, suit>; <black, suit>

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.

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 (see also Rogers, 2003, whose notion of *local tree* is in principle compatible with our characterisation of *cycles*, although his is defined strictly in configurational terms without the actual terminals and their properties playing a role). 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, 2019):

a. A predicative basic expression *p*
b. Temporal and aspectual modifiers of *p* (cf. Bravo et al. 2015's *functional auxiliaries*)
c. Nominal arguments of *p* (subjects, objects, non-argumental clitics)

In García Fernández & Krivochen (2019) we referred to each structural unit containing elements (a), (b), (c) as the *extended projection* of *p* (in the sense of Grimshaw, 2000; Abney, 1987: 57). This idea is weakly equivalent to the so-called *Condition on Elementary Tree Minimality* (CETM) in *Lexicalised Tree Adjoining Grammars* (LTAG; Joshi & Schabes, 1991; Frank, 1992, 2002, 2006):

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

More recently, Frank (2002: 22) elaborates on this perspective, also based on the notion of *extended projection*, slightly reformulating the CETM in the following terms:

> *The syntactic heads in an elementary tree and their projections must form an extended projection of a single lexical head.*

The restriction on the size of elementary trees proposed by Frank is essentially what we are going for, provided that the elements in the *extended projection* of a lexical head are the ones in (a-c): note that nominal arguments are part of the extended projection of the verbal predicate that select them; in this sense, simple NPs do not configure independent cycles (things change with Complex NPs, see **Section 8**). Moreover, there is not really a notion of 'projection' in our system, because there are no heads or intermediate non-terminals ('bar-levels', in X-bar parlance). Thus, we are making *two* claims:

A. Cycles are irreducible, single-rooted graphs
B. These correspond to the 'extended projection' of a single predicate: the smallest set of connected nodes that contains all three (a-c) above.

Let us go back to the analysis. In (29), 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 realised 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 –categorial, say- being percolated up 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, where instead of an *elementary tree* we will refer to *cycles* (i.e., local units defined as single-rooted graphs). 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, in our case 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 subcategorising 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 (28). In the structural description (29), the relation <a man_2, a man_1> 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): λPP{^a man}, or *a man*'. 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 (in Sarkar & Joshi's 1997 terms, they are assigned the same Gorn address). Linking both cycles operates in a manner that may remind the reader of operations like Unification[16] (in the sense of Shieber, 1986): nodes that

[16] It is important to note that *Unify* is an operation over feature matrices, which makes it particularly well-suited for feature-rich theories like HPSG and LFG. Shieber (1986: 14) defines the operation as follows:

> *In formal terms, we define the unification of two feature structures D' and D" as the most general feature structure D, such that D' ⊆ D and D" ⊆ D. We notate this D = D' ∪ D".*

Some further clarification is necessary here. Within Unification grammars, ⊆ is used to symbolise a *subsumption* relation between feature structures, in which a feature structure, abbreviated D, contains *part* of the information of another D', such that D' ⊆ D. In more complex terms, the concept of *subsumption* is based on that of *dom(D)* (the *domain* of a feature structure, namely, the features it includes, regardless of their mapped values), such that D' ⊆ D *iff* $\forall(x) \mid x \in dom(\text{D'}), x \in dom(\text{D})$. A more concrete example will clarify the differences between *Unify* and (Chomskyan) *Merge* (taken from Jackendoff, 2011: 276, ex. 10 a, b):

i) a. Unification of [V, +past] and [V, 3 sing] = [V, +past, 3 sing] stand for a single IL translation are identified with no need to resort to an independent indexing mechanism.

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 IL 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, what we can refer to as 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). In principle, we could be talking about a single predicate and more than one grammatical relation or as many predicates as grammatical relations, always with a single argument: the theory allows for this, therefore, we need to verify whether those relations actually hold in empirical analysis. More than one predicate establishing more than one grammatical relation with a unique argument is a well-known situation: consider *equi* and *raising* structures. But one predicate establishing more than a single grammatical relation with a single argument is not any less common: this corresponds, for instance, to *reflexivity* (Reuland & Reinhart, 1993). We will come back to these cases in **Section 5** below. It is important to note that the relation *dominance* is not always to be interpreted in predicate-argument terms: that depends on properties of the relevant nodes connected by a directed edge. The theory of grammar cannot, we argue, focus on configuration alone and ignore how the properties of related elements influence configuration; the descriptive importance of *lexically governed* processes and *exceptions* must not be underestimated (see Lakoff, 1965; Gruber, 1965; Dowty, 1978, among many others).

If this is the case, however, then we need to make some adjustments, because we are *visiting the same node twice*, once per cycle. This approach to *walks* in structural descriptions 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 that there are no *derivations* in the present framework), nor do we require to link two occurrences of an object[17]. We would also want to eliminate the edge *e* <a man_2, a man_1>, because if we are dealing with the same node (each visit to which materialises as a *token*) we would be having ρ(*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: given the fact that 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 (29); 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 our preliminary analysis above.

(not *[[V, +past] [V, 3 sing]]*, as with Merge)
b. Unification of [VP V NP] and [V, +past] = [VP [V, +past] NP]
(not *[[V, +past] [VP V NP]]*, as with Merge)

[17] 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-.

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 *generalised* transformations (*adjunction*, if interpreted as a general case for *substitution*; see Kroch & Joshi, 1985: 11). We can provide the following semi-formal definition:

30) *Linking (definition):*

*If a sub-graph G contains v_1 and a sub-graph G' contains v_1 (that is, if $\rho(v_G, v_1)$ & $\rho(v_{G'}, v_1)$ hold, for v_G and $v_{G'}$ arbitrary nodes in G and G'), then G and G' are **linked** at v_1*

Transitively, anything that an arbitrary node *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 node S. The revised set of relations in (29) would then be (29'):

29') Cycle 1: [*a man*' [Cycle 2] entered]
Cycle 2: [*a man*' was wearing a black suit]
ρ_1 = {(entered, *a man*')}
ρ_2 = {(wearing, *a man*'}; (wearing, suit); (black, suit); (was, 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 G and G', we are in the presence of a *conjoining transformation*. If *linking* targets an embedded node (i.e., a non-root), we have *embedding*, as in the RCE case above. In both cases, because dominance relations are transitive, anything dominated by the node G and G' are *linked* at is accessible for syntactic relations at G and G'. The notion of *linking* is essential to formulate constraints on possible relations between cycles; in other words, it will prove very useful when dealing with issues of *locality* and *opacity*: we will attempt to answer the question 'when is a node or sub-graph accessible for relations involving a node outside the first's cycle?' We will see that *linking* may overgenerate in terms of allowing for connections between graphs to hold, if not appropriately restricted: in **Section 6** we will introduce the notion of *self-contained graph* to formalise such a restriction.

In the following sections we will focus on the explicit analysis of various English constructions, and in these analyses compare our model with competing syntactic approaches based on their empirical adequacy and the extent to which additional entities and relations need to be invoked.

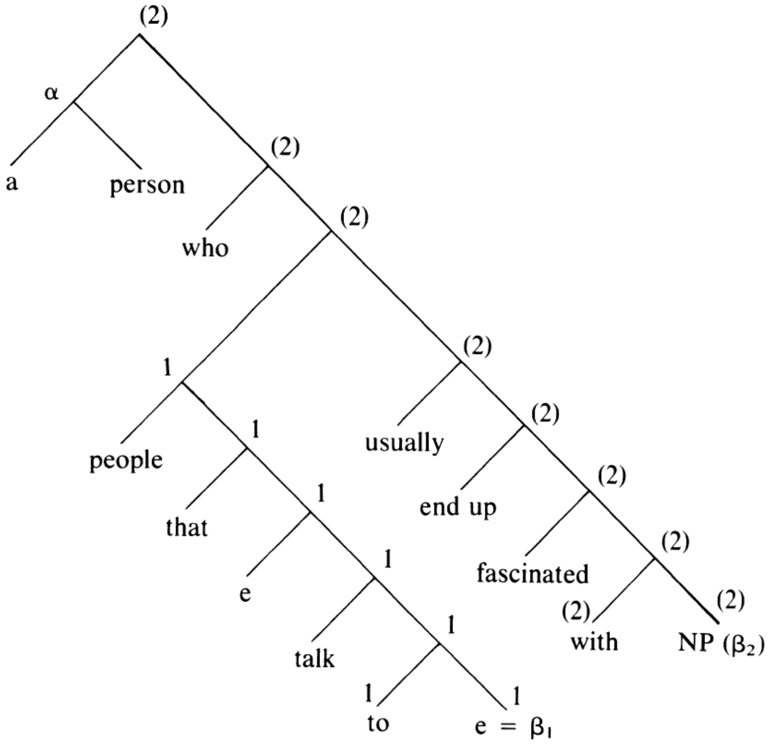
4. Comparing models

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 unless something like 'self-Merge' is explicitly allowed- 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 (31):

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

The GB-MP representation is (32):

32)



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 $G_{\beta 1}$, and $G_{\beta 2}$* [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:

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

However, despite sharing denotata (or, in the framework under discussion here, *despite having identical IL translations*), those are all *different nodes*. This means that extra elements must be introduced in the representation to encode semantic-referential identity: indices. In such a theory, syntactic objects which share indexes receive the same semantic interpretation, in the sense that they denote the same object. The sub-module of the grammar that is in charge of defining and limiting the possible relations between indexed referential expressions in MGG is known as Binding Theory (Chomsky, 1980, 1981; Reinhart, 1983; Reinhart & Reuland, 1993; see Culicover & Jackendoff, 2005: Chapters 11 and 12 for a different view, but still generative) Note that the SMC 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 (immediate dominance) that the relevant node establishes with other nodes.

Moreover, in transformational PSG-based models, local cycles are stipulated in terms of 'government projections', 'bounding nodes', 'barriers', 'phases' (among other notions, usually mutually translatable; see e.g., Boeckx & Grohmann, 2004) and there is no natural way to capture predication domains in these approaches (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 (31) above:

34) Cycle 1: [A person #]
Cycle 2: [who [Cycle 3] usually end up fascinated with [Cycle 1]]
Cycle 3: [people that talk to [Cycle 1]]
ρ_1 = {(a, person)}
ρ_2 = {(end up, who); (fascinated, who); (fascinated, with); (with, who); (fascinated, who)}
ρ_3 = {(people, that); (people, talk); (talk, to); (to, #)}

Where '#' is a placeholder to which cycles can be adjoined (*à la* Chomsky, 1955); basically, it works as a 'wild card node' to be targeted by *adjunction* or *substitution*, as appropriate (Joshi, 1985; Kroch & 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, λPP{^a person}. Similarly, we will use *people*' as an abbreviation of λPP{^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, 2018). There are *not* two distinct nodes, but *a single node* which can be visited as many times as necessary in a trail[18]: all differences pertain to *context* (that is, which node was visited immediately before, and which node will be visited immediately after; or, in other words, the strict ordering relations established between nodes within a cycle). Then, we can simplify (34) as (35):

35) Cycle 1: [*person*']
Cycle 2: [*people*' usually end up fascinated with *person*']
Cycle 3: [*people*' that talk to *person*']
ρ_1 = {Ø}
ρ_2 = {(end up, *people*'); (fascinated, *people*'); (fascinated, with); (with, *person*'); (fascinated, *person*'); (end up, fascinated)}
ρ_3 = {(*people*', that); (talk, *people*'); (talk, to); (to, *person*')}

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

36) Cycle 1 ∩ Cycle 2 = {P'}
Cycle 1 ∩ Cycle 3 = {P'}
Cycle 2 ∩ 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 (31). Note that the 'pivot' is *not* the root of any of the cycles involved in the process, which is not allowed in the

[18] Incidentally, this means that the antecedent is in a way contained within the relative clause. This is not to be interpreted as supporting the 'external headedness' hypothesis (Bresnan & Grimshaw, 1978; Larson, 1987 and related work), but it does entail that there is a close articulation between antecedent and relative clause: the antecedent and the *wh*-operator are not two distinct nodes but only one, visited more than once in a trail. We need to note immediately that this analysis does not extend automatically to antecedent-less relative clauses (*free* relatives, *transparent* relatives).

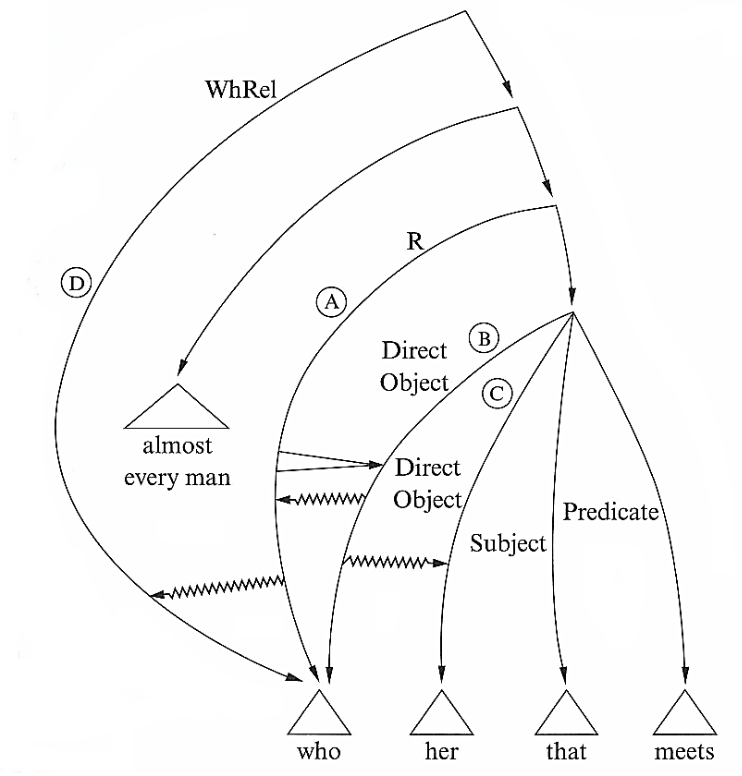
frameworks of Joshi (1985); Frank (2002); or Rogers (2003): sub-graphs need not be linked through their roots-frontiers (as they must in TAGs), in principle they can be linked *at any node*.

It is interesting to briefly compare our representation of a multiple-gap construction with that proposed in Postal (2010: 41) in the framework of Metagraph Grammar. Specifically, Postal considers the following sentence:

37) Carla is a woman [who almost every man that meets her falls in love with]

And assigns the following arc structure to the bracketed relative clause (note that arcs are labelled with the appropriate grammatical relation plus a letter; the letter serves no purpose to the formalism, and only provides us with a way to refer to specific arcs by their letter):

38)



Much discussion about the place of resumptive pronouns in the theory of locality aside (but see **Section 13.3** below), we will focus on the differences between the assumptions in APG/MG and our model rather than on explaining every detail of the simplified structure in (38) (see Johnson & Postal, 1980 for extensive discussion). Here we are interested in two aspects:

i) The notion of *pronominal arc* (and *arc antecedence*)
ii) The rules *sponsor* and *erase*

The relation *erase* is represented by double arrows (such that A erases B), and *sponsor* is represented with the wiggly arrow (such that B sponsors A, A sponsors D, and B sponsors C). Because erased arcs are not part of the Surface Graph (S-graph) of the sentence (Postal, 2010: 27, 30; Johnson & Postal, 1980: 88-89), arc B does not belong to the S-graph. The notion of S-graph is crucial in APG and MG, and thus it is worth defining it briefly so that the differences between S-graphs and our graphs are fully explicit. First we need to define R-graphs. R-graphs are finite sets of arcs and nodes which represent all structure of a sentence except that involving *erase* and *sponsor*, and include all grammatical relations between elements; R-graphs need not respect the SMC, do not contain loops, and are single-rooted (Johnson & Postal, 1980: 51, ff). The result of considering *erase* and *sponsor* relations is a *subset* of R-arcs, and S-graph:

> *An S-graph:*
>
> *a. is connected,*
> *b. is rooted, and*
> *c. contains no overlapping distinct arcs*
>
> […] *S-graphs do not have parallel edges, do satisfy the Single Mother Condition, and have no circuits.* (Postal, 2010: 28)

Furthermore, Johnson & Postal (1980: 90) prove that every S-graph is an R-graph.

It is important to note that in Postal's representations (R-graphs and S-graphs alike), *who* and *her* are two distinct nodes, with *her* being the tail of a nonreflexive resumptive copy arc. Arc B sponsors C and A, and B is deleted; this establishes a relation between A and C called *arc antecede*, which is what in MG (roughly) corresponds to the 'coindexing' mechanism of MGG (Postal, 2010: 34). Note that in order to get correference it is still assumed that the grammar requires a binary relation between arcs, *arc antecede*(A, B), in turn dependant on two other relations: *sponsor*(C, A) and *sponsor*(D, B) and *primitively antecede*(C, D). Postal's conditions and relations are formulated over *arcs* rather than over *nodes* and *trails*, which is already a major difference between his approach and ours. But the point here is that in the graph theoretic proposal we formulate here there is no need to multiply nodes, arcs, or relations: as a matter of fact, correference is one of the major sources of *simplification* in graphs, where 'simpler' is understood in a pre-theoretic way, intuitively as 'having less nodes and/or less edges'.

In the framework advanced here, *who* and *her* are a single node which links several cycles, just like *people* and *person* do in the structural description (35) above. Because of this, in our model there is no need to have either pronominal arcs or sponsor arcs: endophoric dependencies are defined in terms of strict ordering between nodes in trails. Provisionally (we will refine this intuitive characterisation in **Section 5.2.1** and **Section 7**) we can say the following:

Identity-ordering condition on pronominalisation:

v_j can pronominalise v_i iff

a. v_i and v_j denote sortal entities
b. $i = j$, and
c. $(v_i, v_j) \in \rho^*$, such that there is a unique trail that visits v_i and v_j in G and in this trail v_i is ordered before v_j.

This condition captures not only the transformational intuition that for A to pronominalise B they have to be coindexed nominals (Lees & Klima, 1963; Postal, 1969; more recently Hornstein & Idsardi, 2014: 14, ff., among others), but also that pronominalisation is a relation that requires a certain order between the elements it applies to. Note that we said 'order' and not 'c-command' or anything of the sort: it is crucial to note that the notions are not equivalent (particularly in GB-MP incarnations of MGG); allowing for parallel arcs and circuits in local graphs is, as already mentioned, incompatible with a view of phrase markers based on Kaynean 'unambiguous paths'. As we saw in **Section 1**, a model of syntax based on projecting binary-branching trees in which asymmetric c-command underlies linearisation statements at the phrasal and morphological levels (Kayne, 1994, 2013; Moro, 2000; Di Sciullo, 2005, 2014) allows for a very limited range of structural configurations, based on an *a priori* phrasal template (for constituent order, if the only order allowed by UG is Spec-Head-Compl; then that would be SVO, with any 'deviation' -e.g., SOV- being transformationally handled). It is worth repeating that the relations we allow for in our graphs (including, but not limited to, *parallel* edges, multidominance, and globally multi-rooted structures) arise from empirical necessity rather than from axiomatic statements.

The framework we present here maintains some descriptive aspects captured by the notion of 'constituent' from Immediate Constituent analyses, although with important changes: our structural blocks are sub-graphs, not labelled phrases. These sub-graphs are *cycles* containing one and only one predicate with its functional and nominal dependants, and neither their size nor their connectivity patterns are determined beforehand, they cannot be. This flexibility is not a feature of modern versions of transformationally-enriched phrase structure grammars, but it is an important feature of frameworks like Categorial Grammar and some versions of Dependency grammar; in turn, these have major differences between them. It may be observed that Categorial Grammars (which, as shown 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'. As noted above, in some versions of CG, the algebra of a language contains a set of syntactic rules that recursively assign expressions of a language L to indexed categories. Indexes are important in this sense, because categories are simply sets of expressions. It is the index of a category that allows us to refer to it. Thus, we can use, say, NP, VP, COMP, etc. as *indexes*, that is, as 'names' for sets of expressions of the language (not as *labels* in the sense of formal language theory, or as the projection of categorial features in lexical items, as in MGG). In this work we will be consistent with the CG use.

It is important to stress now that, while the internal constitution of each sub-graph may superficially 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 (see also Krivochen, 2015a; 2018) 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: the point is to note that the approaches are not notational variants of one another, and that there are non-trivial theoretical and empirical differences that deserve careful analysis.

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

39) Walk really fast

Looks like (40a) and *not* like (40b):

40) 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 in fact *should*- incorporate exocentric structures in order to be descriptively adequate). 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 (39). 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:

41) ρ = {(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, 2018a), *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:

42) 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 (42c), 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. As observed above, this is not a new idea, it actually predates both PSGs and DGs: Jespersen (1985 [1937]: 6, 25) talks of 'composite verbal expressions' for things like *wait on* in *She waits on us* or *talks with* in *He talks with himself* and assigns *wait on* and *talk with* a single primitive symbol W; similarly, he recognises 'composite prepositions' like *on account of*, and similarly assigns them a single primitive symbol *pp* (1985 [1937]: 6, 32). We have discussed the problems that things like *would rather* pose for monotonic PSGs (including GB/MP grammars) in Krivochen (2016a), 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 *orthographic 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 and the basis of GPSG/HPSG (see Gazdar, 1982 for some early discussion), 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^*[19]), 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 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 theoretic approach have the relation *dominate* as an essential device, 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. Summarising, there are local compatibilities between our graphs and Dependency graphs, and between our graphs and PSGs, but the systems are far from equivalent.

Outrageous and alien 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 *across-the-board* (ATB henceforth)[20] extraction (which owes much to McCawley, 1982; Levine, 1985):

43) I wonder what Hansel read and Gretel recommended

---

[19] 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, ...}.*

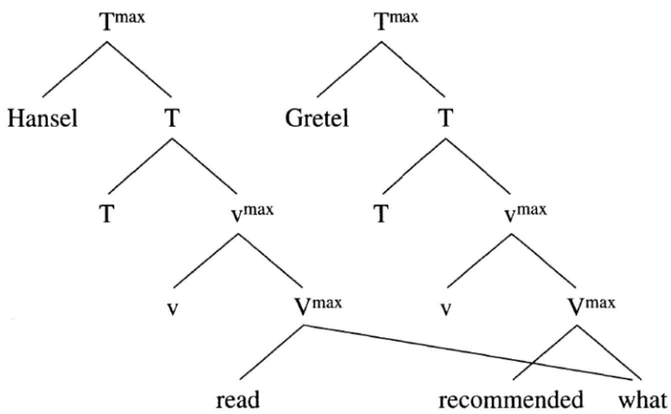
[20] A rule applies *across the board* (ATB) if and only if it affects *all* terms of a true coordinated structure. Williams (1978: 32) proposes the following generalised definition:

i) *The structure*

$$\begin{bmatrix} [X_1]_{C_1} & \\ \ldots & and \\ [X_n]_{C_n} & \end{bmatrix}$$

Is a well formed labelled bracketing if $X_1, \ldots, X_n$ are.
We will say that a string containing structures defined by [(i)] *is in ATB format.*

Williams' definition anticipates elements of Goodall's (1984) conditions over transformations in terms of *parallel structures*.



The node T^{max} can be safely characterised as a *root* in each sub-tree corresponding to the structural descriptions of *Hansel read what* and *Gretel recommended what* (we do need to note that the full derivation proposed by Citko eventually adds more functional structure in the form of a unique Complementiser layer CP, which yields a single-rooted P-marker; here such additional structure is deemed unnecessary). (43) is a representation that follows the classical MGG clausal template, including phonologically empty functional (terminal and non-terminal) nodes. Minus non-audible functional structure (nodes like T, *v*), our *local* description is not too far from that:

44) Cycle 1: [I wonder [Cycle 4]]
Cycle 2: [Hansel read *what*]
Cycle 3: [Gretel recommended *what*]
Cycle 4: [[Cycle 2] and [Cycle 3]]
ρ_1 = {(wonder, I); (wonder, *what*); (wonder, and)}
ρ_2 = {(read, Hansel); (read, *what*)}
ρ_3 = {(recommend, Gretel); (recommend, *what*)}

A fundamental difference between our representation and Citko's (see her ex. (13), p. 482) is that in our representation the coordination in (43) 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, 2016a for detailed discussion; also **Section 11** nelow). In any case, a declarative sentence featuring RNR, 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 (43) 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 (45a) (a 'flat structure' of the kind argued for in Culicover & Jackendoff, 2005: Chapter 4) or an endocentric coordination phrase as in (45b) (see, e.g., Progovac, 1998; Camacho, 2003):

45) a. S* → S and S (and S)
b. &P → S &'
&' → and S

Which generate the following structural descriptions, respectively:

45’)

In bottom-up models (mainstream MP), the TPs corresponding to each clause are dominated by a Complementiser 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 radicalisation 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 LFG’s *Economy of Expression* (or any such constraint). This restriction on the theory raises questions pertaining to the kinds of structural descriptions that can be assigned to expressions, and the expressions that can be proven to be well-formed expressions of the language, in our case, of English. In traditional terms, these questions pertain to the *strong generative capacity* of the theory. The following section will deal with the issues of adequately restricting the power of our graphs.

4.1 On the ‘generative power’ of graphs

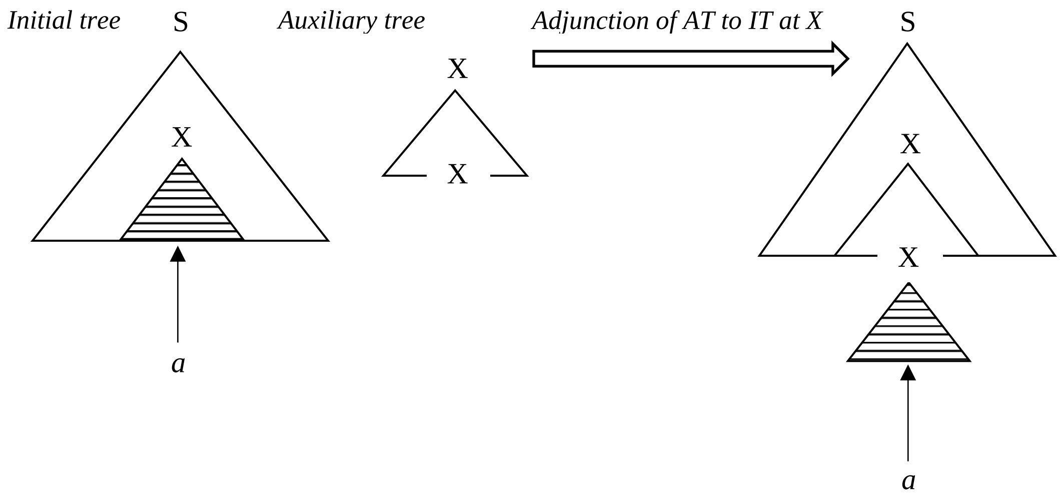
It seems clear that we can generate graphs for all CF languages (as can the base component of a transformational grammar, the *c*-structure component of LFG, and GPSG). But, what is the true generative power of our theory? This question is somewhat tricky: because we are not recursively enumerating strings, our ‘syntax’ is, strictly speaking, *not generative*. But there is a sense in which the notion of *strong generative capacity* (in the sense of Chomsky, 1965: 60) is meaningful: we want to know what kinds of structural descriptions for natural language strings we can 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 Tree Adjoining Grammars (TAG) 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[21]. We must first introduce some basic aspects of TAGs.

---

[21] It is worth noting that there are versions of TAG which are rather close to MGG (e.g., Frank, 2006) in some aspects, particularly in accepting binarity as a guiding principle (but not global monotonicity, clearly, which is disrupted by the operations of *substitution* and *adjunction*), the use of traces and empty nodes, and the use of ‘bar-levels’ in structural descriptions (Kroch & Joshi, 1987; Frank, 2004, 2006).

A TAG is a set G = (I, A), where I is a (possibly unary) set of *initial trees* and A is a (possibly empty) set of *auxiliary trees*. Initial trees are single-rooted structures which contain a non-terminal node which is identical to the root of an *auxiliary tree*. *Auxiliary trees* are also single-rooted structures, which contain a node in their frontier which is identical to their root: this allows for *auxiliary trees* to be adjoined to *initial trees* at the nonterminal that corresponds to the root of the AT. We diagram the process of *adjunction* in (46), where S is a sentential root, X is a nonterminal (or 'intermediate') node, and *a* is a terminal node:

46)



TAGs with links can generate *limited* crossing dependencies, by virtue of having elementary trees annotated with links which are preserved under *adjunction*. These *links* characterise binary relations between nodes in an elementary tree. Joshi's definition of *linking* is worth citing, cf. our own definition in (30) above:

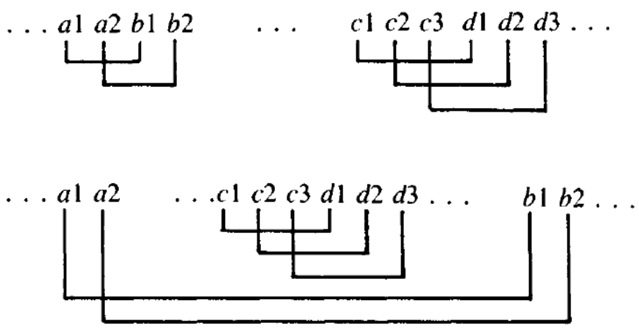
If a node n_1 is linked to a node n_2 then

(i) n_2 c-commands n_1 (i.e., n_2 precedes n_1 and there exists a node m that immediately dominates n_2 and also dominates n_1),

(ii) n_1 dominates a null string […] (Joshi, 1985: 214)

Links are crucial when evaluating the strong generative power of TAGs. A TAG can generate the CS language L = {a^n, b^n, e, c^n, d^n | $n \geq 0$} *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):

47)



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 | $n \geq 0$}, and double-copy languages, e.g., L = {$w\ e\ w\ e\ w$ | $w \in \{a, b\}^*$}.

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, 1998 and Sag, 2010)
II. To indicate dependencies between overt elements (agreement, correference)

Because there are no *gaps* in the present model (because there is no movement or slash-features), 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 the application of the transformation *Pronominalisation* in a Lees & Klima-Ross-Langacker view, for instance- do not arise in our theory because those relations are formulated in terms of trails that visit a *single N node* more than once, *not* as a relation between two distinct N nodes which share an index (or some other diacritic indicating identity). But (II. a.) is worth looking at. This is the kind of dependencies that arise in the now famous Dutch examples considered by Joshi:

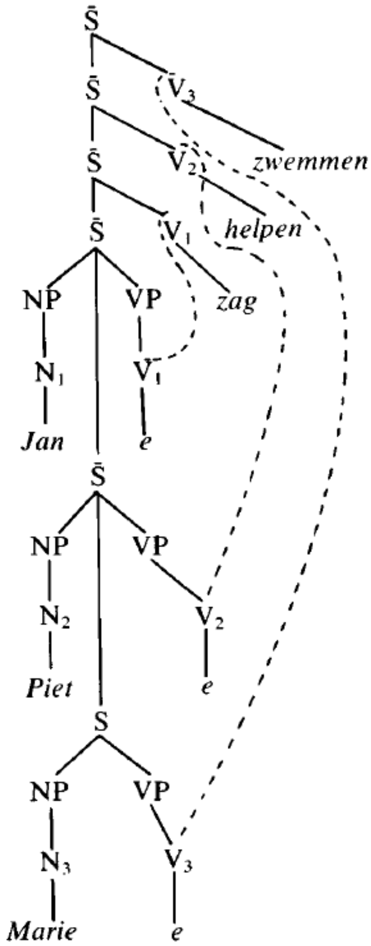
48) Jan Piet Marie zag helpen zwemmen (Dutch German)
Jan Piet Marie saw help swim
‘Jan saw Piet help Mary swim’

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

49) Jan Piet Marie zag helpen zwemmen

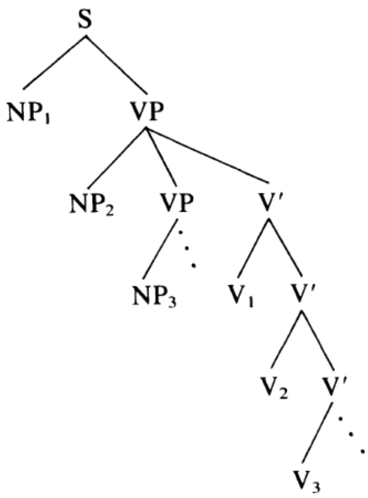
The abstract form of (47) is $a^1a^2a^3b^1b^2b^3$, which displays crossing dependencies (indicated with superscript integers). The structure required is more complex than it would seem *prima facie*:

50)



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 phrase marker (Bresnan et al., 1982: 619):

51)



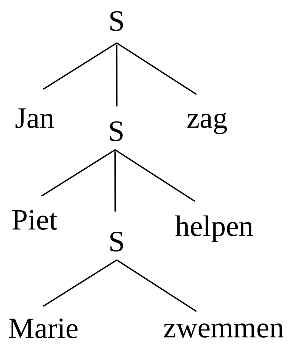
Despite some differences, both works conclude that there is no need to push the grammar to *strong* context sensitivity (i.e., to make the grammar powerful enough that it can generate *all* context-sensitive languages); *weak* context-sensitivity, a.k.a. *mild* context-sensitivity[22], seems to do the trick. Joshi’s structural description is based on the linear position of elements, *as well as* their referential dependencies, which is an important difference with our model: recall that the notion of *order* that is essential in the present proposal is not *linear order* (a.k.a. *precedence*), but rather the existence of trails connecting nodes in *digraphs*.

Let’s see if we can assign an adequate structural description to (48) in our graph-theoretic model. To begin with, let us consider an incorrect, but initially 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:

52) Cycle 1: [Jan zag [Cycle 2]]
Cycle 2: [Piet helpen [Cycle 3]]
Cycle 3: [Marie zwemmen]

But things are more complex (the inadequacy of (52) is indeed prefigured by Joshi). As it is, and assuming that structural descriptions encode aspects of linear order, as Joshi does, (52) describes (*weakly generates*, using traditional parlance) the string *Jan zag Piet helpen Marie zwemmen*. This is certainly not what we have in (48). Equally, an approach based on strict center-embedding (manipulating only *cyclic monostrings* in the sense of Lasnik & Kuppin, 1977: 176-177) is also inadequate, despite the fact that PSGs can indeed capture dependencies between objects which are ‘indefinitely far apart’ at the cost of adding ‘invisible’ structure in the form of nonterminals (see Lasnik, 2011: 356. 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]-):

52’)



Cycle 1: [Jan [*Cycle 2*] zag]
Cycle 2: [Piet [*Cycle 3*] helpen]
Cycle 3: [Marie zwemmen]

Both verbs *zag* 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

[22] Bresnan et al. (1982) do not define *weak context sensitivity* formally. In Joshi’s work, ‘mild context sensitivity’ includes the following properties (as summarised in Krivochen, 2018: 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) 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 (53), with the full set of dominance relations made explicit:

53) ρ = {(zag, *Jan*'); (zag, *Piet*'); (zag, helpen); (helpen, *Piet*'); (helpen, *Marie*'); (helpen, zwemmen); (zwemmen, *Marie*')}

Note that we have made no use of empty nodes, intermediate non-terminals (i.e., 'bar levels'), or headedness. If there is something like 'heads' in the Dependency Grammar sense, they arise from the consideration of dominance relations, but they are not primitives of the theory (therefore, 'projection' is ruled out).

Of course, the system described here can also accommodate facts from languages where clausal complementation does not yield (limited) crossing dependencies (thus, mildly context-free), but rather center-embedding (thus, strictly context-free), as in the Turkish counterpart of (48):

48') Merve Ömer'in Esra'nın yüzmesine yardım ettiğini gördü (Turkish)
Merve Ömer Esra swim help give saw
'Merve saw Ömer help Esra swim'

The dependencies in (48') are as in (49'):

49') Merve Ömer'in Esra'nın yüzmesine yardım ettiğini gördü

And, of course, the English pattern, in which each verb takes a complement clause to its right (head-tail recursion):

48") John saw Peter help Mary swim

In the following section we will examine the problem of grammatical functions defined in dominance sets.

4.2 Ordered relations and grammatical functions

It is widely acknowledged that adequate structural descriptions for natural language sentences need to represent, at some level, the grammatical relations that are established between arguments and predicates. There are differences between theories with respect to the status of grammatical relations in the theory of grammar (primitive vs. derived) and the *locus* at which these relations are determined (whether at a pre-transformational level of Deep Structure, or an independent stratum like *a-structure* or Conceptual Semantics, which can map onto other levels by means of homomorphic interface operations); the descriptive usefulness of grammatical functions for a grammar of English (our focus here) is however rarely questioned.

This, of course, only means that we must be able to represent grammatical relations somehow. This is not a trivial task: if there are edges from –say- *zag* to both *Jan* and *Piet*, how do we know which one is the subject and which the object? A possibility is that the difference is marked by means of morphological agreement. If verb-subject agreement was encoded in an edge *e* <Subj, V>, there would be no problem: there would be a cycle between every verb and its subject, but not between every verb and its object in (48). That is: there is a trail Subj-V-Subj but not Obj-V-Obj. This revised picture would yield the relations in (53'):

53’) ρ = {(zag, *Jan*’); (*Jan*’, zag); (zag, *Piet*’); (zag, helpen); (helpen, *Piet*’); (*Piet*’, helpen); (helpen, *Marie*’); (helpen, zwemmen); (zwemmen, *Marie*’); (*Marie*’, zwemmen)}

However, there seems to be no strong reason to attempt to represent morphological agreement in terms of dominance. As a matter of fact, given that our focus is to build a fully explicit grammar of English in terms of maximally connected graphs, giving morphological agreement a prominent place in the identification of grammatical functions does not seem to be supported. In other words, defining the ‘subject’ of a clause as the element that agrees morphologically with the verb is –in our opinion- more trouble than it’s worth.

Another possible course of action is to define grammatical relations configurationally, a classical move adopted in Chomsky (1955: 254; 1965: 69)

> *It is necessary only to make explicit the relational character of these notions by defining "Subject-of," for English, as the relation holding between the NP of a sentence of the form NP⁀Aux⁀VP and the whole sentence, "Object-of" as the relation between the NP of a VP of the form V⁀NP and the whole VP, etc. More generally, we can regard any rewriting rule as defining a set of grammatical functions* (Chomsky, 1965: 69)

In Chomsky’s view, grammatical functions are read off syntactic configuration. An obvious difficulty in adopting a configurational view of grammatical function arises: there are no phrase structure rules in our system. But there are deeper problems. One of those is that, in PSGs there is no way to represent a single syntactic object with more than a single grammatical function unless additional mechanisms are invoked. Let us limit ourselves to the brief sketch in Chomsky’s citation: we have Subject-of and Object-of defined in configurational terms. Let *G* be a PSG with alphabet Σ = {Det, N, NP, V, VP, S} and the following set of rules:

S → NP, VP
NP → Det, N
VP → V, NP

G generates the string Det⁀N⁀V⁀Det⁀N. Here, the first Det⁀N substring is assigned the GF Subject, and the second one is assigned the GF Object, where ‘first’ and ‘second’ are to be interpreted in terms of *linearity as precedence*. However, there are several problems with this view (which have been reviewed in detail in the RG/APG/MG literature); here we will simply point towards some of the simplest ones. Consider for instance the case of *reflexivity* (to which we will come back below): a predicate is *reflexive* if its subject and object are correferential (Reuland & Reinhart, 1993). In this case, a grammar along the lines of Chomsky’s proposal must necessarily indicate correference by means of additional devices, diacritics or indices, such that NP-Sister of VP and NP-Sister of V are assigned the same index. So the entities in the representation are multiplied in two ways: we have *two* NPs instead of one (because of how the rules are formulated), and because of this we need to extend the grammar by adding referential diacritics. This multiplication, we argue, is neither desirable nor required or motivated empirically. In the empirical side, Johnson & Postal (1980: 16) remark that the Chomskyan view, which is closely related to a rigid and aprioristic view of clause structure (an aspect of the generative theory that, as we have observed several times now, has only became stronger in Minimalism) has remained insensitive to difficulties raised by languages that do not conform to the Spec-Head-Compl template determined by Universal Grammar (which translates to SVO in GF terms); in these cases not only entities but also operations have multiplied: reordering transformations must be invoked in order to restore a putative underlying SVO representation (Kayne, 1994: Chapter 4 and much subsequent work). Any internal complexity pertaining to the GFs themselves is also erased: the characterisation in Chomsky's quotation and the work stemming from it does not distinguish Direct from Indirect Objects. Of course, any new GF can be incorporated into the template, but only at the cost of introducing new nonterminal nodes: just add a new XP within the VP that excludes an NP sister of V and dominates an NP, and define Indirect Object in terms of the relation between NP and XP. But that is an unnecessarily *ad hoc* move; and further articulation in the definition of GF in different languages would require an explosion of nonterminal labelled nodes; in the worst case scenario, one *per* GF[23].

Are there other possibilities? Yes, there are. Even without adding to the theory levels like *a-structure* for roles of participants or *f-structure* level for grammatical functions (as in LFG; Bresnan, 2001: 19, ff.; Dalrymple, 2001: 8, ff.; see also Williams, 2003 for a theory rich in levels of representation and mapping functions) 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* or *labels* to nodes or edges. Examples of the latter strategy is to be found in RG, APG, and Metagraph grammar (Perlmutter & Postal, 1983a, b; Johnson & Postal, 1980; Postal 2010 respectively), where arcs indicate the grammatical relation that is instantiated between the nodes connected by that arc. For instance, a basic (and admittedly incomplete, although in ways that are at present irrelevant) general format for local L-graphs could be (54):

54)

$$\begin{array}{c} a \\ \text{GR}\big\downarrow \\ b \end{array}$$

Perlmutter & Postal (1983b: 83)

(54) 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), where numbers make reference to the place of a specific relation in a hierarchy of grammatical functions.

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:

[23] Such an explosion is not deemed undesirable in some versions of Minimalism, paradoxically, particularly those grouped under the term 'cartographic approaches'. Ramchand (2008) is an example of a cartographic analysis of the verb phrase; see also Ura (2000: 13-14, 25, ff.), for a minimalist theory of GF as the result of checking operations. It is very important to note that orthodox generative views on GF assume that the inventory of GF is given *a priori*, by Universal Grammar (however that works). The similarities between the cartographic approaches and a late GB approach, in which core grammatical functions were associated to distinct Agreement projections is striking, but rarely acknowledged. The agreement system in late GB assumed functional agreement projections AgrS, AgrO, AgrIO for Subject, Object, and Indirect Object case marking respectively; see Chomsky & Lasnik, 1993; for AgrIO, see Radford, 1997: 242. Despite being a textbook, Radford at least presents step-by-step derivations –a rare feature in mainstream Minimalist grammar of the Chomskyan persuasion-, and is left with no choice but assume a further Agr projection to keep the system working.

53”) e_s<zag, Jan>, e_o<zag, Piet>, e_o<zag, helpen>…

However, (53”) is not quite it either. We find this notation cumbersome and unnecessarily stipulative, and –what’s more- not really useful: adding diacritics to the set of tools required by the grammar is only justified if there is no other way to encode grammatical distinctions in the system itself, making use of elements and relations already available. 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. Also, the representation of grammatical functions should not involve the multiplication of elements, rules, or principles (cf. the so-called Extended Projection Principle); it should optimally make use of the basic tools of the formal system already available to us and which have independently proven useful in empirical analysis. This is what we will now try to do.

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 <v_1, v_2> 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 organised in a *hierarchy* The RG hierarchy, as presented in Bell (1983: 148) is as follows:

55) 1 (Subject) >> 2 (Direct Object) >> 3 (Indirect Object)

In RG and APG there is a rich inventory of relational signs, of which we will only consider some. Specifically, and following Bell, we will consider *core* relational signs, and within these, *terms*: adjuncts and *chômeurs* (in APG and MG, these are elements that have been *demoted* to non-term relations) will be left aside. There is a reason for this, and it is that the syntax of non-core functions can be either finite-state (thus, ‘flat’) or higher level (context free, perhaps even mildly context-sensitive) depending on a variety of factors, as argued in Uriagereka (2008) and Krivochen (2015a, 2018). In contrast, the syntax of *terms* seems to be much more amenable to hierarchical organisation. In the hierarchy (55), *terms* outrank *non-terms*, and *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: 198, 250; Postal, 1982: 346 for APG translations of the RG hierarchy). Because the RG hierarchy is remarkably robust when it comes to the empirical analysis of English (see e.g. Postal, 2010) we will maintain not only the *core* functions in our model, but also assume that the hierarchy between them is correct.

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):

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

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 (57):

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

This means that the relation (a, b), where *a* dominates *b*, is ordered before the relation (a, c); in other words, (a, b) outranks (a, c). This is so not because of any property of the nodes, but because of the ordering imposed over the relations between them. The sceptical reader may wonder: what have we gained? Quite a bit, I would argue. If the order between dominance relations is made to correspond with the order between grammatical relations, we have the following relations, more or less formally[24]:

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

Let a = V, A, and b, c, d = N (and $b \neq c \neq d$), then *b* is the Subject / 1 of *a*, *c* is the Direct Object / 2 of *a*, and *d* is the Indirect Object / 3 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: we can go from one to the other no questions asked). 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,

---

[24] The picture of clause structure that emerges from (58) is necessarily oversimplified. It is not fatally so, though. A notion of strict order can, in principle, accommodate the richer array of Object relations recognised and extensively argued for in Postal (2010). A revised hierarchy, including the richer landscape of Object relations, could initially go along the following lines:

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

However, things are rather complicated. The interpretative rule in (58) 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-graded 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) italicised:

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

Even if the italicised NPs could be otherwise grouped under a certain category label, 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 scepticism about the putative advantages of including Postal's extended typology of objects into (58).

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; for a perspective much closer to MGG which still makes use of locally flat structures, see Emonds, 2007: Chapter 6). It is important to note, as Postal (1982) does, that the hierarchy in (52) holds for English (and, we add, for Spanish), but it must not be assumed that it is universal. What *is*, however, is the notion of *ordered grammatical relations*.

In this light, let us revise the set of dominance relations in (53), as (59):

59) ρ = ⟨(zag, *Jan*'); (zag, *Piet*'); (zag, helpen); (helpen, *Piet*'); (helpen, *Marie*'); (helpen, zwemmen); (zwemmen, *Marie*')⟩

The structure of (59) has *Jan* as the subject of *zag*, and *Piet* as its object. The relation *zag-helpen* is not contemplated in the functional hierarchy in (58) because we are not dealing with a V-N relation: the functional hierarchy in (55) and the interpretation of (49) from ordered relations in (59) pertain to functions that NPs or constructions with their functional potential (say, S') play in the A-structure of Vs. We can use a classic DG notion to make this somewhat clearer: *valency* (or *valence*), such that a predicative basic expression with valency *n* must head at least *n* edges whose tails are all assigned distinct GF in its local cycle G. Tesnière (1959: Chapter 97, §3) presents the notion with a chemistry analogy as follows:

> *The verb may* […] *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*$_{\mathrm{V}}$, but *John saw Peter*$_{\mathrm{NP}}$[25]. In *object raising* alternations of monotransitive Vs, it is the whole event (say, *John saw* [*Peter run*]; the aspectual distinction between the bare infinitive and the present participle –as in *John saw Peter running*- is not relevant here) 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*<zag, helpen>). But it is important to note that *helpen* does *not* receive a grammatical function in the minimal sub-graph that contains *zag*. We will come back to these issues in **Section 5**, which deals with the grammar of complement constructions.

4.3 A categorial excursus on existentials and unaccusatives

Before closing this section, it is important to complete the descriptive paradigm: it seems clear how (58) applies to monotransitive and ditransitive constructions, and intransitive unergatives: in RG terms, they all have an initial 1 arc. But we have said nothing about intransitive predicates without external arguments: *unaccusatives*. The problem posed by –some- unaccusatives is that they can appear in two alternations: with a pre-verbal definite or indefinite NP subject or a pre-verbal expletive *there* and an indefinite post-verbal NP with which the verb agrees. The paradigm is very well known, but we will illustrate it anyway:

60) a. $\begin{Bmatrix} A\ man \\ John \end{Bmatrix}$ arrived

[25] In the Standard Theory days, it was usual to have clausal complements S dominated by NP nodes (see, e.g., Rosenbaum, 1965; Ross, 1967). This theoretical device captured the same observation we make in the text.

b. There arrived $\left\{\begin{matrix} a\ man \\ *John \end{matrix}\right\}$

The question is, what is the ρ-set that corresponds to (60b)? Is it the same that corresponds to (60a)? If not, how do they differ?

We can be reasonably sure that the correct representation of (60a) is (61a):

61) a. ρ = ⟨(arrive, *man*’)⟩

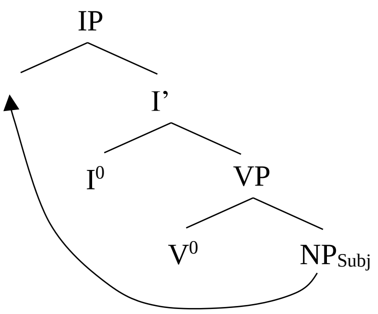
But (60b) is trickier: *prima facie*, it would seem that we need something like (61b) to represent the dominance relations between basic expressions in the expression:

61) b. ρ = ⟨(arrive, there), (arrive, *man*’)⟩

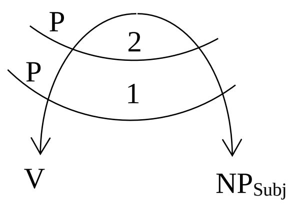
Where expletive *there* is assigned the GF *subject* by virtue of its position in the ordered ρ-set, and *man* being an object. Having *man* as an object actually captures at least an aspect of Perlmutter’s (1978: 160) *unaccusative hypothesis*:

Certain intransitive clauses have an initial 2 but no initial 1.

However, we need to consider that the *unaccusative hypothesis* was formulated within an RG framework, and that the initial 2 underwent *advancement* to 1, such that at a different *stratum* the 2 becomes a 1 (an initial object becomes a subject). Transformational generative grammar models this ‘advancement’ in terms of *movement*, such that an NP base-generated as the complement of the V moves to the Specifier position of Inflection / Tense to receive Case and satisfy the so-called Extended Projection Principle, which requires that the Specifier position of Inflection / Tense (Spec-IP / TP) be filled (Chomsky, 1982, 1995; Belletti, 1988). Both approaches are illustrated below:



Belletti (1988); Chomsky (1995)



Perlmutter (1978)

When a sentence features an overt expletive *there*, the EPP is satisfied by lexical insertion of the expletive (Perlmutter does not consider existential sentences with expletive *there*). In other frameworks, however, things are different. For instance, Arc Pair Grammar, RG’s successor, takes expletive *es* in a German sentence like

62) Es wurde hier getantz (German)
It became here danced
‘Dancing took place here’

to be heading an arc that *self-erases* (Johnson & Postal, 1980: 228); note that strictly speaking, *expletives do not have a grammatical function*, GF being assigned rather to their associates (thus the ‘advancement’).

We do not have *strata* or *movement* / *advancement* rules, so neither of these proposals is compatible with the framework presented here. But there are some things to learn from them, particularly RG/APG: expletives do not seem to be assigned a GF. Now, it is crucial to remember that nodes in our graphs correspond to IL translations of *basic expressions* of the language. In Ajdukiewicz-Montague style Categorial Grammar, expressions belong to indexed categories, and these categories indicate the combinatory properties of expressions: an expression of category X/Y needs to combine (concatenate) with an expression of category Y to yield an expression of category X. The set C of categories is defined as follows (see Schmerling, 2018a: 28, 1983b: 4):

C is the smallest set such that:

a, b, …n ∈ C (where *a, b, …n* are *basic expressions*)
a/b and *a\b* ∈ C (where *a/b* and *a\b* are *derived expressions*, and /, \ are abbreviations for the operations of *left* and *right* concatenation; see also Bach, 1979: 516)

What we can do now is rephrase our question slightly, and ask whether *there* is a basic expression of any category. If it is, then there must be a node in the graph that corresponds to the IL translation of *there*, but not otherwise. The reasoning that follows is loosely based on Krivochen & Schmerling (2018), adapted to the present context.

Let finite clauses be assigned to the category FC. An intransitive verb needs to combine with a noun phrase to form a finite clause; we will say that intransitive verbs are expressions of the category FC/NP[26]: in this way, we get *A man arrived* by left-concatenating a basic expression of category NP (*A man*) with a basic expression of category FC/NP. The question now is: what happens with *there arrived*? Well, in and of itself it is not a well-formed expression of category FC: we still need an NP, in this case *right-concatenated* to *there arrived* to form *there arrived a man*. This is important: if we segment the expression as *there arrived* [there [arrived]] and were to assign *there* to the category NP, given *arrived* FC/NP, that would mean that *there arrived* is a well-formed expression of category FC, contrary to fact. Now, this is not the place to discuss the nuances of an appropriate CG analysis, but we do want to highlight two things:

(a) Basic expressions do *not* correspond with orthographical words (i.e.: there are multi-word basic expressions)
(b) CG is flexible enough that the definition of categories in different languages need not be the same. This contrasts sharply with MGG's aprioristic universality.

Much research pending, we propose the following: expletive *there* is *not* assigned a GF because they are not basic expressions of any category. Rather, unaccusative verbs in existential alternations (that is, sequences *there + unaccusative V*) are expressions of category FC\NP: they need to right-concatenate with an NP to yield an expression of category FC. This category may be derived via a very limited set of rules like (63):

63) **Categories:**
FC (Finite Clause)
IV (Intransitive Verb Phrase)
PN (Proper Noun)
CN (Common Noun)

---

[26] We could equally have defined NPs as FC\IV (similarly to Schmerling, 1983b), but for the present purposes we take the verb to be the functor. In CG, the term *functor* designates the constituent of a derived expression whose category is written with a slash (X/Y or X\Y).

NP (Noun Phrase)
FC/NP (the category of expressions that left-concatenate with an NP to yield a FC)
FC\NP (the category of expressions that right-concatenate with an NP to yield a FC)

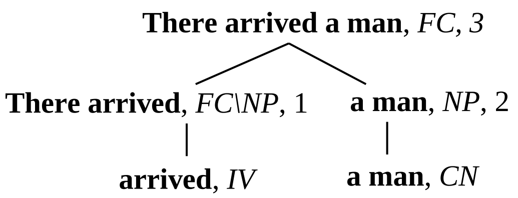
Basic expressions:
IV = {shines, walks, arrives, …}
PN = {John, Mary, Susan, …}
CN = {a man, a lamp…}

Formal operations:
$F_0(\alpha) = \alpha$, for all α.
$F_1(\alpha)$ = the result of concatenating **there** to the left of α, for all α.
$F_2(\alpha, \beta)$ = the result of concatenating α to the right of β, for all α, β.

Basic rules:
S_0. If $\alpha \in P_{IV}$, then $F_0(\alpha) \in P_{FC/NP}$, for all α.
S_1. If $\alpha \in P_{IV}$, then $F_1(\alpha) \in P_{FC\backslash NP}$, for all α.
S_2. If $\alpha \in P_{CN}$, then $F_0(\alpha) \in P_{NP}$, for all α.
S_3. If $\alpha \in P_{FC\backslash NP}$ and $\beta \in P_{NP}$, then $F_2(\alpha, \beta) \in P_{FC}$ for all α, β.

A Montague-style analysis tree for *there arrived a man* would then look along the lines of (64), assuming the toy grammar above:

64)



We have oversimplified the grammar, of course, but we believe our point has emerged relatively unscathed: the operation that introduces *there* yields an expression of the derived category FC\NP, which still needs to combine with an expression of category NP to yield a FC. What this means for us is that there is no node that corresponds to *there* in the ρ-set of *there arrived a man*. The correct structural description for this string is rather (65), ignoring matters of inflection as usual:

65) *There arrived a man*
ρ = ⟨(arrive, *man*')⟩

This analysis implies that *there* is a *syncategorematic* element: it has no meaning of its own, but every expression in which it occurs does have a meaning (see Schmerling, 2018a: Appendix A). The distinction between *categorematic* and *syncategorematic* is an ancient one, as the following quotation from Priscian illustrates:

> *…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 consignificantia, appellabant* (Priscian, *De Partibus* §54.5)

Quite more recently, Schmerling (2018a: 44, fn. 10) very clearly summarises 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

Categorematic elements are assigned to syntactic categories and as such can play GF in syntactic construal, but syncategorematic elements cannot; in the present model this means that nodes in a graph correspond only to *categorematic* elements (which can be multi-word expressions, as pointed out above). As desired, in (65), as in (61), the categorematic NP *a man* is assigned the GF Subject, by (58). *There*, however, does not belong to a category and does not receive a GF, also as desired.

Still in the spirit of comparing models, it is crucial to point out that CG analysis trees and our graphs give very different information. Note also that categories in CG are more informative than syntactic labels in PSGs: given the interpretation of the slash-notation introduced above, if we know that an expression is of category X/Y and that one of its constituent expressions is of category Y, we can deduce that the category of the other constituent expression must be X; we do not have that information in category labels like VP or NP. As pointed out by Schmerling (2018a: 28-29) an important emphasis of Ajdukiewicz (1935) is that his CG allows us to *discover* previously unknown categories of expressions; this is referred to as the *heuristic* character of CG. Here, we have taken advantage of that *heuristic* to determine whether a sequence of words constituted a basic expression or a derived expression (the definition of the category is a plus that is not needed for the study of connectivity patterns), which in the present framework translates into that sequence being a single node or not. In a way, then, pure-CG and the graph-theoretic approach argued for here are complementary: CG analysis trees are proofs that expressions belong to categories of the language; they do not represent dominance or connectivity between nodes. Strictly speaking, trees could be replaced by proofs in a different format, for example:

A man is a basic expression of category CN
A man is a well-formed expression of category NP, by the application of rule S_2

This is exactly equivalent to the right side of the tree in (64) (Montague, 1973: 227). In contrast, our theory is a theory about *graphs*, which must not be confused with *diagrams of graphs*: graphs are sets of nodes and edges, diagrams are graphical/typographical tools which make use of lines and letters.

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 if the size of elementary trees is limited by something like Frank's 1992 CETM). **Sections 5** and **6** will come back to this constraint and the kind of structures it allows for.

5. <u>On English Predicate Complement Constructions</u>

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_{fin}+NP+V_{inf}. The classification of so-called 'predicate complement constructions' has been a concern since the early days of Generative Grammar in both empirical and theoretical fronts: Rosenbaum (1965) distinguishes two kinds of such constructions, classified essentially in terms of VPs taking either NPs dominating Ss or VPs as complements; Perlmutter (1968) and Postal (1974) substantially expand on and modify this proposal. These foundational works, which set the standard for subsequent generative analyses (e.g., Chomsky & Lasnik, 1977; Kayne, 1981; Lasnik & Saito, 1991; see Davies & Dubinsky, 2004, 2007 for extensive discussion), differentiate between the following kinds of non-finite complements:

66) a. Subject-to-Subject raising
b. Subject-to-Object raising (a.k.a. Exceptional Case Marking ECM)
c. Object-controlled Equi
d. Subject-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 an adequate grammatical theory for the English language is that it correctly distinguishes between these structures and assign distinct structural descriptions where applicable. We will proceed in the order indicated in (66), starting from Subject-to-Subject raising.

5.1 Raising to subject

Consider the following example:

67) 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?*', or, in other words, is (67) a monoclausal or a biclausal structure? The traditional view in transformational generative grammar is that in (67) 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. This analysis is also applied to predicative constructions like (68):

68) John seems sick

However, things work rather differently in adjectival predicative constructions and infinitival complement constructions from a semantic point of view: only in (68) *seems* behaves as an evidential expression. Let us first deal with examples like (67), and then briefly comment on (68).

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, 1994: 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 (see Wurmbrand, 1999 for the exact opposite conclusion); as such, *seem* is part of the Cycle in which the modifiers and arguments of *read* are expressed (crucially, this also holds in Wurmbrand's view). The map of connectivity relations in (67) is as follows:

69) Cycle 1: [John seems to have read the book]
ρ = ⟨(seems, *John*'); (seems, read); (to have, read); (read, *John*'); (read, *book*')⟩

The relations between *read, John*', and *book*' illustrate the interpretation rule in (58). *John* is the subject of *seem* and of *read*, but, crucially, there is *no relation* between *John* and the perfective auxiliary *have*, since *have* simply modifies the lexical verb that takes *John* as an argument.

How, if in any aspect, does the ρ-set of (68) differ from that assigned to (67), in the light of their semantic differences? This question is important insofar as it pertains to the level of granularity to which our proposal applies: to what extent should graphs (and, more generally, structural descriptions) represent aspects of lexical meaning? A possible answer is simply 'to the extent that those aspects are determined configurationally'. A follow-up question arises immediately: is *evidentiality* a configurational relation or a lexical specification? Here we will assume the latter, in the sense that it is not a property of the connections that *seem* establishes with other nodes in a local graph. Evidentiality in English is related to modality, although, as Aikhenvald (2004) correctly observes, evidential expressions can co-occur with modals, which makes it difficult to say that evidentiality is a subset of epistemic modality (cf. Palmer, 1986), given the well-known restrictions that English sets on modal expressions. This is important because if *seem* is a non-grammaticalised evidential marker in at least some configurations, then it can be assimilated to modifiers of a predicative expression; in that case, lexical verbs modified by *seem* would determine only one cycle. Asudeh & Toivonen (2017: §3.1) indeed discuss English copy-raising and *flip* verbs (Lakoff, 1965: A-15) as marks of non-grammaticalised evidentiality, which would support an analysis in which raising predicates are modifiers of the lexical predicate that selects the nominal arguments in a single-rooted elementary graph (see **Section 3** above for a brief comparison with Frank's 1992, 2002 Condition on Elementary Tree Minimality, formulated within a TAG framework). Recall that in **Section 3** we argued that each *cycle* in a structural description is a single-rooted graph which contains the following elements (based on García Fernández & Krivochen, 2019):

a. A predicative basic expression *p*
b. Temporal and aspectual modifiers of *p*
c. Nominal arguments of *p* (subjects, objects, non-argumental clitics)

The crucial thing here is to determine whether a predicate takes nominal arguments or simply modifies another predicate, one which does take arguments. We argue that raising predicates are modifiers of the event-denoting predicate, which do not take nominal arguments. Thus, there is no need to assume a multi-cyclic structural description in constructions of the kind *raising predicate + lexical predicate*: nominal arguments are arguments of the lexical predicate, and the raising predicate is a (modal, evidential) modifier of that predicate.

In the light of the preceding discussion, the ρ-set of (68) would be:

70) ρ = ⟨(seems, *John*'); (seems, sick); (sick, *John*')⟩

As in (67), theta and agreement requirements are satisfied by different predicates, but we see no evidence to propose the existence of a further cycle. Evidentiality notwithstanding, there doesn't seem to be a difference in *connectivity* between (67) and (68); differences pertain to aspects of lexical structure that are at a smaller level of granularity than that we are concerned with here.

We may note that the relations in (70) can also be extended to other predicates, like *remain* or *be likely* in sentences like (71a) and (71b) respectively:

71) a. Her bank account remained active after her death
   b. Mary is likely to get that job

Note that because we allow for multi-word *basic expressions* (with Schmerling, 1983b) and nodes correspond to the IL translation of basic expressions (not to orthographical words), the analysis of *be likely* is in all respects identical to that of *single-word* raising predicates, the segmentation we argue for is certainly *not* [$_{\text{VP}}$ be [$_{\text{AP}}$ likely]] or any such extension of X-bar theory.

5.1.1 <u>Copy-raising in a multidominance framework</u>

An analysis of Subject-to-Subject raising that allows for multidominance has additional advantages when we look at so-called *Richard*-constructions (Rogers, 1971, 1972, 1974). ‘Richard’ is a syntactic copying transformation (in the sense of Ross, 1967) first formulated by Rogers (1971: 217) and worked out in subsequent publications by him, which relates sentences like (72) and (73)

72) It $\left\{\begin{matrix} seems \\ looks \\ sounds \end{matrix}\right\}$ like Richard is in trouble

73) Richard $\left\{\begin{matrix} seems \\ looks \\ sounds \end{matrix}\right\}$ like he is in trouble

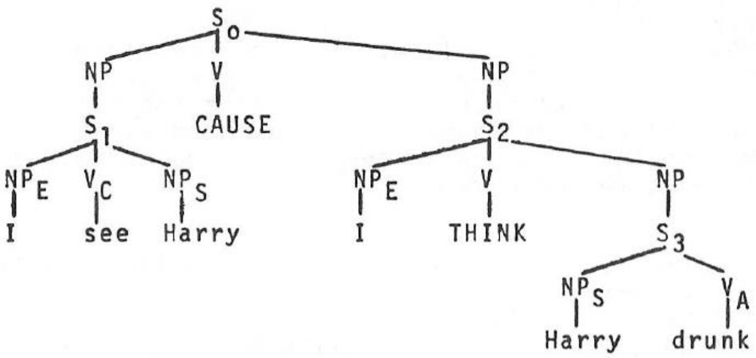
The transformation…

> *…copies the subject NP of the <u>like</u>-clause in the structure underlying sentences such as those in* [63] *as subject of the main clause, resulting eventually in sentences such as those in* [64] (Rogers, 1974: 551)

In more modern, but equivalent terms (Potsdam & Runner, 2001; Asudeh & Toivonen, 2012), a *raising to subject* rule that leaves behind a resumptive pronoun is referred to as *copy raising*. It is important to note that the analyses of Rogers (1972) and more recently Asudeh & Toivonen (2012) and Asudeh (2012) are *biclausal*: in Rogers’ paper, the Deep Structure that corresponds to a Richard-sentence contains *two* S nodes; in Asudeh & Toivonen’s LFG-based works (following Bresnan, 1982), *raising* constructions are XCOMP, that is, internally insaturated complements. Crucially for the LFG analysis, copy-raising Vs are single-word terminals, which take *like* complements (Asudeh & Toivonen, 2012: 325).

In the terms we are working with here, a biclausal analysis would translate into two initial sub-graphs which are linked at the ‘copied’ node, whose IL translation would be *Richard*’ in (73). However, if we take (copy-)raising and *flip* verbs to be evidential markers, there seems to be no support for proposing more than a single cycle here either: the raising predicate is a modifier of the lexical verb that assigns grammatical functions to nominal expressions. Our objections to that proposal notwithstanding, there are some aspects of the biclausal analysis that deserve some discussion. In Roger’s analysis, one sentential S node corresponds to the perception involved in *flip* Vs, which generates *sui generis* presuppositions, and the other to the object of perception (which is presupposed). The syntactic-semantic structure that Rogers (1972: 310) assigns to the sentence *Harry looks (to me) like he was drunk* is (74):

74)



If we were to translate Rogers' structure *directly* into a maximally connected graph, there is quite a bit of pruning (Ross, 1969b) we can do. First, we can delete all intermediate nodes. Then, we can identify identical nodes, such that *I* and *Harry* need not be distinct. But there is a more fundamental problem, which is the presence of phonetically empty semantic primitives (as was usual in Generative Semantics; see McCawley, 1968; Dowty, 1979): here CAUSE and THINK. There is no more justification for those than for intermediate phrasal nodes, which means that a more radical restructuring of the structural description must take place. Before doing that, however, we need to distinguish between different classes of Richard (and Richard-like) constructions, which will require tweaks to the structural descriptions we assign each of them. Initially, we can identify two big classes of Richard-constructions: *surface Richard* and *true Richard*. Let us characterise them in some detail.

Surface Richard constructions:

Surface Richard constructions are, predictably, sentences which only look like they are instances of Richard, but have syntactic-semantic properties which set them apart from true Richard constructions. Interestingly, languages which do not feature Richard, like Spanish, do allow *surface* Richard constructions, which suggests that there is further ground to draw the distinction.

A. *Active constructions without internal experiencer:*

75) Richard acts (*to me) like he's in trouble
76) *It acts (to me) like Richard's in trouble

Here we have copying without raising (the matrix verb is a thematic assigner) and without perception either. The structure is thus biclausal, if by 'clause' we understand a unit of predication with a thematic assigner. If each of the predicates select nominal arguments, then, by the definition of *extended projection* that we introduced in **Section 2** and repeated above, we should have a sub-graph *per* predicate.

The interpretation of (75) can be either progressive or habitual, by virtue of 'act' being a dynamic verb. Note that the expletive *it* alternation is ungrammatical, cf. (72-73). The ρ-set for (75) is (77):

77) Cycle 1: [Richard acts like]
Cycle 2: [Richard is in trouble]
ρ_1 = ⟨(act like, *Richard*')⟩
ρ_2 = ⟨(be, in trouble); (in trouble, *Richard*')⟩

B. *Middle constructions (without internal experiencer):*

78) This bread cuts (*to me) like it's stale (does not licence Richard; non-factive)
79) *It cuts (to me) like this bread's stale

Non-Flip Vs do not license an experiencer, that is straightforward enough. In this case, the interpretation is clearly non-active, it's a state; this is the main difference with the previous class. It refers to an individual-level property of the bread. The ρ-set for (78) is (80):

80) ρ = ⟨(cut like, *bread*'); (be, stale); (stale, *bread*')⟩

C. *Middle constructions (with internal experiencer):*

Examples like (81), taken from Rogers (1974: 552), differ from those like (78) in that it is possible to have an internal experiencer:

81) The soup tastes (to me) like Maude has been at the cooking-sherry

We need to note that the subject position of the lower clause is not occupied by a resumptive pronoun, but by a referential NP: it is *not* possible to claim that (81) has been derived by a *copying* rule (a fact also pointed out by Rogers, 1974: 552). These verbs, which allow for disjoint reference between the subjects of the perception verb and the embedded verb are referred to as *perceptual resemblance verbs* (Asudeh & Toivonen, 2012: 325): for these cases, a biclausal analysis seems to be more strongly supported. Let's thus tweak the example, and make both subjects co-referential:

82) The soup tastes like it's been left cooking for too long

This is still *not* a Richard sentence: we cannot have the expletive-it alternation

83) *It tastes like the soup has been left cooking for too long

Maude-sentences can (but crucially, do not need to) display at least one of the defining properties of Richard-sentences, which is the copying procedure that gives us co-referentiality; as well as the presence of an experiencer *to*-phrase. However, we have seen that there is also at least one defining property of Richard-sentences which Maude-sentences do not display: the *expletive-it* alternation. This suggests that we need to make the conditions under which Richard is licensed stricter. This is what we will do next.

True Richard constructions:

Middle constructions with internal experiencers:

84) Richard looks/sounds/seems to me like he's in trouble (licenses Richard; non-factive)
85) It looks/sounds/seems to me like Richard's in trouble

Now we formulate the additional requirements for a construction to be identified as a Richard-construction:

a) It must be possible to have the *expletive-it* alternation (cf. (83); (84))
b) The subject of the perceptual resemblance V and the subject of the complement clause of that V *must* be coreferential

Requirement (b) holds for NPs, which can bear referential indexes; but some additional discussion is in order. Consider the following example, from Rogers (1974: 551) –recall that, at this point in the theory, expletive 'there' was transformationally inserted-:

86) There looks like there's gonna be a riot

Expletives have no referential indexes, so requirement (b) –co-reference- cannot be satisfied. This means that expletive 'there' cannot have an index or, rather, that it has an index 'Ø', as a way of indicating that it has no reference. But there is a unique element in the indexing alphabet that can have no reference in context C, meaning that any *n* instances of 'there' in a cycle must be 'co-referential' in that they share the prohibition of having a referential index.

More practically: saying that 'the subjects *cannot have disjoint reference*' avoids the problem altogether, because even if we get the expletives 'it' and 'there', they do not (in fact, they *cannot*) have disjoint reference by virtue of there being a single element in the indexing alphabet with no reference in C; and thus we need not invoke null indexes.

Requirement (b) captures Rogers' and Ross' characterisation of Richard as a *copying transformation*. The requirement for co-referentiality comes from the fact that Richard is essentially a *copying* rule. If we copy an NP and leave behind a resumptive pronoun, it will bear the same index as the NP. Without such a requirement, and as Rogers (1974: 551) points out, we could not adequately exclude (87b) as the result of copying applied to (87a), in the idiomatic sense:

87) a. The shit$_i$ looks like it's gonna hit the fan.
b. *The fan$_i$ looks like the shit's gonna hit it$_i$.

(87b) cannot be filtered unless we assume that Richard is a copying rule: only in this way is the idiomatic unit [the shit hit the fan] preserved. (87a) is grammatical precisely because there is a reordering copying rule which displaces [the shit] and leaves behind a copy in the form of a resumptive pronoun. Moreover, (87b) provides further evidence that Richard as a transformation needs to be limited to *subjects*: this follows if Richard-Vs are raising Vs and if an analysis of raising as multidominance is along the right lines: the subject is dominated by the Richard-V, with which it agrees in person and number, and by the embedded predicate, where it is thematically interpreted. Under these conditions, co-referentiality is guaranteed, as desired.

5.2 Raising to object

We have seen some basic aspects of *Raising-to-Subject* constructions. Next up in (66) are *Raising-to-Object* (a.k.a. Exceptional Case Marking, ECM) 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):

88) a. Mary saw John run
ρ = ⟨(see, *Mary*'); (see, *John*'); (see, run); (run, *John*')⟩

b. The judge believed John to have committed the crime
ρ = ⟨(believe, *judge*'); (believe, *John*'); (believe, commit); (commit, *John'*); (commit, *crime*')⟩

This is a much simpler example than the crossing dependencies we saw in (48) (the famous example *Jan Marie Piet zag 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 (58)) *and* the embedded V *run*. As we said above, *run* itself does not satisfy the valency of *see*, rather, the clause containing *run* and its subject *John* does. 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 *object raising* structures are very similar to those of structural descriptions assigned to small clause complements, as in the examples in (89):

89) a. They consider John an idiot
b. John wants Bill dead

c. I like my tea strong

The ρ-sets for the sentences in (89) are given in (89'):

89') a. ρ = ⟨(consider, *they*); (consider, *John*'); (consider, idiot); (idiot, *John*')⟩
b. ρ = ⟨(want, *John*'); (want, *Bill*'); (want, dead); (dead, *Bill*')⟩
c. ρ = ⟨(like, *I*); (like, *tea*'); (like, strong); (strong, *tea*')⟩

The main difference between *object raising* and small clause complements for purposes of this work is that in *small clause complements* the accusative NP is a 2 of the embedded as well as the matrix predicates, there is no need to resort to *demotion* or any such operation: the accusative NP is an object in both clauses *always*. However, this is only the case in ECM clauses when the embedded predicate is unaccusative or ergative (that is, when the predicate does not have an initial 1 arc; see Perlmutter, 1978: 160), or passive.

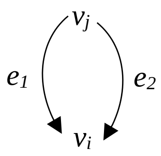
We follow Rosenbaum (1965) and particularly Postal (1974) in considering the accusative NP in ECM and small clause (secondary predication) structures to be an argument of the embedded predicate, which gets *raised* to an object position in the matrix clause. In the present model, this simply means that there is an edge from the matrix predicate to the accusative NP, not that there has been any kind of 'movement' (see also Postal, 2010). 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, and in which the NP is a 'constituent' of both the matrix and the embedded clauses. In the terms we have been working with here, this means that *object raising* structures involve a *single cycle*. This is desirable for several reasons, of which we will briefly review two: anaphora licensing and locality conditions (the latter, in comparison with Object-controlled Equi constructions, in the following section).

5.2.1 A note on reflexive anaphora

There are two kinds of examples of reflexivity that we are interested in for purposes of the present section, which we illustrate in (90) and (91):

90) The defendants proved themselves innocent during the trial
91) *John believes Mary to like himself

These cases are interesting insofar as they allow us to test some preliminary considerations about binding, which we will return to in detail in **Section 7**. What we are interested in here is that the claim that *object raising* structures are single cycles, with the additional condition that nodes in a cycle be *strictly ordered*, yields a descriptively adequate account of sentences like (90) and (91). The first consideration we need to make is that *the defendants* and *themselves* in (90) and *John* and *himself* in (91) correspond to subsequent visits to the same node in a trail; if it is possible, there is a graph that corresponds to a well-formed expression of the language. The question arises, at this point, whether there is a legitimate trail in a single-rooted graph, such that the adequate structural descriptions for the licensing of reflexive and reciprocal elements. In order to answer this question, we need to consider a proper characterisation of 'reflexivity'. Here we will follow Reuland & Reinhart (1993), and define reflexivity as a property of *predicates* rather than of arguments: a predicate is reflexive if two of its arguments, the external and an internal argument, are coindexed. Concretely, we will make use of the relation between edges '*parallel*', defined in **Section 2** and repeated here as (92) for the reader's convenience:

92) 

<u>Parallel</u>: edges e_1 and e_2 are parallel iff e_1<v_i, v_j> ∧ e_2<v_i, v_j>. In this configuration, v_i dominates v_j establishing two different relations with it.

Let v_j be a dyadic predicate, and v_i a nominal argument of that predicate. Note that the dominance relations in (92) translate into ordered notation as follows:

93) $\rho = \langle (v_j, v_i), (v_j, v_i) \rangle$

As per the interpretative rule (58), this means that v_i is *both* the Subject and the Object of the predicate v_j; the predicate in node v_j is thus *reflexive* in the sense of Reuland & Reinhart (1993). Within our graph theoretic model, then, we can define *reflexivity* in terms of the *parallel* relation between edges, such that a simplex sentence (in the sense of Lees & Klima, 1963) contains a reflexive predicate if and only if the structural description assigned to that sentence features a *parallel* relation between edges. The centrality of *parallel* is due to the fact that this allows a single node to be in two different grammatical relations with another node. In Postal's (2010: 18) terms,

> *one good reason to assume parallel arcs is that for certain types of grammatical relations a single phrase has the possibility of bearing more than one to the same larger constituent.*

Note that this advantage holds even if no notion of *immediate constituency* is recognised in the formalism, just predicates and arguments.

So far, we have been dealing with monoclausal sentences, in which we have only one lexical predicate. Can we extend now these considerations to more complex structural descriptions, containing more than a single lexical predicate? Here is where (90) becomes relevant. Its ρ-set is as follows:

90') ρ = ⟨(prove, *defendants*'); (prove, *defendants*'); (prove, innocent); (innocent, *defendants*')⟩

The considerations we have made about the relevance of *parallel* to the representation of reflexive predicates extend without the need to resort to further theoretical devices to sentences with more than a single predicate, as in the case of *object raising*. Note that in (90) *defendants*' is the Subject and Object of *prove*, here a reflexive predicate, and also the Subject of the embedded adjectival predicate *innocent*. If we had chosen a verbal predicate to be embedded instead of an adjective, the structure would not modify the conditions under which reflexivity is licensed in the simplest English cases (see **Section 7** for further discussion about Binding Theory within the present framework). The crucial thing to bear in mind is that *object raising* structures define a single cycle, *ceteris paribus*.

A look at the ρ-set of (91) is now in order. Note that in this case the anaphoric element is the Direct Object of the embedded predicate, which means that this predicate must be reflexive. Concretely, it means that, in order to license a *self*- object, a configuration like (92) must be an adequate characterisation of the embedded clause. Let *himself* in (91) be interpreted as coindexed with *John*. In that case, what we have in (91) is the following

94) ρ = ⟨(believe, *John*'); (believe, *Mary*'); (believe, like); (like, *Mary*'); (like, *John*')⟩

Note that by virtue of *Mary* and *John* being distinct nodes, we cannot say that the embedded predicate *like* establishes two distinct relations, by means of two distinct edges, with a single node. In other words, there is no *parallel* relation in the structural description of (91). The reflexive anaphora is not licensed due to the absence of *parallel edges*, and (91) is thus adequately excluded as a well-formed expression of the language.

5.3 Object-Controlled Equi

Let us now focus on the following structure in (66), *object-controlled Equi*. The reason we are dealing with them after *object raising* is that there are some superficial parallels between them, in the sense that they are linearly sequences of the kind $NP_{Nom} + V_{fin} + NP_{Acc} + V_{inf}$. It is important to be able to distinguish between *object-controlled Equi* structures and *object raising* structures both syntactically and semantically in a descriptively adequate theory of the grammar.

It is interesting to note that already in Rosenbaum (1965: 10) we see some relevant conditions formulated over the rule *Identity Erasure* (better known as *Equi NP deletion*) which operate over *paths* in tree-like graphs rather than over strings (or sequences of variables, cf. Lees & Klima's, 1963 formulation of the *pronoun* and *reflexive* rules):

A NP_j is erased by an identical NP_i if and only if there a S_α such that

- *i) NP_j is dominated by S_α*
- *ii) NP_i neither dominates nor is dominated by S_α*
- *iii) For all NP_k neither dominating nor dominated by S_α the distance between NP_j and NP_k is greater than the distance between NP_j and NP_i, where the distance between two nodes is defined in terms of the number of branches in the path connecting them* (Rosenbaum, 1965: 10)

The importance of the node S_α should not be underestimated: the clausal root S was a cyclic node in the Standard Theory, which in turn played a central role in the definition of locality conditions: reordering transformations which established dependencies across cyclic nodes were in general blocked (see, e.g., the *Subjacency Condition* in Chomsky, 1973: 247[27]), and in order to get around this strict cyclicity in long distance extraction, the theory had to incorporate a notion of *punctuated movement*. Relations between nodes could take place (i) within the boundaries of cyclic nodes (their Specifiers, in X-bar parlance: if XP is a cyclic node, then a phrasal node immediately dominated by XP is accessible for operations triggered outside XP), thus being *local* (e.g., the kind of configurations that license anaphora or NP raising), or (ii) across cyclic nodes, thus being *non-local* (e.g., Wh-movement; see Postal, 1972). For purposes of the present work, two structural conditions are particularly relevant: *non-locality* and *distance*. The former implies that in the kind of deletion configurations that we are dealing with here, *Subject*- and *Object-controlled Equi*, there is more than a single cycle. This is already a feature that differentiates *object raising* from *Equi*. The latter pertains, in Rosenbaum's account (and, truth be told, in any framework in which nodes correspond to words or phonologically defined basic expressions, including APG and MG) to the structural configuration in which co-indexing takes place (see also Rizzi, 1990 and subsequent work). Let us see how the present work deals with these conditions.

---

[27] The original formulation of Subjacency is the following:

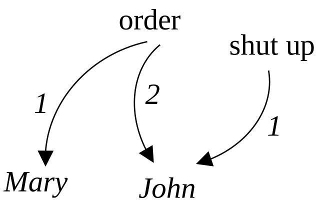
> *If X is superior to Y in a phrase marker P, then Y is subjacent to X if there is at most one cyclic category* [NP / S] *$C \neq Y$ such that C contains Y and C does not contain X.* (Chomsky, 1973: 247)

Structural descriptions assigned to *object raising* structures in our graph-theoretic account are different from those assigned to *Object-controlled Equi*, as desired: the ρ-set for an Object-controlled Equi sentence like (95) is given below:

95) Mary ordered John to shut up
Cycle 1: [Mary ordered John [Cycle 2]]
Cycle 2: [John shut up]
ρ_1 = ⟨(ordered, *Mary*'); (ordered, *John*')⟩
ρ_2 = ⟨(shut up, *John*')⟩

There are several things to note here. First, note that we are dealing with two cycles, not just one: there are two single-rooted sub-graphs, connected at the node *John*'. This is consistent with the traditional generative observation that Equi constructions are *biclausal*, as opposed to *raising* structures. Second, note that the relations 'object of *order*' and 'subject of *shut up*' are indeed defined in different cycles, but there is no need to multiply the nodes because grammatical functions are read off the ordered ρ-set. Alternatively, we could annotate the edges *à la* APG, an option already considered in **Section 4**:

96)



Unlike in APG / MG, however, there is no multiplication of nodes here, just like there is not for pronouns and anaphors. The diagram above also helps see that, whereas there is a strict ordering within each cycle (such that every node in each cycle is ordered with respect to every other node within that cycle), this strict ordering *does* not extend beyond the limits of a cycle: *order* and *shut up* are not ordered with respect to one another. This is a potential ambiguity in the structural description that gets avoided by our definition of *cycle*.

It is also important to note that the cycles are *linked* (in the technical sense) at *John*, which eliminates the need for additional operations: there is no Equi-NP-deletion under a view of the grammar where cycles are connected at identical nodes. The framework explored here allows us to capture Rosenbaum's locality condition as well as the fact that in Equi structures the controller and the subject of the non-finite form can have distinct GF (as in Object-controlled Equi) by defining GF in terms of ordered relations in *digraphs*. In (96), for instance, we see that *John* is the object of *order* and the subject of *shut up*, where each of these predicates defines a cycle: (96) is an example of a multi-rooted graph.

Finally, and in contrast to our *object raising* structural descriptions, we may note that the matrix V *order* only dominates *John*', but 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 object raising: this correlates with well-known extraction asymmetries between these constructions, exemplified in (97):

97) a. John wanted the guests to leave → $\begin{array}{c} What\ did\ John\ want? \\ *Who\ did\ John\ want\ to\ leave? \end{array}$

b. John ordered the guests to leave → $\begin{array}{c} ??What\ did\ John\ order? \\ Who\ did\ John\ order\ to\ leave? \end{array}$

c. John saw Mary cross the street → $\begin{matrix} \mathit{What\ did\ John\ see?} \\ \mathit{Who\ did\ John\ see\ cross\ the\ street?} \end{matrix}$

d. John told Mary to finish the report → $\begin{matrix} *\mathit{What\ did\ John\ tell?} \\ \mathit{Who\ did\ John\ tell\ to\ finish\ the\ report?} \end{matrix}$

Object raising 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, by virtue of the matrix and embedded clause being *linked* at the node corresponding to the IL translation of the relevant NP. In MGG, control structures require the introduction of a null category PRO as the subject of the infinitive, which is controlled by the object of the matrix clause. The motivations for the abandonment of Equi in favour of a base-generated empty category are not always clear, although in GB/MP theta-theory and Case are frequently invoked: because an index cannot be assigned more than one thematic role in the course of the derivation (Chomsky, 1981: 139, fn. 15; see Jackendoff, 1987 for discussion), and given the fact that control verbs are thematic assigners, the incorporation of an empty category PRO which can be assigned a theta role within the embedded clause and be controlled by an NP with an independent theta role from the matrix clause seemed to be a good move. However, the requirement that an index cannot be assigned more than one theta role derivationally is seldom justified, and in fact it has been challenged in the so-called *Movement Theory of Control* (Hornstein, 1999): in this view, theta roles are features, which an NP can check in several structural positions in which it has landed via punctuated movement. All in all, the multiplication of entities is a problem in both proposals: in one, we have a new element in the lexicon (PRO), plus a set of principles that govern its distribution (Control Theory, see Bouchard, 1982: Chapter 5 for useful early-GB discussion), plus an indexing mechanism; in the other, we have traces / copies left behind by NP-movement, plus indexing, plus an operation to selectively delete copies at Phonetic Form (the movement operation itself is not an innovation of the Movement Theory of Control, and therefore while it is an overall complication of the theory, it is not the MTC's fault).

The relevant IC segmentations for a surface string $V_{fin}+NP_{Acc}+V_{inf}$, which can correspond to either *object raising* or *object-controlled Equi* are as follows (see Rosenbaum, 1965 for a classic argument in terms of distinct *Deep Structures* for *object raising* and Object-controlled Equi; McCawley, 1998: 133, ff. for descriptive discussion and examples galore):

98) a. [… V_{fin} [NP_{Acc} V_{inf}]] (*object raising*)
b. [… V_{fin} NP_{Acc} [PRO V_{inf}]] (Object-controlled Equi)

We think that our approach has several advantages over MGG-IC analyses. 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; furthermore, by having nodes stand for IL translations of basic expressions, we also eliminate the need for an independent indexing operation. These changes notwithstanding, *object raising* 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).

5.4 Subject-Controlled Equi

We have deliberately left Subject-controlled Equi (a.k.a. 'Control') for last, because of the problems that its analysis presents to syntactic theory; Equi is trickier than it seems. 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* (and indexing) 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, the complement is *not* open: we don't have XCOMP but COMP, with a phonologically empty category 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) where there is no need to have the same *f-structure*. The subject of the matix clause and the subject of the embedded clause are semantically related, but *are not the same syntactic object* (as opposed to the Raising cases). Relevantly, the treatment of Equi using *anaphoric control* is not uncontroversial: whereas Dalrymple (2001, 2007) defends an *anaprohic control* analysis, Falk (2001) and Asudeh (2005), among others, propose an analysis of Equi which involve *functional control* (see Bresnan, 2001: 297, ff. and Dalrymple, 2001: 325, ff. for an accessible introduction to *anaphoric* vs. *functional control*). Crucially, as Dalrymple (2007) points out, the answer to the question whether an adequate analysis of Equi involves *anaphoric* or *functional* control (and even if the same analysis is valid for all cases) is not given by the theoretical framework of LFG, but by the linguistic facts. This is important because it allows for different languages to follow different strategies when building predicate complement constructions; given the strong empirical focus of the present work this is clearly a view we are sympathetic towards.

The reason we have spent some time with LFG's analysis of Equi is that we need to ask the question whether *subject raising* and *subject-controlled Equi* should receive distinct structural descriptions in terms of the number of cycles or the connectivity patterns in the graphs or whether the difference can be codified in lexical specifications (at least in English; as pointed out above, it is possible and theoretically consistent that other languages may choose to encode the distinction by different means). In the latter case, these specifications would 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 either Raising or in Equi, but their differences arise as a matter of *lexical semantics*. Dowty's take on lexically governed rules seems 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 in terms of lexical specifications is precisely the kind of framework that complements the one sketched here perfectly; we will come back to this in **Sections 12** and **13**. As highlighted in **Section 3** above, the conception of the lexicon in Categorial 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 IL translations of basic expressions in the algebra of a language L (in our case, English).

The question now is whether we can extend this analysis to *raising* and *subject controlled equi*. Above, in **Section 5.1**, we argued that *subject raising* structures receive a monoclausal analysis, which translates into there being a single-rooted graph as the structural description for the relevant expressions of the language. The argument was, partly, that raising verbs are (in specific contexts) modifier expressions of lexical predicates rather than lexical predicates themselves, therefore, they are part of the cycle defined by that lexical predicate. Can we say the same for *equi*-governing verbs? Consider examples (99) and (100):

99) Mary *tried / wanted* to finish the paper (Subject-controlled Equi)
100) Mary *was likely / seemed* to finish the paper (Subject-to-Subject raising)

The question is, then, whether *try* is a modifier of *finish* or a different predicate altogether; if the latter, then we need to propose a multi-rooted structure linked at the node corresponding to the subject NP, which is the subject of both predicates.

Unlike *raising* predicates, *equi* predicates impose restrictions with respect to their co-occurring NPs; this much is already a traditional observation. (101a) is ungrammatical not because *winter* cannot *arrive*, but because it cannot *try*:

101) a. *Winter tried to arrive early this year
b. Winter seemed to arrive early this year

In the present context, this means that the subject NP is a nominal dependant of the matrix predicate as much as it is a nominal dependant of the embedded predicate: therefore, we need to conclude that each predicate, the *equi-governing* V and the embedded V define a cycle each, as per the discussion in **Section 3** and **5.1** above. In the light of this discussion, then, we can specify the ρ-sets for (99) and (100) as follows:

99’) ρ = ⟨(try, *Mary*’); (finish, *Mary*’); (finish, *paper*’)⟩
100’) ρ = ⟨(be likely, *Mary*’); (be likely, finish); (finish, *Mary*’); (finish, *paper*’)⟩

Only (100’), the ρ-set of a *subject raising* sentence, defines a single-rooted structure, with a root *be likely*. (99’), the ρ-set of a *subject controlled equi* sentence receives a different structural description, with two cycles linked at *Mary*’.

We will now analyse some more complex cases. Consider first the following sentence:

102) Melvin expects to want to win (Johnson & Postal, 1980: 540, ex. (164))

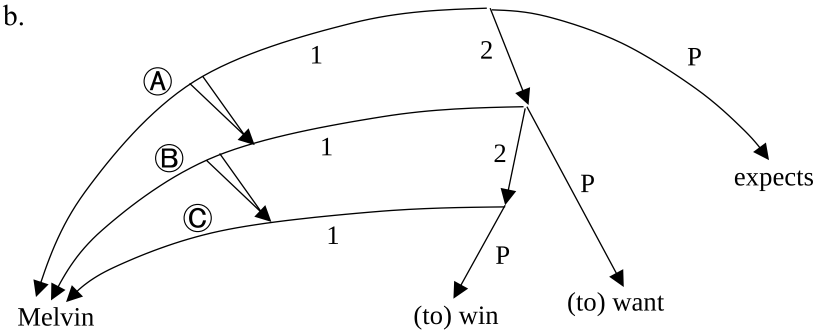
In this case, we have *three* predicates, *expect*, *want*, and *win*. Two of those take non-finite complements, *expect* and *want*. Let us compare three structural descriptions for (102): the MGG one, the APG/MG one, and ours so as to assess the advantages and disadvantages of each. We begin by considering the structural description assigned to (102) by MGG (based on Chomsky, 2001 and related work):

102’) a. [$_{CP}$ [$_{TP}$ Melvin [$_{T}$ [$_{vP}$ ~~Melvin~~ [$_{v'}$ expect+*v* [$_{VP}$ ~~expect~~ [$_{CP}$ [$_{TP}$ PRO [$_{T}$ to [$_{vP}$ ~~PRO~~ [$_{v'}$ want+*v* [$_{VP}$ ~~want~~ [$_{CP}$ [$_{TP}$ PRO [$_{T}$ to [$_{vP}$ ~~PRO~~ [$_{v'}$ win+*v* [$_{VP}$ ~~win~~]]]]]]]]]]]]]]]]]]

Note the proliferation of copies (an occurrence of NP in *v*P guarantees theta-marking to NP, whereas movement to Spec-TP satisfies T’s EPP feature), and the necessity of an independent indexing mechanism to establish the relation *control* between *Melvin* and all instances of PRO (such that the entity denoted by *Melvin* is understood to be the ‘expecter’, the ‘wanter’ and the ‘winner’). While we reject non-audible structure and empty categories (two major aspects of the structural description (102’a)), this representation is useful to see the monotonically recursive structure of a sequence of predicates with similar lexical specifications insofar as syntactic context goes.

Next up is APG’s representation (from Johnson & Postal, 1980: 540, which we reproduce with minor typographical changes):

102’) b.



This structural description features three Predicate arcs, one per V. *Melvin* is the 1 of all three predicates, and the 2 of *expect* and *want* corresponds to a clause (i.e., a predicate plus its own 1 and 2 arcs). There is no need to resort to PRO or indexing at all in APG; rather, it is arcs that get erased (here marked with double arrows): A erases B, B erases C. It is interesting to see that there are arcs that do not join terminal nodes (or *vertices*): because there are no arcs between predicates, the formalism is forced to admit that an arc may not join two elements in the ‘superficial form’ of the sentence.

The conditions we impose over edges, by definition, are stricter: an edge *must* join two and only two nodes, and every node must be justified in terms of *economy of expression*, regardless of their formal convenience. Our structural descriptions do not resort to non-audible structure in labelled projections and empty categories, or erase / sponsor arcs. The structure will have as many roots as lexical predicates; each of those lexical predicates will form its own sub-graph with its functional dependants and nominal dependants, and these sub-graphs will be linked at those nodes whose IL translation is identical. The cyclic analysis is as follows:

Cycle 1 = [Melvin expects [Cycle 2]]
Cycle 2 = [Melvin (to) want [Cycle 3]]
Cycle 3 = [Melvin (to) win]

With this in mind, the global ρ-set (after *linking* the three cycles at the common node) for (102), in our terms, is (102’c):

102’) c. ρ = ⟨(expect, *Melvin*’); (want, *Melvin*’); (win, *Melvin*’)⟩

At this point, we can shake things up, and proceed to the analysis of something like (103):

103) John seems to want Mary to tell Peter to try to finish the paper

(103) combines *subject raising*, *subject controlled equi*, *object raising*, and *object controlled equi*. We admit it is somewhat heavy, but it is grammatical. A laboratory sentence, sure, but it helps us illustrate the possible dependencies that we allow for. In this case we will go directly to our own proposal, step by step, leaving the MGG and APG structural descriptions as an exercise to the reader.

Above we said that raising predicates did not take their own dependants, but rather inherit those of the lexical predicate they modify. Therefore, raising predicates do not define cycles: they are nodes within cycles defined by lexical predicates. With this in mind, let us make the cyclic analysis and ρ-set of (103) explicit:

103') Cycle 1 = [John seems to want [Cycle 2]]
Cycle 2 = [Mary tell Peter [Cycle 3]]
Cycle 3 = [Peter try [Cycle 4]]
Cycle 4 = [Peter finish the paper]

ρ = ⟨(seem, *John*'); (seem, want); (want, *John*'); (want, *Mary*'); (want, tell); (tell, *Mary*'); (tell, *Peter*'); (finish; *Peter*'); (finish, *paper*')⟩

The graph defined by (103') is multi-rooted: there are four lexical predicates *want*, *tell*, *try*, and *finish*, of which *want* is modified by *seem* (bear this structural pattern in mind, for we will come back to it in our extension of the present system to Spanish auxiliaries in **Section 6**). The GF of each nominal element can be read off the ρ-set, assuming (58).

The English cases we have analysed so far suggest that *subject raising* and *subject controlled Equi* must receive a different structural analysis, beyond what may be encoded in lexical specifications. However, this conclusion must not be generalised to other languages automatically: just like the English analysis was motivated empirically, so must the structural descriptions assigned to *subject raising* and *subject controlled equi* sentences in other languages. Let us see an example. In the Spanish grammatical tradition, the possibility of a verbal predicate to combine with a meteorological V is taken to be an indication of its status as an auxiliary, and thus akin to *raising* predicates (see, e.g., García Fernández, 2006; Bravo, 2016). As the argument goes, raising verbs do not impose restrictions on their co-occurring NPs, and nor do auxiliaries. Thus, the ill-formedness of (104a, b) is due to the incompatibility between the subject NP and the embedded predicate, not the raising V:

104) a. *La piedra parece correr (Spanish; raising V)
The stone seems to run
b. *La piedra ha corrido (Spanish; auxiliary V)
The stone has run

The problem with the sentences in (104) is not that a stone cannot *seem+V* or *have+V-ed*, but rather that it cannot *run*. This is represented in the theory in terms of a lack of argument structure in *raising V* and auxiliaries, they do not assign thematic roles because, *stricto sensu*, they do not take arguments. And because meteorological verbs do not take arguments either, auxiliaries and raising verbs can freely combine with them. Lexical verbs (including *equi* verbs) do restrict the class of NPs that they can co-occur with, and thus –so the argument goes- they *cannot* combine with meteorological verbs. Or, rather, they *shouldn't*. However, we do get examples like the following, both of which are perfectly natural and acceptable sentences:

105) a. Quiere llover
Want$_{3SgPres}$ *rain*$_{INF}$
b. Parece llover
Seem$_{3SgPres}$ *rain*$_{INF}$

The English equivalent of (105a) would be *it wants to rain,* which is ill-formed due to the incompatibility of *want* and *rain*: in simple words, things that can *want* cannot *rain*. The grammaticality of (105a) seems to support a view in which configurational information is supplemented with lexical specifications: (105a) is, in our opinion, better analysed in terms of LFG's *functional control* (which is lexically specified) rather than as a garden-variety *equi* construction. Of course, this must not be overgeneralised: any run-of-the-mill *equi* construction with *querer* receives (in principle) the same ρ-analysis as the English examples; (105a) simply cannot be assigned the same structural description as (106), despite superficially featuring the same verb:

106) Juan quiere terminar el trabajo pronto
Juan wants to finish the work soon

We hope that this necessarily brief discussion serves at least as part of an argument in favour of combining structural information and lexically governed process in grammatical analysis, putting empirical adequacy before formal apriorisms.

5.5 A note on raising and polarity

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:

107) *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**):

108) 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 COMP$_{[+Int]}$ in love was **ever** happy? (the interrogative operator COMP$_{[+Int]}$ as a local, accessible trigger; the NPI is thus licensed)

We will now analyse 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 (see, e.g., 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, which follows from X-bar theory, that material inside 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-adjoined (Chomsky, 1955), thus, when we consider a *kernel sequence*[28], they are still not 'there' in the derivation (see also the relative timing between *constituent structure rules* and *conjoining* and *embedding generalised transformations* in Fillmore's 1963 cyclic architecture of the grammar).

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 (which is a nice argument for its extra-theoretical reality); 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* […]
>
> *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: 34)

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

109) a. **No** policeman seemed to all of the reporters to have **any** chance of solving the case

[28] 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 generated by the [Σ, F] grammar –a PSG in normal form- (Chomsky, 1957: 45).

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 case[29]

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 (109a), 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 (109c), 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 (109b) 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' PSG/CG approach in Dowty, 2003): note that in general it can be omitted and when it is materialised its linear position is not as restricted as with complements. Let us illustrate the situation with a *Richard* construction, already familiar to us:

110) 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 (109a) above presents similar positional freedom:

109') 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. Summarising (109b, c, d) in a single example, we get both peripheral positions yielding severely degraded or directly ungrammatical results:

109'') (*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)

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 (111):

111) <u>*Licensing (first preliminary formulation)*</u>

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

Which entails that there must be a *trail* between v_i and v_j. 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

[29] For ease of contrast, compare (109d) 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 below.

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 (109”) is (112), using S to denote the root node for convenience:

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

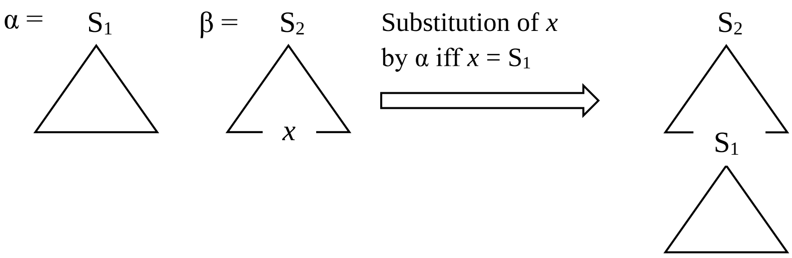
That is: both left and right peripheral topic adjuncts are assumed here to be adjoined to the root (see also Baltin, 1982; Kroch, 2001), whereas the intermediate adjunct is somewhere in the VP, in phrase structure terms. Here, what matters is simply the root vs. non-root distinction, because if the adjunct 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):

113) <u>*Licensing (second preliminary formulation)*</u>

Let G and G’ be sub-graphs and v_i and v_j be nodes. Then, $v_i \in G$ may license $v_j \in G'$ iff

i. *(v_i, v_j) $\in \rho^*$, and*
ii. *G’ is* **not** *adjoined to the root of G, and*
iii. *There is at least a node in G that is identical to a node in G’*

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 generalised embedding transformations (in the sense of Fillmore, 1963; these could also be understood in terms of *substitution* in Kroch & Joshi, 1985; Frank, 2002: 17); for instance:

114)


S_1 is still the root of the sub-graph α (whose internal structure does not concern us now), which is what matters. Strictly speaking, (114) illustrates the *substitution* of *x* by S_1 under identity, because the node *x* is in the frontier of β, and it does not dominate anything else. If *x* dominated some structure, there are additional requirements should the TAG definition of *adjunction* be used (Kroch & Joshi, 1985: 9, ff.; 1987: 111): in this case, the structure is not extended in the same way as in Chomsky-Adjunction (such that, for instance, Chomsky-adjoining an XP to VP necessarily extends the VP; see Frank, 2002: 20 for some discussion about differences between TAG-adjunction and Chomsky-adjunction). In TAG-*adjunction*, the structure dominated by *x* in β would be ‘pushed downwards’ by the adjunction of α, which in turn needs to have a node labelled *x* in its own frontier in order to preserve structural relations.

Now, let us consider a contrast that we mentioned in passing above: that between (109d), repeated here for convenience as (115a), and the new example (115b)

115) 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 fronting *without* subject-auxiliary inversion in (115a) as *topicalisation*, and fronting *with* inversion as *focalisation*. This choice of terminology is not innocent: the kind of argument we put forth here contrasts with Rizzi's (1997: 295-296) pertaining to the mechanisms of *topic* and *focus*. In the light of the conditions (107 ***i***-***iii***) above for licensing, we can interpret the contrast in (115), in which an NPI is licensed under fronting + inversion (*focus*) but not under simple fronting (*topic*). Let us discuss this contrast in some more detail.

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 (in Rogers' 2003 terms, we have a *composite* tree). 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 (113*i*-*iii*) are sufficient to filter out NPI licensing from a *topic*. But then what happens with (115b)? 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: focalised structures only require a single local graph

In a non-transformational framework, in which things do not move around literally leaving copies or traces behind, the relevant difference between adjunction to the non-root and no adjunction is that between *non-monotonic* structure and *monotonic* structure respectively: a graph grows monotonically if new connections are established within a cycle (see also Uriagereka, 2002 for a perspective based on MGG phrase structure and justified on the need to linearise local derivational units via Kayne's Linear Correspondence Axiom). In these terms, the choice between I and II depends on the answer to the following question: 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 focalised 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 *focus* that we see in (115b) is *base-generated* (but this does not extend to other kinds of *foci*, see Jiménez Fernández, 2015 for a classification of *foci*, and García Fernández & Krivochen, 2019 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).[30]

[30] As we have emphasised 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 (115a) and (115b), only in the former do

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 generalisations 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. Relations across cycles: parentheticals and Clitic Climbing in Spanish

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 aprioristic 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 (that is, relations between nodes which belong to distinct local single-rooted graphs). We can take a sub-graph and embed it in another sub-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 the mechanism of *substitution / adjunction* in a Tree Adjoining Grammar (Joshi, 1985; Kroch & Joshi, 1985; Frank, 2002), but because there is no restriction that the target for embedding is the root of a sub-graph, we can speak of *generalised adjunction*, which also sounds rather cool.

Recall that we started our inquiry with a distinction between *relation changing transformations* (RCT) and *relation preserving transformations* (RPT), following McCawley (1982); the former modify GF, whereas the latter only affect linear order. More generally, we extended the notion of 'relation' such that it applies to local connectivity in a graph: let *a*, *b* and *c* be nodes in G, and let there be a relation R(*a*, *b*). If we want to establish a connection between *a* and *c*, an RCT would *replace* R(*a*, *b*) by R'(*a*, *c*), in which case we are left with only one relation; whereas an RPT would *add* a new relation R' to the set of relations in G without modifying R.

Assume now that we have a string *abcd* that displays the following dominance relations:

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

And assume also that there is an embedding transformation inserts the substring *efg* at *c* via *adjunction*, yielding *ab*[*efg*]*d* ('pushing' *d* downwards; see 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

---

we have a separate tonal unit for the fronted 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*'.

edge from any individual element in the original string to some individual element in the adjoined string? If so, under which conditions?

The answer to this question relates to the more general problem of how to establish relations *across* cycles, a problem that has been at the core of syntactic theorising within both transformational and non-transformational frameworks. In some cases, the problem has been addressed from the perspective of constraints on inputs: if there is a rule *r* that relates syntactic objects, how can we specify the structural description that an expression of the language has to satisfy in order for *r* to apply? For example, *binding principles* can be formulated in such a way, and in general *indexing* mechanisms are approached from this perspective. Standard Theory-style constraints (like the *Tensed S Condition*, below) often adopted the form *rule r cannot apply in configuration C*, which is essentially a condition on the input of *r*:

Tensed S Condition:

No rule can involve X, Y in the structure

…X…[$_{\alpha}$…Y…]…

Where α is a tensed sentence

(Chomsky, 1977: 89)

In other cases, it is the *output* of the rule that is under scrutiny. Conditions which are formulated over *outputs* are sometimes referred to as *representational constraints* (e.g., in Müller, 2011), and they apply to the result of applying *r*, which in turn implies that the structural description for *r* has been met (in other words, that the *input* of the rule was well-formed). An example of such *output* conditions is the *Empty Category Principle* (ECP) as formulated in Chomsky (1981):

[$_{\alpha}$ e] is properly governed

(Chomsky, 1981: 250)

The ECP, in turn, was invoked in order to justify punctuated movement in GB, since traces needed to be antecedent-governed and thus *operator-variable* relations needed to be *local* (i.e., within the limits of a cyclic node, S' / NP).

This brief discussion serves as a background for the following questions: What would a condition over cross-cycle dependencies look like under present assumptions? Does the distinction between RCT and RPT interact with the formulation of these conditions?

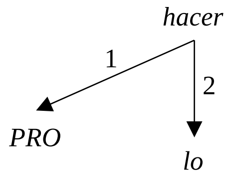
We would like to propose, initially, that the insertion of an element in a graph that *only disrupts linear order* without changing constituency does *not* count as defining a cycle. In other words, the configurations that are allowed under our definition of *licensing* are sensitive to the distinction between RCT and RPT. Thus, we suggest the following preliminary constraints (from Krivochen et al., in preparation):

117) *[X… […α…]…Y] yielding R(X, Y) is a legitimate configuration iff*
i. […α…] has been introduced non-monotonically, ***and***
ii. The rule introducing […α…] 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, focusing our discussion on legitimate relations between argumental *clitics* and the predicates that select them under specific syntactic operations: *clitic climbing* and *parenthetical insertion*.

Let *a* be a node corresponding to an argumental clitic and let *b* be its governing V; thus, we have the following relation: R(*a*, *b*) = R(Clitic, V). Now we need to ask: what exactly is R? Here we argue that a clitic-governing V structure is modelled in terms of *dominance*, thus, something like *hacerlo* (Lit. do_{INF}+it_{CL}) displays the dominance relation ρ(hacer, *lo*), where the accusative clitic is the direct object of the verb *hacer*. We will leave discussion pertaining to the subject aside for now, let it be PRO_{arb} or some such (as in *hacerlo es impensable* '(for anyone) to do it is unthinkable') simply for expository purposes. In Relational Grammar / APG terms, we would have something along the lines of the following preliminary representation:

118)



Or, using our *ordered* notation:

118') ρ = ⟨(hacer, PRO); (hacer, *lo*)⟩

These considerations hold at a very local level, and for *a-structure* (that is, the syntactic representation of predicates and their selected arguments). 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 trail[31]). This means, as a provisional observation, that parentheticals (if derived along the lines of Emonds, 1979; McCawley, 1982, via post-cyclic *adjunction*) should not count as intervening nodes for clitic climbing, because the clitic would not be ordered with respect to any of the nodes that constitute the parenthetical by virtue of it not being part of the same *derivational current* as the clitic. This is indeed verified empirically:

[31] The space in which our graphs are defined 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)

119) 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$_i$ dejan, a Ana, preparar e_i algunas veces

(119) can only be acceptable if *a Ana* is derived in parallel, not monotonically (note that in (119b) *a Ana* appears between commas). If this is the case, the clitic is ordered with respect to *dejar* and *preparar*, but *not* with respect to *a Ana*. Therefore, the annotated string (119b) complies with the requirements in (117) over legitimate licensing configurations, and is thus predicted to be grammatical.

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 subcategorised for nor monotonically assembled) is not always clear. Initially, we would expect that (120b) below would not work, but surprisingly, it does:

120) 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 (120b) 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 via some sort of *generalised transformation* into the matrix clause *Juan puede hacerlo*, then a node belonging to this adjoined sub-structure 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, 1973 in a Subjacency-inspired view; Uriagereka, 2002 for a different perspective; also De Vries, 2007), 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). It is important to note that we are dealing with two distinct problems: one pertains to how parentheticals are linked to matrix sentences, the other to the internal structure of parentheticals and their opacity. McCawley and Emonds shed light on the former issue, but not on the latter (at least not directly). Further refinement of the analytical machinery 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:

121) 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

122) 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 (121) 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 (122) is more than simply an order-changing rule: the possible adjunction sites are very much restricted. One way of looking at this pays close attention to lexical semantics: the *locus* of parenthetical adjunction depend on whether there is a semantic relation between the adjoined domain and the matrix clause; specifically, we can look at the auxiliaries in the adjoined phrase and the matrix clause, since these are the possible hosts for clitic climbing and since the parenthetical seems to be a 'correction' affecting the modal in the matrix clause. In the specific case (122), the modal auxiliary in the adjoined clause is related to that in the matrix clause on a scale, such that *tener que* means deontic obligation, whereas *poder* only means deontic possibility, a weaker notion than obligation. 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*, which belong to different semantic scales), the construction ceases to be acceptable:

123) ??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 (121) not an intervening object between the 'gap' and the coindexed clitic, but the adjoined auxiliary in (122) *is*? Note that the lexically-oriented proposal can shed some light on where the parenthetical can appear in terms of scalar relations between auxiliaries in the matrix clause and the parenthetical, but it says nothing about syntactic dependencies between objects in the matrix clause and objects within the parenthetical (other than the root). Of the two problems we identified above, the second (the opacity of cycles for purposes of relations with objects in other cycles) remains unaddressed; we will now focus on it.

Let us go back to the contrast between (121) and (122): when does an embedded (TAG-adjoined) cycle 'get in the way' of links in the matrix cycle? The possibility we will explore here is that if the adjoined graph 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 (AT) is a *self-contained unit*, it does not intervene for purposes of operations at the Initial Tree to which the AT is adjoined (Joshi's 1985: 214 expansion of TAGs by means of *links* preserved after *adjunction* yields results that are similar to ours in weak generative capacity). Of course, making what 'self-contained' means explicit is paramount. We define the notion as follows:

124) *Self-containment (definition)*

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

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

ii. v_i *receives a grammatical function 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 (58)) that is dominated by a node outside that graph (i.e., an argument of a predicate in G', where G' ≠ G). Note that *self-containment* is the notion that we appealed to informally in condition (113*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 (113*iii*) formulated in terms of *self-containment* (which we need anyway):

125) Licensing (final formulation)

Let G and G' be sub-graphs and v_i *and* v_j *be nodes. Then,* $v_i \in G$ *may license* $v_j \in G'$ *iff*
i. $(v_i, v_j) \in \rho^*$*, and* [alternatively, *there is a trail T that contains* v_i *and* v_j *and* v_i *is ordered before* v_j]
ii. G' is not adjoined to the root of G, and
iii. Neither G nor G' are self-contained

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 (120), the adjoined 'parenthetical' contains a node such that ρ(*tener que*, Cl) holds, as well as ρ(*hacer*, Cl). We can give the relevant set of *local* dominance relations for (114a) as follows:

126) Cycle 1: ρ = ⟨(poder, *Juan*'); (poder, *lo*); (poder, hacer); (hacer, *Juan*'); (hacer, *lo*)⟩
Cycle 2: ρ = ⟨(tener que, *Juan*'); (tener que, *lo*); (tener que, hacer); (hacer, *Juan*'); (hacer, *lo*)⟩

The adjoined object, which corresponds to Cycle 2 in (126), 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 multidominance approach to RNR (McCawley, 1987, 1998; Levine, 1985; see also **Section 13.1** below). This last claim requires some unpacking: were we working within a transformational framework, the D-Structure of (120a) would need to look like (120a') in order to get a proper semantic interpretation:

120a') 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.

6.1 Discontinuity and clitic climbing in Spanish auxiliary chains[32]

A crucial point is that so-called 'clitic climbing' (discontinuity relations between clitics and their governing verbs) occurs across sequences of verbal predicates under specific conditions (Rivas, 1974; for a Relational Grammar treatment, which is now classical, see Aissen & Perlmutter, 1983). In this section we will focus on clitic climbing through *auxiliary chains* (in the sense of Bravo et al., 2015; García Fernández et al., 2017). 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:

127) *An auxiliary chain* CH_{AUX} *is a string* $\{\{x \frown y \frown z \ldots n\} \frown VP\}$ *where*

[32] Much of the data and discussion in this section is adapted from Krivochen & García Fernández (2018). We are grateful to Luis García Fernández for allowing us to use that material here.

i) {x, y, z…n} ∈ Auxiliary Verb
ii) n > 2

(Bravo et al., 2015: 75)

Consider now the following examples, where traces have been added for expository purposes only:

128) a. Podrías gustarle y decirlo a todo el mundo
Could$_{2SgCond}$ like-CL$_{Dat3Sg}$ and say-CL$_{Acc3Sg}$ to all the world
'He/she could like you and you could tell it to everybody'
b. Le$_i$ podrías ser t_i infiel y decírselo a todo el mundo
CL$_{Dat3Sg}$-could$_{2SgCond}$ be unfaithful and say-CL$_{Dat3Sg}$-CL$_{Acc3Sg}$ to all the world
'You could be unfaithful to him/her and tell it to everybody'
c. *Lo$_i$ podrías serle infiel y decir t_i a todo el mundo (the clitic climbs to the auxiliary from the second term of the coordination; the result is ungrammatical)
CL$_{Acc3Sg}$-could$_{2SgCond}$ be unfaithful and say to all the world

129) a. Nos estás molestando y mirando
CL$_{Acc1Pl}$-Are$_{2SgPres}$ bothering and looking
'You are bothering us and looking at us'
b. *Estás molestándonos y mirando (intended reading: same as (129a))
Are$_{2SgPres}$ bothering-CL$_{Acc1Pl}$ and looking
c. *Estás molestando y mirándonos (intended reading: same as (129a))
Are$_{2SgPres}$ bothering and looking-CL$_{Acc1Pl}$

In (128a) we have the auxiliary *poder* followed by two coordinated infinitives, *gustar* and *decir*, each of which hosts a clitic which corresponds to its internal argument. (128b) features the clitic which depends on the first terms of the coordination having climbed above *poder*, whereas the second term of the coordination has both accusative and dative clitics *in situ*. The contrast between (128c) and (128d) is particularly interesting, since it shows that there is no problem in distributing an *auxiliary* between coordinated infinitives (such that (128c) means *estás molestándolo y estás mirándonos*), but it is *not* possible to distribute a *clitic* between coordinated terms. In (129a) we have the progressive auxiliary *estar* followed by two coordinated infinitives, but unlike the examples in (128) there is only one clitic proclitic to the auxiliary which is nevertheless interpreted distributively with respect to the coordinated infinitives (as can be seen in the English translation). The analysis presented in this squib explains why some elements can be distributed across coordinated lexical verbs in periphrastic constructions while others cannot. Our goal is to provide adequate structural descriptions for these sentences which capture their syntactic and semantic properties.

An important constraint to bear in mind is that it is impossible to make two clitics climb from different terms of the coordination. (130) illustrates the result of making a clitic from each term climb, yielding an ungrammatical sentence[33]:

---

[33] It is relevant to note that there is no *a priori* constraint on making clitics climb from different predicates, as seen in (i):

i) Se los deja traer al examen (from 'Les deja traer el diccionario al examen')
CL$_{Dat}$ CL$_{Acc3Pl}$ lets bring to the test
'He/she lets them bring it to the test'

130) *Se$_i$ lo$_j$ podrías ser t_i infiel y decir t_j a todo el mundo
CL$_{Dat3Sg}$CL$_{Acc3Sg}$-could$_{2SgPres}$ be unfaithful and say to all the world

These restrictions have nothing to do with lexically governed processes pertaining to either the auxiliaries or the lexical verbs (it is also worth pointing out that argumental and non-argumental clitics behave exactly the same for all present intents and purposes), since both (131a) and (131b) below are well-formed as individual sentences:

131) a. Le$_i$ podrías ser t_i infiel
CL$_{Dat3Sg}$-could$_{2SgCond}$ be unfaithful
b. Lo$_i$ podrías decir t_i a todo el mundo
CL$_{Acc3Sg}$-could$_{2SgCond}$ say to all the world

The relevant condition that needs to be invoked to account for the ungrammatical cases is, we argue, Ross' (1967) *Coordinate Structure Constraint* (CSC), to which we will return in detail in **Section 11**:

> *In a coordinate structure, no conjunct may be chopped, nor may any element contained in a conjunct be chopped out of that conjunct.* (Ross, 1967: 428)

It is necessary to state this more precisely, because the CSC has been understood in at least two different senses, not always yielding equivalent results. As correctly pointed out by Postal (1998: 83),

> *It seems correct to divide Ross's original formulation of the CSC into separate principles. The one I called the Conjunct Constraint* […] *forbids the extraction of coordinate conjuncts themselves. The other, the CSC, bans (non-ATB) extraction from true conjuncts.*

Here, we use the term 'CSC' to refer strictly to the *second* of Postal's principles: non ATB extraction from *true* conjuncts is to be filtered out[34]. We will come back to the mechanisms of ATB dependencies in **Section 13.1** in the context of the discussion of Right Node Raising (RNR), but it is worth introducing some aspects of that discussion here.

Let G and G' be sub-graphs, each of which corresponds to a single cycle, and let G" = G ⋃ G' (where ⋃ stands for Shieber-style *unification*). Let *R* be a relation between nodes *a*, *b*, *c*, *d* in G and G' such that R(*a*, *b*) holds in G and R(*c*, *d*) holds in G', where *a* and *c* are lexical predicates. *R* is an ATB relation *iff b = d*. Note that, because relations hold of ordered pairs of nodes, *b* and *d* are assigned a GF in G and G' respectively. This formulation is rather clumsy, however. We can simplify things by appealing to the notion of *linking* defined in (30) above, and repeated here:

> *If a sub-graph G contains v_1 and a sub-graph G' contains v_1 (that is, if $\rho(v_G, v_1)$ & $\rho(v_{G'}, v_1)$ hold, for v_G and $v_{G'}$ arbitrary nodes in G and G'), then G and G' are **linked** at v_1*

For ATB, we need to further specify (i) that v_1 needs to receive a GF in G and G', and (ii) that G and G' are sub-graphs of G". Note that we have no need of indexing operations to keep track of what has 'moved' where, and of course the multiplication of copies is also avoided: an issue when considering traditional movement-based approaches to ATB operations is that for *n*-coordinated terms affected, *n*-1 occurrences need to be deleted. Thus, in order to get *what did John buy and Mary break?*, there has

---

The relevant constraint applying to (130) must thus pertain to the structure of coordination, not to the fact that there is CC from two different predicates.

[34] As a reminder, Ross (1967: 176) first defined ATB phenomena as a class of rules that affect (e.g., displace, delete) identical constituents of all the conjuncts of a coordinate structure at once.

to be a way to get two NPs in object position (of *buy* and *break*), transform them into *wh*-phrases, and then move only one to the left periphery, deleting the other. Or leave the object NPs *in situ*, lexically insert a *wh*-phrase in the left periphery of the clause, assign the same index to both object NPs and the *wh*-phrase, and delete the NPs. Other options are also possible, but they all involve multiplying indexes, or occurrences, and require *ad hoc* selection and deletion rules. We will return to these considerations in **Section 13.1**, for the time being it is worth noting that if we allow for multidominance in maximally connected local graphs the conditions that account for the relevant data involving clitic climbing in coordinated auxiliary constructions can be captured without the need to invoke any further principles or operations.

If we see clitic climbing as an operation that creates new edges in addition to the primitive relation between the lexical V that assigns the clitic a GF, this should also apply to climbing through cycles. The only constraint to bear in mind is the CSC in Postal's sense: non-ATB climbing is not allowed. Thus, (132a, b) in which clitic climbing does not affect elements in both conjuncts are correctly excluded:

132) a. *Lo podrías serle infiel y podrías decir $t_{\mathrm{CL.ACC}}$ a todo el mundo
b. *Se lo podrías ser $t_{\mathrm{CL.DAT}}$ infiel y podrías decir $t_{\mathrm{CL.ACC}}$ a todo el mundo

In these examples as well as in (130), the CSC strictly understood is violated, and the results are ungrammatical. So far so good, but, can we say something about how complex grammatical cases are analysed? After all, in the grammatical cases in (128) and (129) clitic climbing is not all that is going on, there is also *gapping* of the auxiliary. That is, the corresponding interpretations for (128a) and (129a) are (128'a) and (129'a) respectively:

128') a. Podrías gustarle y podrías decirlo a todo el mundo
129') a. Nos estás molestando y nos estás mirando

The auxiliaries *poder* and *estar* are distributed with respect to the lexical verbs: a structural description in which the auxiliary only occurs in the first conjunct cannot adequately represent the semantics of the sentence (see Krivochen & García Fernández, 2018 for further discussion). We should be able to provide an adequate structural description without multiplying the occurrences of the auxiliaries or the clitics. Let us, then, specify the ρ-sets for (128a) and (129a), leaving aside for the time being the structure of coordination:

133) Cycle 1: ρ = ⟨(poder, *pro*); (poder, gustar); (gustar, *pro*); (gustar, *le*)⟩
Cycle 2: ρ = ⟨(poder, *pro*); (poder, decir); (decir, *pro*); (decir, *lo*)⟩

134) Cycle 1: ρ = ⟨(estar, *pro*); (estar, molestar); (molestar, *pro*); (molestar, *nos*)⟩
Cycle 2: ρ = ⟨(estar, *pro*); (estar, mirar); (mirar, *pro*); (mirar, *nos*)⟩

Note that cycles 1 and 2 in (133), where each term of the coordination has its own clitics which are not correferential, are linked at the auxiliary *poder*; in this case there is no need to involve *deletion* in *Gapping*: the operation is licensed when two cycles are linked at a verbal node. It is worth pointing out that, because the clitics are not correferential, climbing is not licensed since it is not possible to make it happen *across-the-board*; in other words, because the cycles are not linked at the node that corresponds to the clitic's IL translation, however we choose to formulate the operation Clitic Climbing, it cannot possibly affect both terms of the coordination (see also (132)).

The case of example (134) is slightly more complex, because the cycles are linked at *two* nodes: *estar* and *nos*. We predict, then, that it is possible to apply Clitic Climbing *across-the-board*, as well as to have *Gapping* of the auxiliary. In other words, we predict that (135) below (= (129a)) should be grammatical and furthermore that it should indeed be interpreted as (129'a), with the auxiliary and the clitic being distributed over the conjuncts:

135) Nos estás molestando y mirando
CL_{1PlAcc} $are_{2SgPres}$ bothering and looking

Both these predictions are borne out, without the need to invoke additional conditions or theoretical tools: the notion of *linking* can naturally capture the requirement for cross-cycle operations to apply ATB in the relevant cases.

Our graph model can provide adequate descriptions for at least a set of expressions belonging to languages other than English; the present section has focused on some aspects of the dynamics of auxiliaries and clitics in Spanish. From this it must not be inferred that we are giving our conditions the status of universals, or that they apply to any additional phenomena: these are empirical claims, and as such cannot receive an *a priori* answer.

The analysis of auxiliaries, preliminary as this one was, required us to be able to articulate relations within and across cycles: it is not a new claim that some auxiliaries have a syntactic behaviour similar to *equi* constructions and others are closer to *raising* structures (García Fernández, 2006 for Spanish). Within the present model, the heterogeneity of the class of auxiliary verbs identified for Spanish in Bravo et al. (2015) and subsequent work can be readily captured: in those works we identified Spanish auxiliaries which, within a sequence of auxiliary verbs, could both modify other auxiliaries and be modified themselves (dubbed *lexical auxiliaries*) and auxiliaries which could only modify, but not be modified (dubbed *functional auxiliaries*). The first group includes *modals*, *phasal aspectual auxiliaries*, and iterative *volver a*; the second includes *progressive 'ser'*, *perfective 'haber'*, *temporal-aspectual 'ir a'*, and some more. The following table summarises the proposal:

Transparent / functional	**Opaque / Lexical**
Progressive *estar* 'to be', passive *ser* 'to be', perfective *haber, Eng. have –ed*; *ir a* 'be going to / will'	Phasals (*empezar a*; *terminar de…*), second-position modals, first-position auxiliaries (*soler*, *haber de* 'have to'), *tardar en* 'take time', *volver a* (to do something again)

The distinction is *empirically* motivated, but it has far-reaching consequences for the format of phrase markers, which has been analysed in previous work from both Immediate Constituent and Categorial Grammar perspectives (Krivochen & Schmerling, 2018). We provided arguments that prove that a monotonic approach to sequences of auxiliary verbs (as that proposed in Cinque, 1999 within MGG; and Bach, 1983 within Generalised 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. The cyclic approach pursued here, where each cycle contains one and only one *lexical* node, captures the empirical insights from Bravo et al. (2015); García Fernández et al. (2017); Krivochen & García Fernández (to appear) and related works in a simple manner: *functional auxiliaries* are modifiers of *lexical auxiliaries* (or *lexical verbs*), and therefore belong within the cycle defined by the latter.

The consideration of a specific process that can take place across cycles in auxiliary chains, Clitic Climbing, motivated the analysis of the conditions under which a cycle is internally accessible, or ‘transparent’: concretely, we asked ‘which configurations determine that an adjoined cycle intervenes for purposes of relations at the target of adjunction?’ 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

A crucial point of **Section 6** was that a one-size-fits-all, purely configurational theory of parentheticals (like that proposed in MGG) 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 (118) above. If two cycles are *linked* at some node, then *in principle* they are mutually accessible: of course a closer inspection may reveal that some node in one of these may not be ordered with respect to some node in the other or some further condition may be violated. In this section we will come back to English data, and present further evidence in favour of a ‘mixed’ approach to parentheticals (and other adjoined sub-graphs) in which opacity is not an automatic consequence of adjunction. The data we will consider pertains to pronominal reference and crossover effects, and will provide the basis to rethink the principles of Binding Theory in terms of ordered relations in digraphs.

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

136) 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 materialisation of nodes (i.e., their morpho-phonological exponent) which would be unexpected (and remain unexplained) 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:

137) 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 (137b) and (137c). The MGG literature on Binding Theory (Chomsky, 1981, 1995, and related work; see Truswell, 2014 for a review) would attempt to account for the ungrammaticality of (137b) and (137c) in terms of a Principle C violation: the relevant cases would receive structural descriptions in which the R-expression [John] is bound within its governing category. That proposal would indeed work for (137c), in which the pronoun [he] and the R-expression [John] co-exist in the same derivational space (in other words, they belong in the same *phase*, in Minimalist parlance; see e.g. Wurmbrand, 2011: 60). (137b), 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 together with the rest of the sentence (either by top-down phrase structure rules or bottom-up Merge). 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 (138):

138) *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 (138) could be blamed on a violation of the Complex NP Constraint (Ross, 1967: 127) which bans extraction of X from a configuration of the kind [$_{\text{NP}}$…[$_{\text{S}}$…X]]; 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 (2019a), 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 singulary transformations in a Fillmorean architecture) and (b) whether the adjoined sub-graph is a *self-contained unit*. In that work, which advanced a TAG 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 singulary transformation T_S triggered from γ can affect β* ***iff*** *T_S is intrinsically ordered*[35] ***after*** *an embedding transformation that adjoins β to γ at X.*

What singulary transformations (e.g., *wh*-movement) *cannot have access to*, we argued, is *elements embedded within β* (i.e., dominated directly or transitively by the root node in *β*); only *β* as a whole can be affected by a singulary transformation at γ ordered after *adjunction* of β to γ. We committed

[35] 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.

ourselves to a model of syntax with multiple cycles and ordered rules (building on Fillmore, 1963), which crucially does *not* entail any commitment to a multi-layered model with several levels of representation and corresponding rules of interpretation. Despite the fact that the present proposal is not derivationally (or, more generally, *proof theoretic*) oriented, some empirical consequences of our previous derivational proposals hold also in the present model. To the extent that such empirical consequences pertain to locality conditions in the establishment of dependencies, they must be captured regardless of the formal foundations of the theory, *proof-theoretic* or *model-theoretic*.

Here we go further along the line of reasoning that we adopted in Krivochen (2018, 2019a) as far as the emergence of interpretative domains goes: we propose that a graph G 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):

139) <u>*Full Order Condition*</u>

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:

<u>*Full Leaf Order*</u>*: 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 emphasised in Sternefeld's (1998a, b) 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) → D(a) ∧ D(b); [irreflexivity]
¬(R(a, b) ∧ R(b, a)); [antisymmetry]
R(a, b) ∧ R(b, c) → R(a, c). [transitivity]

R(a, b) may again be written a < b (mod. R). (Krivine, 1971: 7)

The condition in (139) 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 (137) seems to resist traditional accounts. In sum, the problem is that delaying the application of Binding Theory principles derivationally (say, to LF, after the ‘overt’ syntactic computation and covert operations) does not solve the essential, underlying issue: how come sometimes parentheticals are visible for purposes of operations at the matrix clauses to which they are adjoined, 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 (137), 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):

140) ρ = {(to, *John*’), (*John*’, should), (for, *John*’), (*John*’, 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 materialisation 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) and fronted adverbial clauses triggering crossover effects[36]:

141) a. The President, whose son$_i$ is involved in a collusion scandal, betrayed him$_i$
b. *Who$_i$ did the President, whose son$_i$ is involved in a collusion scandal, betray t_i?

142) 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 (141b) is unexpected (and remains unaccounted for) if non-restrictive relative clauses (a.k.a ‘apositive relatives’) are assumed to be opaque to operations at the matrix clause to which they are adjoined, 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). Similarly, with a fronted adverbial clause we have crossover effects that are unexpected if the root-adjoined object is not internally accessible: the NP that causes the crossover effect is embedded inside that clause, and in an orthodox view of representational relations it is not clear how either (141b) or (142b) satisfy the conditions for *strong crossover* (see also Postal, 1971):

> *The central cases of the strong crossover configuration arise when Wh Movement has taken place from some A-position A* asymmetrically c-commanding some pronoun P in an A-position, to an A’-position A* that c-commands P. In a right-branching language like English, this will arise only if P is "between" A’* and A*. The movement can thus be said to "cross over" P.* (Sportiche, 1985: 461)

[36] We are indebted to Susan Schmerling for her help in constructing these examples.

Note that under an orthodox view, it is not possible to explain the ungrammaticality of the (b) cases via crossover, because the NPs *whose son* and *President Trump* do *not* c-command A* (it is important to observe that (141b) features the *wh*-operator coindexed not with the possessive, but with the N complement). Since the problem seems to us to be in the theoretical assumptions about phrase structure and configuration rather than on the definition of *crossover*, we will propose in what follows an alternative formulation of Binding conditions within locally cyclic digraphs rather than defined over phrase structure trees.

Here we combine considerations about derivational timing (essentially following Fillmore, 1963) with the relativisation of the notion of representational opacity: as we saw in the previous section, *a sub-graph is opaque for purposes of the establishment of relations with objects outside that subgraph if and only if it is a self-contained unit*. Sportiche's definition of the configurations that generate *strong crossover* effects means that *strong crossover* is essentially a violation of Principle C of Binding Theory (Chomsky, 1982: 20), which establishes that referential expressions (a label that includes [-anaphoric] [-pronominal] elements; concretely, NPs and Wh-traces) must be free (i.e., not c-commanded by a correferential element). We agree with 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, we must not restrict ourselves to considering connectivity patterns (i.e., purely configurational issues): conditions like Wasow's (1979) *Novelty Condition*[37], 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*, utilising Fregean terminology) must also be captured. 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 *pronominalisation* which makes reference to the ordered occurrences of a particular node when walking a graph[38]. Because edges are directed, we can indeed capture both forward and backward pronominalisation (Ross, 1969a), both of which are formulated by Ross as *cyclic rules* (that is, they apply within sub-graphs). We have indeed pursued this line of reasoning in **Sections 4** and **5**. 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: it is unlikely, however, than a memory stack that goes beyond PDA+ capabilities is required, if even that. Constraints on *forward* and *backwards pronominalisation*, which seem to interact with cyclic conditions (see, e.g., Ross, 1969a: 192), would have to be explicitly formulated,

---

[37] Formulated as follows: '*An anaphorically dependent element cannot have more determinate reference than its antecedent*' (Wasow, 1979: 36).

[38] For instance,

> *An NP_1 may pronominalize an identical NP_2 if NP_1 is to the left of NP_2 under any conditions, or if NP_1 is to the right of NP_2 and NP_2 does not command NP_1*
> (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.

etc.[39] 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 formalising connectivity in syntactic dependencies (creating a snapshot of a dynamical system, more specifically), but it is good practice to point towards alternatives when they could exist. In any case, it is crucial to highlight that the requirement that nodes in a cycle be *strictly ordered* with respect to all other nodes within that cycle is *not* interpreted in terms of linear ordering (i.e., precedence in a string). We will see below that the requirement of *ordering* is sufficient to formulate constraints that can take care of the *crossover* cases we are interested in, given that these constraints interact with the definitions of *linking* and *self-containment* proposed above.

At this point, and considering the argument involving anaphora in **Section 5.2.1** the reader may want to generalise these observations, and thus he/she 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***
>
> *(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)

---

[39] There is also the obvious fact that pronouns can occur in contexts which do not satisfy the structural description for any *pronominalisation*-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).

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:

143) 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, 2019b: 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 formulate some preliminary generalisations:

144) <u>*Binding Graph (definition)*</u>

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

It is necessary to define *binding* formally. In mainstream generative grammar, *binding* is a relation between nodes α and β where

> *α is X-bound by β if and only if α and β are coindexed, β c-commands α, and β is in an X-position* [where 'X' stands for either A or A'] (Chomsky, 1981: 185)

Of the three conditions, we are interested in two: co-indexing and c-command. We have said a few times already that if nodes stand for IL translations of lexical items rather than for lexical items (orthographic words, bundles of features, sound-meaning pairs, depending on the theory), and if *trails* are allowed, then there is no need to have an indexing mechanism in the theory: if α and β share 'index', it means that they are the *same node,* visited more than once in a *trail*. In other words, α *dominates** α or, symbolically:

145) $(\alpha, \alpha) \in \rho^*$

The requirement that there be a *strict ordering* between nodes in a sub-graph determines that there is a unique trail in which α is visited twice, in distinct environments. While this certainly looks initially adequate, this cannot be it. One of the most immediate problems is that the cases of *reflexivity* we looked at in **Section 5.2.1** remain unaccounted for. To remind the reader of the relevant configuration, in reflexivity we are dealing with *parallel edges*:

146) $\rho = \langle (v_j, v_i), (v_j, v_i) \rangle$

For example,

146') a. John admires himself
b. ρ = ⟨(admire, *John*'), (admire, *John*')⟩

And since the 1 and 2 arcs of *admire* are the same node, we can account for reflexivity without the need to resort to indexing. Moreover, because the GF are read off the order between edges in the ρ-set of a given expression, there is no need to assume any extra structure. It seems, then, that we need to incorporate into the notion of *binding* the fact that a single node can establish more than one grammatical relation with a predicate. Sticking to classical Binding Theory, the domain within which those relations are established determines the principle of BT that applies to the distribution of a certain node; if within the target's governing category then Principle A; if outside, then B. This means that Principle B allows for two cycles to be *linked* at the node that corresponds to the IL translation of a 'pronominal' object, as in:

147) John_i wishes that Mary would love him_i

In this case, we have two lexical predicates, *wish* and *love* (the latter, modified by an auxiliary *would*), which define two cycles. These are linked at *John*', which is dominated by *wish* (of which it is the subject) and *love* (of which it is the direct object). In order to pronominalise *John* as *him*, it is necessary (in the Lees & Klima/Ross/Postal view) that *John* is higher up in the tree structure than *him*, and furthermore that they are mutually accessible. In our terms, these conditions can indeed be captured: because trails establish an *order* between nodes, such that ρ-sets feature ordered node-to-node relations, we can propose the following preliminary condition

Binding:

a node v_i binds v_j iff either

a) *$i = j$ and there is a unique trail t such that $i > j$ in t* (where > is read as is 'ordered before')
b) *$i = j$ and v_i and v_j are nodes in parallel edges*

The indexes attempt to capture the requirement that v_i and v_j have the same IL translation; apart from their context, they are identical. Condition (a) includes the case in which sub-graphs are linked at v_i, condition (b) local reflexives. We are assuming that both (a) and (b) are evaluated in digraphs which comply with the *strict ordering condition* in (139) above; in other words, if the nodes in a graph are not *strictly ordered*, it makes no sense to ask further questions: it is not a well-formed graph.

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 analysing 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.

The idea that graphs may be *linked* at inner nodes (that is, at nodes other than root nodes), if not further constrained, can be used to argue that expressions like (137b, c), (141b) and (142b) are well-formed expressions of the language, contrary to fact. Let us look again at the ungrammatical cases, which we repeat below, and see if they are adequately excluded given the assumptions above.

148) a. *Who$_i$ did the President, whose son$_i$ is involved in a collusion scandal, betray t_i?

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'?

There are some things to observe here. First, the adjoined objects are not *self-contained*: the appositive relative in (148a) contains a node that is also dominated by a node outside that sub-graph and so does the adverbial clause in (148b). Also, the ungrammaticality of these sentences is caused by the presence of the adjoined clauses, *regardless of their linear position*. That is, if the adjoined domain is extraposed such that the NP they contain is not 'between' the *Wh*-operator and the trace, the result is still ungrammatical, as long as the indexing is kept the same:

148') a. *Who$_i$ did the President betray t_i, whose son$_i$ is involved in a collusion scandal?

b. *Who$_i$ t_i may be accepting help from a country full 'bad hombres', now that President Trump$_i$ has been offered Mexico's help in the wake of Hurricane Harvey?

In order to account for these cases, we need to consider what exactly is happening in *crossover* cases, and then see how the cases with linked adjoined objects fall into the same category as these. Let us start from a simple case of *strong crossover*:

149) *Who$_i$ does John$_i$ admire t_i?

A look at the ρ-set for (149) should clear things up at least a bit:

150) ρ = ⟨(admire, *John*'), (admire, *John*')⟩

But there is no reflexivity in (149), despite the presence of parallel edges: there is at least a property of a declarative sentence (see (146') above) which ceases to hold once a new relation is created in the Wh-interrogative, which defines *three* contexts for the node corresponding to the IL of *John* (we know it is the same node because *crossover* cases require co-referentiality). But this means that this node must *dominate** itself and *be dominated** by itself in different contexts: the Wh-context dominates the Subject context, but the Subject and the Object contexts are in *parallel edges* (since they are the 1 and 2 of the same predicate *admire*); thus this node cannot be ordered with respect to itself. This violates the condition of *antisymmetry* in the definition of *strict ordering*, and yields an ill-formed graph. The contrast with a Wh-interrogative without *crossover* is stark:

151) Who$_i$ does Mary$_j$ admire t_i?

In this case, there are no parallel arcs, and following the direction of the edges 'from the root down' we get a unique trail that visits *who* twice. No violation ensues, and therefore the sentence is correctly admitted as an expression of the language.

All in all, the take-home message of this section and the previous one is that adjoined clauses are not always as opaque or syntactically independent as MGG would have us believe (Kluck et al., 2014: 4; Haegeman, 2009; De Vries, 2007). If parentheticals were uniformly derived, as analyses based on *phase theory* and *multiple spell-out* propose, then any element embedded in the parenthetical should never be visible outside the parenthetical's root, contrary to fact (in addition to the data considered here, see also Hoffmann, 1998). A configurational approach combined with an appropriate definition of 'opacity', here given in (124) 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 with 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

In this section we will deal with some aspects of the syntax of *non-restrictive*, or *appositive*, relative clauses and their relation with *restrictive* relatives. This continues the topic of the last couple of sections, insofar as we are dealing with cross-cycle relations in non-monotonic structures; our goal is to refine the formulation of the conditions under which these relations are allowed in English.

We 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 process of adjoining parentheticals, which he thinks of in terms of a transformational rule of *Parenthetical Formation* changes grammatical relations (see also Ross', 1973 recursive transformation *slifting*, which applies only sentence-final *that* clauses and right-adjoins the parenthetical to the root, thus extending the phrase marker[40]). 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. It is interesting to note at this point –as an epigraph of sorts- that Emonds himself observes that Parenthetical Formation allows for violations of Subjacency (Chomsky, 1973), for example:

152) We introduced a diplomat who got, $\begin{Bmatrix} \textit{it seemed to us} \\ \textit{in our opinion} \end{Bmatrix}$ too much attention (Emonds, 1979: 212, fn. 1)

The McCawlean view, according to which the rule of Parenthetical Formation (which he refers to as Parenthetical Placement) is an *order changing rule* which does not affect constituency (i.e., it does not change grammatical relations) renders the apparent problem of the violation of Subjacency moot: if there is no syntactic displacement, then constraints on displacement of course do not apply[41]. We will return to this line of reasoning when briefly dealing with Right Node Raising, in **Section 13.1**.

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 (153a), to be contrasted with (153b):

---

[40] Ross' formulation is as follows (1973: 134):

SLIFTING

X – [$_S$ Y-[$_S$ *that*-S]$_S$]$_S$ -Z

SD: 1 2 3 4 5 ⇒ OPT
SC: 1 4#[$_S$ 2 0 0]$_S$ 5 [# indicates Chomsky-adjunction]

[41] For a Minimalist analysis of Italian and (some) English data, according to which there is post-Spell-Out movement of the parenthetical, see Del Gobbo (2007). We will not discuss this analysis here, given the fact that it depends on several theory-internal assumptions (displacement-as-movement; overt vs. covert operations; a Y-model architecture with independent PF-LF levels of representation) none of which is independently justified.

153) 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 (153a) in a strictly paratactic reading, but in that case the structural description is different, not monotonically recursive. In Emonds' own terms,

> *the string John, who goes to MIT is not a constituent correferential 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 (154a), below); this seems to indicate that the structural relation between restrictives and their antecedents is closer than that between appositives and theirs.

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 RC]] structure as their antecedent. (154) illustrates this fact, with all possible combinations of *wh*- and COMP:

154) 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' (excluding determiners and quantifiers),

> *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'*

A semantic argument in favour of an [N' Comp] analysis is to be found in Partee (1975), who argues that a compositional interpretation for restrictive relatives has the head noun of NP and the relative clause (S' / COMP) form a derived category CN (common noun), which is then modified by Det / Quant. In this view, then, the correct interpretation of *the man that I saw* is 'the unique *x* such that *x* is a man and I saw *x*', which requires not that there is a unique man, but rather that there is a unique man that I saw (see also Bach & Cooper, 1978). If, on the other hand, restrictive relatives were adjoined to NP (where NPs are assumed to rewrite as Det, N') would yield an inadequate compositional interpretation (bear in mind that in a Montagovian setting each syntactic rule has an associated semantic interpretation rule; see Partee, 1975: 213, 223 point 9; and for a very similar perspective in a different syntactic framework, the presentation of rule format in Gazdar, 1981: 156).

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

155) 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 ((156a) is taken from McCawley, 1982: 96; (156b) is taken from McCawley, 1998: 450):

156) 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 a restrictive relative (as in (157a)); that is not possible with appositive relatives (as in (157b); examples taken from McCawley, 1998: 446):

157) 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. Emonds (1979) restricts the possibilities of co-appearance in the following descriptive statement:

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

Examples of this structure (Appositive+Restrictive) are given in (158):

158) a. We found that movie, which cost plenty, that you so highly recommended (Emonds' ex. (22))[42]
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. (15b)).

---

[42] 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 indeed noted by Emonds:

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

In contrast to the distributional condition formulated by Emonds (and cited above) with respect to restrictives following appositives, an appositive *can* follow a restrictive relative clause even if it is not the final constituent in the string:

159) 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 (92) 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 Complex NP, CNP, henceforth). Let us consider (148c) again, repeated here as (160):

160) Every psychologist who$_1$ Emily talked to who$_2$ insisted on helping ended up making a disaster

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 (they must have the same IL translation, otherwise they cannot be correferential). 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 antecedent size can be represented as in (161):

161) [[Every psychologist] who$_1$ Emily talked to] who$_2$ insisted on helping

Antecedent of

Antecedent of

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. This view is compatible with the observation in Bach & Cooper (1978) that a relative clause can denote properties of second order: *who insisted on helping* can denote (in these authors' terms) not only the property of insisting on helping asserted of *x*, but also the property of insisting on helping *and* having property P (in the particular case of (161), P is *Emily talked to x*). Because there is no principled limit to this process, this gives us the desired recursive structure, indefinitely (see also Bach & Cooper, 1978: 149).

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 (161), with its recursive semantic interpretation and monotonically growing antecedent size does *not* correspond to the interpretation of (153a), repeated here as (162):

162) *John, who$_1$ goes to MIT, who$_2$ likes math, will get a job

Antecedent of

NOT antecedent of

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 (153), as opposed to the situation in (161). 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 ∋ *who$_1$* was ordered with respect to G' ∋ *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 (163) below for a sequence of *restrictive* relatives is impossible to obtain, because restrictive relatives are monotonically recursive:

163) [[Every psychologist]$_i$ [who$_i$ Emily talked to] [who$_i$ insisted on helping]] ended up making a disaster

But a non-recursive, flat structure like that is precisely the kind of structural description that we have in the case of several *stacked* appositives, as in (164) (an example encountered in the wild):

164) 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, they define a *finite-state* sequence:

164') 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 (2016a), the appositive relatives in (158) and examples of the sort are *que-coordinated*, each being a state in a *Markov chain* (see also **Section 11**). The order among the appositives in cases like (164) is strictly linear, with parataxis being the only structural option to save the representation. All the relative clauses in (164) have the same antecedent, [a man]: note that we could shuffle these relatives around and still get a well-formed sentence. In other words, the sub-graphs corresponding to the relative clauses are all linked at *a man*': the ρ-set of (164), with simplifications (and unifying all cycles), is (165)

165) ρ = ⟨(spare, *man*'); (spare, *effort*'); (do, *man*'); (do, *everything*'); (set up, *man*'); (set up, *things*')…⟩

All the linked relatives form a graph whose global interpretation constitutes the contextually relevant set of properties that define the proper name Donald Trump, denoted by λPP{^*Donald Trump*} (Schmerling & Krivochen, 2017) for purposes of (model-)interpreting (164) at *t*.

Let us now turn to 'mixed' cases, in which we have an appositive and a RRC to be linearly ordered with respect to each other. Relevant examples are like (158a) and (159a), repeated here as (166a) and (166b) respectively (with minor annotations added):

166) 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, (166a) 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 (166a) 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 a *partial* ordering (but *not a strict ordering*) to be imposed in interpretation; however, we don't have a contradictory situation like *crossover* cases, in which a node had to be ordered both *before* and *after* itself in a trail. To the extent that (166a) 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*. It is important to distinguish the anomalies generated by a *partial ordering* imposed over a set of nodes (which can be solved outside of the grammar, by choosing an ordering as an interpretative hypothesis and seeing where things go from there) from the hard violation of *ordering* requirements that we have seen *crossover* cases generate.

Example (166a), with its dubious acceptability, forced us to make some additional considerations. (166b), 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 (167b) below as an extraposed version of (166a) –which does *not* mean at all that (167b) is ungrammatical, it just means that it cannot receive an interpretation in which *Relative clause extraposition* has applied to (166a)-:

167) 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 (167a) is that there is a set of children, a subset of which were brought, and in turn 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 which results 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 strict ordering to be imposed among the sub-graphs (and transitively, to all the nodes in each of these sub-graphs) in (161a). But the same interpretation procedure is not available for (167b): 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 (only *head-tail* recursive, not *freely* recursive; see Uriagereka, 2008: Chapter 6 for some discussion about this point) appositives that we proposed above is called upon again: the sub-graphs which constitute the structural descriptions for the appositives are *linked* at the node whose IL translation corresponds to the CNP *the children that you brought*; the different *wh*-words used (*who*, *which*) make no difference in terms of the connectivity properties of the graph that proves (167b) to be a well-formed expression of the language.

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 of the graph (which is the case in truly recursive restrictive relative clauses which take N+COMP antecedents) yields a straightforward order by having relative operators be functions of n^{th} order in Context-Free (or higher) semantic representations. In contrast, appositive clauses can only be *stacked* in finite-state sequences, displaying strictly paratactic dependencies between themselves. The structural description assigned to these is no different from other cases of non-scopal dependencies involving, e.g., VP adjuncts. As exemplified in (166), restrictive relatives may also establish paratactic dependencies, if they are all linked at a single node which corresponds to the antecedent. Because each local cycle corresponding to a RC is strictly ordered, but the global structural description for multiple appositive relatives does not specify an order among them, permutation of linear order is permitted *salva veritate* and without changing meaning.

9. On wh-interrogatives: aspects of syntax and semantics

We have dealt with *wh*-elements in previous sections (primarily *relative wh*-pronouns), but in those accounts we treated *wh*-phrases as an atom, without paying much attention to their internal structure because it was not required for our previous purposes. Thus, *wh*-phrases like *what* and *which of the books* would have received the same treatment in the initial presentation of *wh*-dependencies, since it made no difference for the cases under consideration. While that uniformity assumption (roughly, '*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'[43].

We need to clarify some issues pertaining to the syntax and semantics of *wh*-interrogatives (we will not deal with *whether*-interrogatives here). *In principle*, the structure of a *wh*-interrogative is not a lot different from that of a declarative sentence, in terms of GF: in (168a) and (168b), *who* and *what* are still Subject and Object:

168) a. *Who* bought a book?
b. *What* did John buy?

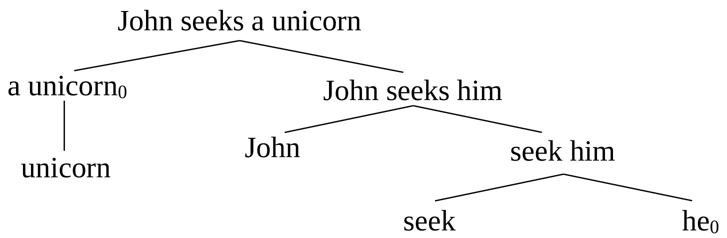
The preliminary ρ-sets of (168a, b) are thus (169a, b) respectively:

169) a. ρ = ⟨(buy, *who*'), (buy, *book*')⟩
b. ρ = ⟨(buy, *John*'), (buy, *what*')⟩

In the ρ-sets (169a, b) the *wh*-words *who* and *what* are assigned GF according to the rule (58). A question that arises now is whether there are aspects of the semantics that are determined by syntactic structure (as this is understood here, namely, connectivity patterns between nodes) and which we have not represented adequately in (169). In the framework of Montague (1973), we can say that *wh*-words bind an occurrence of an indexed pronoun; Karttunen (1977) builds on this and proposes an interrogative operator **?** having scope over a proposition and the *wh*-word as being external to it, then a syntactic rule replaces **?** with an indexed *wh*-word and deletes the pronoun in the preposition that has the same index. This is so because an interrogative sentence translates to an expression which denotes a (possibly unary) *set* of propositions; these propositions constitute an answer to the question. Therefore, the translation of *What did John buy?* translates as an expression which denotes the set of propositions for each thing that John buys, that he bought it. The way in which we have phrased this is not accidental: it mirrors the syntactic structure assigned by an analysis tree to a *wh*-interrogative in Karttunen's (1977) treatment.

This is not a process exclusive to *wh*-interrogatives: *scope* can be represented in this way more generally. For instance, let us consider Montague's (1973: 228) analysis tree for the *de re* reading of *John seeks a unicorn* (omitting the rules that have applied at each point):

170)



[43] 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 Montague, the subindexes simply indicate that those expressions are to be regarded as arguments of a function, that function being specified in the rule that applies to yield a certain expression. In Montague, if we interpreted the analysis tree as a PS tree, the *de re* NP would have scope over the intensional operator, but this is not essential to the analysis. For example, in Keshet's (2010) theory of *split intensionality* there is a scopal position *below* the intensional operator where a NP can be interpreted as *de re*. The take-home message that we want to focus on is that the *de re* reading involves an NP that appears structurally higher than its *de dicto* counterpart; this structural height has to do with the requirement that the NP has *scope* over the indexed pronoun that it binds (and this holds in both Montague's and Keshet's proposals). In this sense, Karttunen's (1977) proposal naturally extends the fragment in Montague (1973), which did not deal with *wh*-elements.

Karttunen's derivation of *wh*-interrogatives is exemplified in (171):

171) a. Who loves Mary?
b. *who*$_0$ **?** *he*$_0$ *loves Mary*
c. *Who loves Mary?*

In this proposal, the treatment that *wh*-phrases receive is does not differ in essence from the treatment that garden-variety NPs receive: *wh*-interrogatives feature a quantified NP (in our example, *wh*+someone$_0$) and a free variable in the form of an indexed pronoun (in our example, *he*$_0$). The interrogative in (171) translates as an expression which denotes a set that contains, for each individual that loves Mary, the proposition that he or she loves Mary. Identification of indexed categories allows for the construction of an appropriate semantic form, in exactly the same way in which *de re* and *de dicto* readings are constructed in Montague (1973); see Karttunen (1977: 24-25) for a precise formulation of the relevant quantification rule in PTQ-style. This much is not very different – superficially- from the MGG way of doing things:

172) a. COMP NP$_{[+wh]}$ loves Mary
b. NP$_{[+wh]i}$ t_i loves Mary
c. *Who* loves Mary?

This move is not without its semantic effects: we are saying that someone loves Mary, and we want to know the identity of that someone. A *wh*-interrogative is, semantically, a request for the identification of an individual or set of individuals via *wh*-phrases which informationally constitute the *presupposition* such that it or they fulfil the requirements specified in the *focus*. Syntactically, this reading is represented by having the *wh*-element over the proposition via a *root transformation* (regardless of whether it is successive-cyclic or not, it always targets the *root* of the tree; see Postal, 1972 for further discussion).

Similarly, in Metagraph grammar (Postal, 2010), scope can be represented in terms akin to Montague's: in a sentence like *Mary saw no student*, whose underlying structure would be [<[no student]$_1$> [Mary saw DP$_1$],] '*each member of the set of coindexed DPs would represent a single phrase, each X occurrence marking a separate arc sharing as head the single phrase represented by X*' (Collins & Postal, 2014: 14). The representation assumed in Collins & Postal's work is, for all present intents and purposes, equivalent to Montague's. In all these frameworks (phrase structure grammars, model-theoretic approaches), the scope of a node is indicated by its height in a tree/graph, so certain weak equivalences are not surprising given otherwise very different foundational assumptions.

In the light of this brief discussion, it seems that we need to revise the preliminary ρ-sets in (169) in order to have a proper representation of scope. But, how exactly?

A cautionary note is in order before continuing. It is almost unnecessary to say that *wh*-movement is one of the core issues in grammatical analyses of English and has motivated an enormous body of literature from generative (transformational and non-transformational) and non-generative approaches; we will not attempt to untie the Gordian knot of English *wh*-dependencies in this work. We will, however, propose a way to look at the problem which, to the best of our knowledge, has not been pursued before (although there are connections with traditional TAG analyses; see e.g. Kroch, 2001; Kroch & Joshi, 1985: §6, 1987[44]; and obviously Montague grammar).

The question we were dealing with is how to amend the ρ-sets in (169) to have a better representation of the semantics as read off by the structural description. A possibility we will suggest here, much research pending, is to have the *wh*-element dominate what would otherwise be the *root* node of the relevant graph (Emonds, 1970; Postal, 1972). That is, to have (173) instead of (169):

173) a. ρ = ⟨(*who*', buy); (buy, *who*'); (buy, *book*')⟩
b. ρ = ⟨(*what*', buy); (buy, *John*'); (buy, *what*')⟩

In (173), *who* and *what* are assigned a GF in the configurations specified in (58); those configurations remain identical in (173) and the preliminary ρ-sets in (169)[45]. We have added another visit to the *wh*-nodes, at the very beginning of the trail: this position corresponds to the place of the operator **?** in the IL translation of interrogatives in Karttunen (1977). The second visit in the ρ-sets in (173) corresponds to the position of the bound indexed pronoun within the proposition, which is the structural context in which the *wh*-element receives a GF. Note that there has been no need to dissociate operator from variable: all that differentiates them is their position in the trail that corresponds to the graph assigned to an expression. There is no need to have empty terminals: in this respect, our approach differs from the treatment that long-distance dependencies receive in TAG (Kroch, 2001; Frank, 2006): elementary trees in a TAG are generated by a PSG which follows the usual constraints we can find in MGG: projection, and the SMC. This is not necessarily an impediment in adapting the concepts of *substitution* and *adjunction* to the present context: it is, however, crucial to amend these definitions such that the notions of *frontier* and *internal nodes* in a tree (sets of terminals and non-terminals respectively) are ditched in favour of *generalised adjunction*: the crucial requirement to be met is identity in IL translation between the relevant nodes in distinct cycles to be linked. It is a good moment to remind the reader that there is no *overgeneration* to worry about, insofar as the present theory is *not generative*: we are not recursively enumerating structural descriptions or building phrase markers via recursive combinatorics, we are simply *describing* the connectivity patterns in expressions of the language.

A relevant question we need to ask at this point is 'what is the nature of *wh*-elements'? The Standard Theory answer was that *wh*-elements emerged transformationally:

T_{w2}: Structural analysis: NP — X

---

[44] The reader may also want to compare the present approach with that of Frank (2006), who implements TAGs in a phase-based Minimalist architecture.

[45] The approach pursued here, in which *wh*-movement is an *order changing rule* (and therefore grammatical relations are preserved without the need to invoke additional elements in representations), trivially satisfies Kroch's requirement without formulating further constraints (local or global) as part of the grammar:

> *we never want to allow derivations under which thematic roles, once established, are altered by further adjunctions, and we will block such derivations by, in every tree, placinga particular local constraint on every node that is assigned a thematic role by a governor.* (Kroch, 2001: 11)

Structural change: $X_1 — X_2 \rightarrow wh + X_1 — X_2$
where wh + animate noun → who
wh + animate noun [sic] → *what*

Chomsky (1957: 112)

That is, argumental *wh*-elements were (indefinite) NPs with a diacritic feature [+*wh*]. This diacritic has remained throughout the history of transformational generative grammar, recently in the form of an uninterpretable feature which has to be checked against a *Wh*- head (Epstein, 2015; Chomsky, 2015). The transformation above is supposed to account for two things: first, that the result is a *wh*-word (*who*, *what*); second, that these *wh*-words have the distribution of NPs. Here we are not concerned with the morpho-phonological form of expressions, but their distribution is something we do care about. Note that in our analysis of relative clauses the so-called *wh-operator* or *wh-pronoun* is actually another visit to the node that corresponds to the antecedent of the relative clause; in other words, the relative clause and the NP it modifies are *linked*:

174) a. The man who came was drunk
b. Cycle 1: [The man was drunk]
Cycle 2: [the man came]
ρ_1 = ⟨(be, drunk); (drunk, *man*’)⟩
ρ_2 = ⟨(come, *man*’)⟩

In (174) cycles 1 and 2 are linked at the node *man*’, one visit to which –we assume- corresponds to what surfaces as *who* given specific structural conditions. We have not gotten into exactly how this happens and, because we are concerned with structure and not with how that structure surfaces, we will not get into it in detail either. The important point here is that the reasoning that we have applied to relative clauses (restrictive and appositive) does *not* hold for *wh*-interrogatives, because there is no *antecedent-operator* relation. It is not possible to say that in (168) *who* and *what* link cycles or indeed that they have ‘antecedents’. They need to be syntactic objects in their own right. This means that interrogative *wh*-words will receive a different analysis from relative *wh*-words, on semantic basis.

We agree with Karttunen (1977: 18) in that the assignment of *wh*-words to a category requires us to have a model for the semantics of the sentences in which they occur. With ordinary NPs that was not a problem, but *wh*-words –as we have seen- present different challenges. In order to flesh these challenges out (and how they can be addressed), we need to introduce some basic aspects of Montague semantics. Montague (1973) defines two atomic categories *e* and *t* for *entity expressions* and *truth value expressions*; *e* and *t* are then used to recursively define categories which are assigned to expressions more familiar to the ordinary working linguist. In PTQ, proper NPs (or ‘terms’) are assigned to the category *t*/IV (they combine with intransitive verb phrases to yield *t*, which we can think of as analogous to *S*(entence) in PSG: PTQ has no category *S* distinct from *t*. The point of using the symbol *t* is its mnemonic value: *t* stands for ‘truth (value)’, and truth values are the kinds of objects that sentences denote. Common NPs are *e//t*, where double slashes are used to indicate that a category differs from its single-slashed counterpart in its semantic role but plays an identical syntactic role (PTQ: 222). In this context, Karttunen (1977: 19) defines interrogative *wh*-words as *t*//IV, that is, a modified *t*/IV; *wh*-words play the same syntactic role as ‘terms’ (recall also that R-expressions and *wh*-traces are assumed to behave in the same way for Binding Theory purposes), but differ in their semantic role: *wh*-words are equivalent to existentially quantified (indefinite) common NPs (Karttunen, 1977; Reinhart, 1998: 44). Reinhart (1998) identifies some problems with Karttunen’s original proposal, and builds on a modified version of it. Interestingly for our purposes, she introduces a *choice function* (CH) for lexically restricted *wh*-phrases, such that (175) is analysed as (176) (taken from Reinhart, 1998: 41):

175) Which lady read which book?

176) a. for which $\langle x, f\rangle$ (lady(x)) and (x read f(book))
b. {P|($\exists\langle x, f\rangle$) (CH($f$) & lady($x$) & P =ˆ($x$ read f(book)) & true(P))}

Where P is a set of true propositions such that the conditions that follow hold. Much detail aside, what matters to us is that the meaning of a bare *wh*-word like *what* or *who* can be assimilated to lexically restricted *wh*-phrases: *which lady* in (175) means 'choose an entity that belongs to the set of ladies (such that…)'; *who*, in the same context, means 'choose an entity (such that…)', with the added specification that the entity be animate (for inanimate entities, we have *what*). The 'such that…' clause depends on the structure where the *wh*-word or phrase appears; in Reinhart's analysis of (175), that would be P =ˆ(x read f(book)) & true(P)). Karttunen (1977: 19) proposes that the IL translation of bare *wh*-words, say *who*, is the same as the IL translation of an indefinite common NP, like *someone*: there is an existentially quantified term that corresponds to a member of a set of entities, the *wh*-operator introduced the choice function that selects an entity from that set based on whether it satisfies the conditions imposed by the predicative structure in which *wh* occurs. In Karttunen's approach, *who* and *what* have IL translations *who*' and *what*': $\hat{P} \bigvee x P\{x\}$[46]; again, whether we get *who* and *what* depends on whether x is animate or not. Crucially, x is the entity over which *wh* quantifies, it is semantically distinct from *wh* itself. This becomes clearer when we consider the IL translation that Karttunen assumes for a restricted *wh*-phrase like *which girl*: $\hat{P} \bigvee x[girl'(x) \wedge P\{x\}]$. This impacts directly in our framework, because we have nodes that corresponds to the IL translation of basic expressions, we can have *who* and *what* as nodes without the need to assume that there is a phonologically empty N *in the syntax*: this view contrasts with Reinhart's, who proposes that bare *wh*-words take a null N complement, yielding a structure like (177):

177) [who [$_N$ e$_i$]] (Reinhart, 1998: 44; see also Panagiotidis, 2001: Chapter 5)

In the generative view, empty heads are nothing to worry about, and their proliferation is not seen as anything undesirable. We, however, want to avoid them at all costs: paraphrasing Postal (1972: 215), the problem is not that descriptive adequacy cannot be attained with empty heads (or formal features); rather, the problem is that the theory becomes unrestricted and unnecessarily stipulative[47].

As an interim summary, the scenario we are left with is then one in which we have the following IL translations:

178) a. *who* / *what* $\equiv \hat{P} \bigvee x P\{x\}$

[46] The symbol $\bigvee$ must not be confused with $\vee$: the former, in Montague (1973: 229), is used to denote *existential quantification* (i.e., $\exists$), the latter is the logical disjunction *or*. In turn, *universal quantification* is denoted by $\bigwedge$

[47] Postal's fragment, which deals with feature marking (not with empty nodes), is well worth citing, as it is as valid today as it was in 1972 and still serves as a reality check for theoretical syntacticians:

> *the whole feature-marking proposal has no independent justification. The point is not that descriptive adequacy is unachievable in this way, but rather that it is achievable under the assumption of successive cyclicity only at the cost of having available the overly powerful device of marking arbitrarily selected nodes with arbitrary rule behavior coding features. It is strange that this powerful device should be appealed to by authors who are often at pains to stress the need for restricting the power of syntactic theory, and who have often objected to other approaches on just this ground* (Postal, 1972: 215)

b. *which* N ≡ $\hat{P} \bigvee x[N'(x) \wedge P\{x\}]$

As an example, the full translation of (171) applying Karttunen's quantification rule would be (171') (adapted from Karttunen, 1977: 20):

171') Who loves Mary?

a. *who*' ≡ $\hat{P} \bigvee x P\{x\}$

b. ?-he$_1$-loves-*Mary*' ≡ $\hat{p}\ [^{\vee}p \wedge p = {}^{\wedge}love_*'(x_1, Mary')]$ (where $love_*'(x_1, Mary')$ is the IL translation of *there is an individual x with index 1 and x_1 love Mary*, *Mary*' is the IL translation of the NP [Mary])

c. who-loves-*Mary'* ≡ $\hat{p}\ [who'(\hat{x}_1\ ?\text{-}he_1\text{-}loves\text{-}Mary'\ (p))]$

In terms of how many nodes we need to represent the relevant connections, note that the IL translation of *which N* includes the IL translation of N, *N'* (or, unabridged, $\lambda P[[^{\vee}P](^{\wedge}N)]$, see **Section 2** for discussion). Also, the complement of *which* can be an object of arbitrary complexity:

$$179)\ which \begin{Bmatrix} boy \\ student\ of\ mathematics \\ picture\ of\ himself\ that\ you\ liked \end{Bmatrix} \ldots$$

It makes sense to have *who* and *what* as single nodes and assign *which*-N a more complex structure, depending on their complements.

There is an aspect of the relation between the **?** operator and the pronoun that is particularly interesting to the effect of formulating conditions over dependencies in our graphs. Let us 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 *lexically* bound: going back to Karttunen's IL translations, **?** must quantify over an indexed *he*. This quantification requirement is met in the IL translation of *wh*-words. But syntactically we need to make sure that *wh*-arguments (i.e., Subjects and Objects which are questioned) can not only be interpreted as operators, but also receive GF within the proposition. In other words, we need to visit each *wh*-element *twice*: at the beginning of the trail, in the position indicated with **?** (ordered first in the ρ-set of the relevant graph), and then as the tail of an arc in which a GF can be assigned to it. In the Montagovian view, it is the pronoun that receives a GF and not –*stricto sensu*- the operator. But in our graphs the operator and the pronoun are simply contextually distinct visits to the same node: a *wh*-node is both, using Ladusaw's (1980: 112) terms, a *trigger* and a *variable*, walked on in distinct contexts (where, recall, the *context* of a node is the set of nodes that it *directly* dominates and it is *directly* dominated by within a cycle)[48].

[48] It is important to distinguish the notions of *semantic* and *syntactic* saturation of a predicate (Chung & Ladusaw, 2003: Chapter 1); here we are primarily concerned with how semantics can inform an adequate theory of syntax: how to appropriately construct and restrict the theory of syntax so that it assigns each expression of the language a structural description that adequately represents semantic dependencies between grammatical objects in that expression.

Specifically, let v_{wh} be a *wh*-node (i.e., a node whose IL translation is either (178a) or (178b); we will not deal with adverbial *wh*-words here). Then, we propose the following well-formedness condition over *wh*-trails:

180) There must be a unique trail *t* in G such that $\rho^* \ni (v_{wh}, v_{wh})$

The idea is simple enough: open scope must be closed, and what closes it is the 'pronoun' bound by **?** in Karttunen's (1977) presentation. Furthermore, scope closing must be *unambiguous*, this is the significance of *unique trails* in the more general context of a *strict ordering* requirement on graphs. The only thing is, for purposes of representing connectivity patterns, the nodes that correspond to the operator **?** (the *trigger* in Ladusaw's terms) and the pronoun $\mathbf{he}_n$ (Ladusaw's *variable*) is *the same node, visited twice in a local trail*, where 'locality' is to be understood as a dependency that satisfies the definition of *licensing* (note that there is no mention of specific nodes in the definition of *locality*: no barriers, phases, or the like). There is a further requirement, which is that there be a biunivocal relation between operators and indexed pronouns: of course, because we are dealing with *trails,* in reality we are dealing with multiple visits to the same node, in distinct contexts. This is important because the requirement, as it stands, relates scope contexts (i.e., root contexts) to restrictor contexts (i.e., non-root contexts) in terms of indexed categories, not lexical items. Because expressions with the same IL translation have the same index in the algebra of the language, by referring to an index we are simply making reference to an expression of the language in isolation from its structural context.

Recall that we have introduced a requirement for graphs to be *strictly ordered*: in the relation between both contexts of occurrence of v_{wh} the instance that is ordered first opens scope, and the one that is ordered after that one closes scope. (179) can thus be interpreted in relation to what Sternefeld (1998a) 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* [Negative Polarity Items] *as well as bound variables must be* [in] *the scope of (i.e., c-commanded by) the operators they depend on* (Sternefeld, 1998a: 151)

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* if we are talking about *transitive dominance*) between the relevant nodes. The other difference (and perhaps a more important one) is that in the case of operator-variable relations, the 'dependency' is not determined by scope since *operator* and *variable* are the same node: the problem must thus be formulated in a different way.

The relevant question, in terms of the present proposal, is whether there is a unique directed trail *t* in G (which can be a single cycle or a graph combining several cycles) containing v_{wh} such that

(a) There are as many visits to v_{wh} in G as there are predicates selecting v_{wh} if v_{wh} corresponds to an argument plus one which dominates the root, or
(b) There are as many visits to v_{wh} in G as there are predicates being modified by v_{wh} if v_{wh} corresponds to an adjunct (*why, how, where* in non-locative construals) plus one which dominates the root

In the case of argumental *wh*-elements, we additionally require that the structural context of v_{wh} in each case satisfies the semantic requirements imposed by neighbouring predicates (in other words:

does v_{wh} in at least one structural context receive a GF from a predicate, thus partially satisfying that predicate's valency?).

It is worth noting that having *wh*-movement be modelled with two visits to the node standing for the IL translation of the *wh*-element captures Postal's (1972) observation that there are reordering transformations that do not proceed in a strictly cyclic manner (which he calls *U-rules*, for *unbounded rules*), whose application need not resort to intermediate reconstruction sites. Postal argues that *wh*-movement is one of these (at least in English)[49]; we would like to restrict that claim by saying that *wh*-movement is unbounded when the structure crossed over by the *wh*-dependency grows monotonically: for example, a sequence of *equi* verbs, propositional attitude verbs, *object*/*subject raising* verbs, and the like (plus possible combinations of these). In this way, the ρ-set for a monotonically growing sentence like (181a) is (181b) (indices and gaps are purely illustrative):

181) a. What$_i$ does Mary think that John believes that he should tell Peter that Susan read __$_i$?
b. ρ = ⟨(*what*', think); (think, *Mary*'); (think, that); (that, believe); (believe, *John*'); (believe, that); (that, should); (should, tell); (tell, *John*'); (tell, that); (tell, *Peter*'); (that, read); (read, *Susan*'); (read, *what*')⟩

In each case, a verb that takes a causal complement directly dominates that clause's root. But what we really care about in (181b) is that there is no need to assume additional occurrences of *what* in intermediate positions: *wh*-movement is *not* modelled as a successive-cyclic rule (*contra* Chomsky, 1973, 1986, 1995 and subsequent work in GB and Minimalism). In the terms of Abels & Bentzen (2012: 433), the dependency set we have in (181) is *uniform* in the sense that all intermediate elements in what we can descriptively refer to as the 'movement path' (the structure that mediates between both visits to *what* in the trail defined by the digraph defined in (181b)) are equally *unaffected* (in this aspect, our proposal is close to TAGs). As argued in detail in Krivochen (2015a, 2016a, 2018, 2019a), in turn based on Uriagereka (2002a), non-monotonicity generates syntactically opaque domains: local monotonically derived units allow for internal accessibility, but as soon as the computational dependencies change that local unit becomes opaque for purposes of operations triggered from the outside. For our present intents and purposes this means that non-monotonic growth of a graph generate local *islands* (in the sense of Ross, 1967), whose internal nodes (i.e., everything dominated by the root of the relevant sub-structure) are inaccessible for operations triggered by external objects: this captured the concept of *locality-as-impenetrability* that was a staple of the Standard Theory and GB but without resorting to designated nodes like barriers or cyclic nodes. The perspective adopted in those works was eminently *derivational* and *proof-theoretic*: the basic idea was that phrase markers display local properties which depend on the semantic dependencies between objects; we called this *mixed computation*. It is useful to consider what things look like from a different point of view from that adopted here, for the perspectives can in principle converge:

> *computationally uniform dependencies define local domains within which we can probe: cycles. In this context, we do not need to appeal to substantive elements to define cycles (including*

[49] Bach & Horn (1976: 273, ff.) identify cases in which a Subjacency-like view of locality results inadequate. Interestingly, they say that '*Wh Movement is not always constrained by Subjacency*' based on examples like (i)

i) Who did Bill believe that John told Ralph to kill?

If *Wh-movement* is an unbounded rule, its immunity to Subjacency -at least in some cases, specifically those not involving *islands*- could be captured in a unified framework which dispenses with obligatory intermediate landing sites.

> *bounding nodes, as in the Standard Theory; barrier nodes, as in Chomsky, 1986a; phase heads, as in Chomsky, 2001, etc.); rather, cycles emerge from the dynamics of the computation.* ***Simply put, when we need to change computational dependencies because a given kind of grammar is either too weak or too powerful, the current cycle is closed.*** (Krivochen, 2018: 224. Highlight ours)

In the present framework, the configurations described proof-theoretically in terms of *non-monotonicity* arise when *self-contained graphs* are connected. Note that we can restrict the possible dependencies across cycles without resorting to non-audible structure or multiplying the levels of representation. The possibility of describing structures with varying kinds of computational dependencies (combining sub-structures of finite-state, context-free, and context-sensitive complexities in structural descriptions) is a crucial factor when determining the empirical adequacy of a theory of syntax: we don't want to assign expressions structural descriptions that are 'too rich', nor 'too restricted' in terms of the semantic dependencies that those structural descriptions establish between elements, as we briefly discussed in (9) above (see also Lasnik, 2011; Lasnik & Uriagereka, 2012; Saddy, 2018).

9.1 Simple wh-questions

In this section we will present sample analyses for a variety of *wh*-interrogatives involving a single *wh*-word or phrase.

Subject and Object *wh*-interrogatives are straightforward enough, we already saw examples with bare *wh*-words in (169) and their analysis in (173). We have not, however, given an explicit analysis of Subject and Object *wh*-interrogatives with lexically restricted *wh*-phrases. The relevant examples are of the type (182a) and (182b) respectively:

182) a. Which student read that book?
b. Which book did John read?

In the line of what we said above about the semantic structure of *which*-N, in which the *wh*-instantiates a *choice function* over the set of entities denoted by the N, we will analyse *which N* as ρ = ⟨(which, *N'*)⟩: the *wh*-word dominates the N. In this relation, the *wh*-word quantifies the N, and the N restricts the *wh*-word[50]. Transitively, anything that dominates *which* will dominate *N* also. The ρ-sets of (182a, b) are thus (183a, b) respectively:

183) a. ρ = ⟨(which, *student'*); (which, read); (read, which); (read, *book*')⟩
b. ρ = ⟨(which, *book'*); (which, read); (read, *John*'); (read, which)⟩

If we turn our attention to more complex examples, like (184a, b), the advantages of an approach that maximises local connectivity become even clearer:

184) a. [Which picture of himself$_j$]$_i$ did John$_j$ say Mary likes __$_i$ (Uriagereka, 2011: 5, ex. 11)
b. John$_i$ wondered [which picture of himself$_{i/j}$]$_k$ Bill$_j$ saw __$_k$ (Chomsky, 1995: 205, ex. 36 a)

There are several things to note in these examples. One is that the *self*-forms that appear are *not* anaphors in the sense of **Section 5.2.1**, there are no *parallel* edges in the graphs that define the structural descriptions for these expressions. These *self*-forms have the interpretation of Binding-

[50] 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 hypothesise, the operator and its complex must be connected at *c-structure*.

theoretic *pronouns*, there is no *reflexivity* (which, in Reuland & Reinhart, 1993, requires identity of the 1 and 2 in a sentence). Another important aspect of the sentences in (184), and perhaps more importantly for our current purposes, they display a complex structure as complement of *which*; this structure contains a node that needs to be multidominated in order to yield the appropriate correference: *himself* is not a reflexive, but it is referentially bound to an expression that is superficially outside the *wh*-complex (which we have indicated with indices).

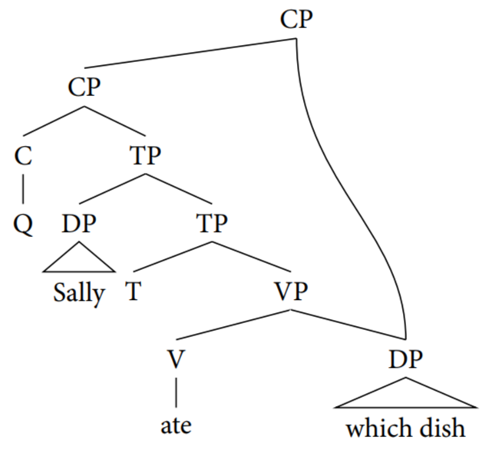
Let us focus on the first example for concreteness. The derivation of (184a) along classical MGG lines requires, at least, the following:

- A set of phrase structure rules to generate the string *John said Mary likes which picture of himself*
- A movement rule that displaces the syntactic term [which picture of himself] from its base position as the complement of *like* to the 'left periphery', call it *Wh-movement* (Chomsky & Lasnik, 1977: 434), *Move-α* (Lasnik & Saito, 1992), *Internal Merge* (Chomsky, 2000), etc.
- An indexing rule that keeps track of occurrences of syntactic terms. It needs to be able to assign the same index to *John* and *himself*, but also to *which picture of himself* and the gap in the complement position of *like*.
- A rule that inserts the auxiliary *do* to spell-out tense and agreement features

Interestingly, much of this complication emerges because interrogatives are assumed to derive from declaratives via movement transformations mapping trees onto trees. The declarative version is assumed to be generated by PSR or some equivalent mechanism (e.g., Merge), based on an *a priori* clausal skeleton (simplifying quite a bit, Complementiser-Tense-Verbal projections –but see Cinque, 1999 and much related work for an expanded functional structure-); transformational rules then move things around extending the phrase marker and leaving derivational crumbs behind which allow interpretative systems to keep track of what has gone where (traces, copies, slash-features…). Somewhat recently, there has been work emerging on what is called 're-Merge' theory of movement (e.g., Johnson, 2014, 2016), whose generative power is exactly the same as the Copy+Merge theory as far as we can tell, only differing in the diagrams they allow for L-trees (not in the L-trees themselves: in other words, the drawings change, but not the formal relations between elements in the tree as a mathematical construct). The re-Merge theory would assign a phrase marker like (185b) to (185a):

185) a. Which dish did Sally eat?

b.



Prima facie, the representation in (185b) is rather close to our graphs, minus non-audible structure (all nonterminal nodes and labels corresponding to phrasal projections). But it is not. Johnson (2014: 268) argues that

Note, then, that a phrase which resides in two positions, as which dish does in [(185b)], *need be semantically interpreted in only one of those positions. The normal requirement that everything in a syntactic representation must be interpreted by the semantic component must be allowed to permit (29).*

(29) If a term is in more than one position in a phrase-marker, it need be semantically interpreted ***in only one of them***. (highlighting ours)

Our model is based on the diametrically opposite view that if a node occurs in a certain context (where, remember, the *context* of a node is the set of nodes it is directly connected to) then it is because that node is assigned an interpretation in that context. For example, in the ρ-set that corresponds to (185a),

186) ρ = ⟨(which, *dish*') (which, eat); (eat, *Sally*'); (eat, which)⟩

which dish receives a GF in the context *e*<eat, __>, and opens scope in the context *e*<__, eat>. An advantage of *dependency*-based approaches, including ours (to a certain extent) is that semantically-vacuous positions or occurrences are eliminated, because the motivation to have an element in a certain position in a structural description *is semantic*. Building a compositional semantic interpretation for (185) requires *both* an occurrence of *what* in a scope position and in a position where a GF can be assigned, *neither is more basic than the other*. This contrasts with the transformational approach, whereby semantic interpretation was initially restricted to Deep Structure because transformations did not change meanings (Katz & Postal, 1964), then Surface Structure played a role since it contains all the same information as Deep Structure *plus* traces and aspects of suprasegmental phonology determining *focus-presupposition* dynamics (Chomsky, 1970a; also Schmerling, 2018: xii), and in orthodox vanilla Minimalism, an external component, C-I. Two assumptions are behind these treatments: (i) interrogatives are derived structures, and (ii) structural descriptions assigned by the grammar are phrase-structure based trees. Neither assumption is empirically justified, thus these are better thought of as *axioms* of transformational generative grammar.

But, what happens if we take interrogatives and declaratives as equally 'basic' structures (given the fact that there are no 'derivations' in model-theoretic approaches), and only care about providing an exhaustive map of connectivity between nodes in a given expression of the language? In that case, we can do away with the requirements that derive from PSGs, and begin by considering (a) how many cycles we have and (b) where they are linked. This allows us to provide the following ρ-set:

187) Cycle 1: [John say [Cycle 2]]
Cycle 2: [Mary likes which picture of John]

ρ_1 = ⟨(say, *John*'); (say, like)⟩
ρ_2 = ⟨(like, *Mary*'); (like, which); (which, picture); (picture, *John*')⟩

The two cycles are linked at *John*; and the matrix verb dominates the root of what is its complement, the clause [Mary likes which pictures of himself] (thus, it transitively dominates every node that root dominates).

It is important to note that we have left aside a number of questions pertaining to the semantics of *picture*-Ns; in particular, we treated *picture* as if it was an argument, but it would be also possible to take it as a function with domain *John*. In this case, and if we assume Montague's (1973) *translation rule* 1, given *John* in the domain of *picture of x*, then *John*' translates as *picture of* (*John*). As for prepositions (in general), we follow Schmerling (2018a: 63) in considering them *case* *desinences* rather than *case assigners* (as is common in MGG, see e.g., Uriagereka, 2008: 146); it is worth highlighting that the mechanisms of Case in CG are quite different from Case as seen in MGG (see Schmerling, 2018a: Chapter 5 for further discussion of Case in German).

As for simple *wh*-interrogatives where the *wh*-element is not an argument, we will assume here that so-called VP adjuncts dominate the predicate that they modify. This goes along the lines of some versions of (neo-) Davidsonian semantics for VP adjuncts (Davidson, 1967, 1969; this framework is assumed, e.g., in Uriagereka, 2008; also Higginbotham, 2000 and much related work; Maienborn, 2011 gives a general perspective of event semantics including Davidsonian and Neo-Davidsonian approaches) in a specific (and perhaps *sui generis*) sense. Consider the following sentence:

188) Mary read a paper quickly in the park

A rough Davidsonian representation of (188) would go along the lines of (189):

189) ∃(*e*) [READ(Mary, a paper, *e*) & QUICKLY(*e*) & IN(*e*, the park)]

Because we have made no use of thematic roles in (189) to account for the relation of the event and its participants, that representation does not count as 'neo' Davidsonian (Maienborn, 2011: 811; Parsons, 1990), but this does not really matter in the present context. There is no need to invoke thematic roles for our purposes either, and as a matter of fact the order in which arguments are presented in Davidsonian representations is the same we have used in our strictly ordered sets (Subjects > Direct Objects > Indirect Objects). But here we are concerned with a different point: note that in (188) the modifiers *quickly* and *in* take the event as an argument (in the case of *in*, we are in the presence of a dyadic predicate, so the event is one of its arguments, *the park* being the other). Part of Davidson's contribution to the theory of semantic representation for natural language sentences is that

> *Adverbial modification is* […] *seen to be logically on a par with adjectival modification: what adverbial clauses modify is not verbs but the events that certain verbs introduce*.
>
> Davidson (1969: 298)

The relevant part of Davidson's quote is *logically*: the semantic interpretation of something like *really fast* (see (39)-(41) above) is licensed by the same syntactic mechanisms that the interpretation of *black suit* (see (29) above): nodes corresponding to adjectives dominate nodes corresponding to the nouns they modify, just like nodes corresponding to adverbs dominate nodes corresponding to the verbs they modify.

In this context, the ρ-set that we propose for (188) is (190):

190) ρ = ⟨(read, *Mary*'); (read, *paper*'); (quickly, read); (in, read); (in, park)⟩

The syntactic similarities between (190) and the Davidsonian representation (189) are rather evident. Now, let us consider two interrogative sentences that can be formed, questioning *quickly* and *in the park*:

191) a. Where did Mary read a paper quickly?
b. How did Mary read a paper in the park?

The ρ-sets of (191a) and (191b) are (192a) and (192b), respectively:

192) a. ρ = ⟨(*how*, read); (read, *Mary*'); (read, *paper*'); (in, read); (in, park)⟩
b. ρ = ⟨(*where*, read); (read, *Mary*'); (read, *paper*'); (quickly, read)⟩

The dominance sets for (192a) and (192b) are, for all syntactic intents and purposes, exactly parallel to those assigned to *wh*-interrogatives in which an argument is questioned; the crucial difference is whether the *wh*-element receives a GF in the graph; if not, then there is no pronoun for **?** to bind, in contrast to the structures considered above.

It is important to bear in mind that the only parallel between adverbial and adjectival modification is *syntactic* in nature (defined in terms of *dominance*), and has nothing to do with the generative custom of defining so-called 'lexical categories' using binary features [N] and [V]. We are *not* claiming that adjectives and adverbs belong to the same lexical category or anything of the sort; the only *categories* we acknowledge are CG-style indexed categories, which bear *no resemblance* to MGG's categories.

There are some points that we would like to insist on before closing this section, which pertain to what we see as advantages over transformational generative treatments of *wh*-constructions and other 'filler-gap dependencies' (Sag, 2010). In contrast to transformational treatments, there has been no need to multiply the entities: no traces or indexes have been invoked. The absence of indexes is particularly important, because free indexing can quickly get out of control in terms of strong generative power. But also it requires some specific level where indexes are assigned; for example, in Lasnik & Saito's (1984) encoding of the Empty Category Principle as feature-marking this level is Logical Form, however, the specific name that we give this level is inconsequential. The crucial thing is that by multiplying the levels (which in turn is needed by the multiplication of entities and derivational steps) a further need of mapping operations (usually stated as 'principles', e.g., Huang, 1982: 220; see also Williams, 2003: Chapter 1 for discussion within a framework that multiplies the levels of representation, and thus linking rules, even more).

In this context, there are some things to bear in mind. First, it really should not matter whether we have a node morpho-phonologically realised 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., Groβ & 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* phemonena 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.

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

In this section we will consider some English sentences which will force us to explore the limits of the formal mechanisms explored here. A theory that maximises 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). Relevant examples are like (193) and (194):

193) The man_i who shows he_i deserves it_j will get the $prize_j$ he_i desires
194) Every $pilot_i$ who shot at it_j hit the MIG_j that chased him_i

It has been noted in the literature that getting the indexes right on these sentences is problematic for a transformational approach to pronominalisation (famously, by Bach, 1970), but the difficulties do not end there: even if we left the specifics of indexing mechanisms aside, it is necessary to account for the referential dependencies probing anaphorically and cataphorically into domains which usually behave like islands, like relative clauses (e.g., *who shows he deserves it*, [which] *he desires*, *who shot at it*, *that chased him*) 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, 2015a, 2016a, 2018; 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*[51] to these sentences *requires* the system to be sensitive to local changes in computational complexity; what we need is 'mixed computation'. Let us illustrate this point using (193) as our example:

195)

a) The⁀man; the⁀prize; he⁀deserves⁀it ⟹ **Finite-state**

b) [The man *S'* [will get [the prize *S'*]]] ⟹ **Context-Free**

c) [*the man* [who shows [he deserves **it**]]] will get [**the prize** [*he* desires ***e***]] ⟹ **(Mildly) Context-Sensitive**

Note that the computational differences arise *within local derivational units* and pertain to relations of co-reference within or across graphs. From a *proof-theoretic* perspective, we have shown in previous work that 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 (see also Lasnik & Uriagereka, 2012; Lasnik, 2011). From this perspective, we can define syntactic *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 (see Saddy & Uriagereka, 2004; Saddy, 2018, and related work): as we briefly saw in **Section 9**, the proposal in Krivochen (2017) was that the change in computational complexity is what delimits local syntactic domains, not the presence of designated nonterminal nodes. In such a 'mixed' system, computational uniformity (the norm in MGG; see Jackendoff, 2011:

[51] 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 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 to capture semantic dependencies), but not procrustean. See Krivochen (2015a; 2016a, 2018); Krivochen & Saddy (2017) for discussion.

277-279 for some additional discussion) would in fact have to be *stipulated*. In (193a) we have uniformly binary branching minimal treelets, following Greibach (1965: 44) and Uriagereka (2012: 53), a finite-state grammar can capture all relevant dependencies within these substrings. (195b) features phrasal ‘constituents’ related within structures containing placeholders for root nodes (the roots of the relative clauses); center embedding pushes us up to CFG. The appearance of crossing dependencies in (193c) (note that the linear distribution of indices is *i…j…j…i…j*) forces us up at least ‘half’ a level, to ‘mild’ CSG (Joshi, 1985 and much related work; see **Section 4.1** above).

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. Under a Lees-Klima view, Bach (1970) argues, the Deep Structure of (193) would have to be:

> *The man who shows that the man deserves the prize that the man who shows that the man deserves the prize that the man... (ad infinitum) will get the prize that the man who shows that the man deserves the prize that the man who shows ... (ad infinitum)*
>
> Bach (1970: 121)

While it is possible to save the transformational view by restricting the applicability of *Pronominalisation* to specific contexts (e.g., pronouns in root clauses need not be transformationally derived, as argued by Lakoff, 1976: 329, ff. and Postal, 1969, 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 should a ‘transformation’ *Pronominalisation* be formulated, it 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, as we have highlighted a few times already, the surface elements *the man*, *who*, and *he* in (191) are *the same node* visited three times in a trail, 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, it is possible to recast the *Pronoun Rule* and the *Reflexive Rule*[52] of Lees & Klima (1964: 23) in terms of trails. Grouping together the conditions formulated in **Section 4** and **Section 5.2.1**, we can capture the Lees-Klima approach (recently revamped, e.g., in Hornstein & Idsardi, 2014: 14, ff. within a Minimalist framework) as follows:

196) **Pronoun rule:**
v_j can pronominalise v_i iff

a. v_i and v_j denote sortal entities
b. $i = j$, and
c. $(v_i, v_j) \in \rho^*$, *such that there is a unique trail that visits v_i and v_j in G and in this trail v_i is ordered before v_j.*

---

[52] Lees & Klima’s formulations are as follows:

> *Reflexive Rule: X-Nom-Y-Nom’-Z → X-Nom-Y-Nom’+Self-Z where Nom = Nom’ = a nominal, and where Nom and Nom' are within the same simplex sentence.*
>
> *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).

197) **Reflexive rule:**
Let v_j be a predicate and v_i an argument of v_j. Then, v_i is a reflexive iff
$\rho = \langle(v_j, v_i), (v_j, v_i)\rangle$

Or, alternatively, and under the same conditions,

…*iff v_i is the tail of parallel arcs*

In Bach-Peters sentences it is particularly interesting to see how many cycles we have and how they are linked such that the appropriate correference relations hold. Let us now give the cyclic analysis and ρ-set for (193): recall that (building on García Fernández et al., 2017) each *cycle* contains a single lexical predicate, plus its temporal / aspectual modifiers and nominal arguments:

198) 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 all cycles together we get (199):

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

And after *linking* all cycles we get the following ρ-set, where nominal nodes get assigned GF:

200) ρ = ⟨(show, *man*'); (show, deserve); (deserve, *man*'); (deserve, *prize*'); (get, *man*'); get, *prize*'); (get, desire); (desire, *man*'); (desire, *prize*')⟩

Note that the object position 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; the analysis here follows the lines of **Section 5**.

It is interesting to note that some aspects of the present analysis are prefigured in McCawley (1970), and it is possible that only minimal adjustments are needed in order to make it compatible with the variable-free semantic proposal of Jacobson (2000). McCawley (1970: 176-177) considers the following sentence, attributed to S. Kuno:

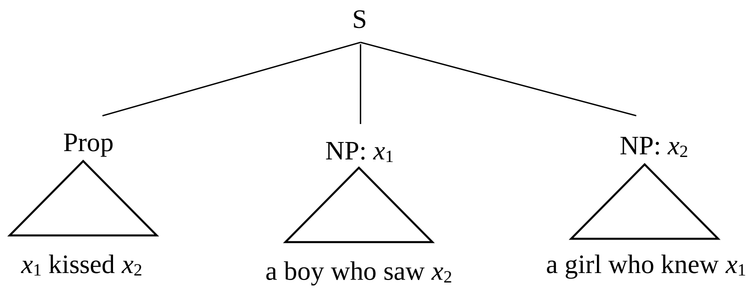
201) A boy who saw her kissed a girl who knew him

(201) is a run-off-the-mill Bach-Peters sentence (to the extent that these creatures can be said to be 'run-off-the-mill' at all). McCawley notes -like Bach, even though neither cites the other- that an approach to *pronominalisation* like the one presented in Lees & Klima (1963) requires an *ad infinitum* proliferation of antecedents if applied to *all instances of pronominal reference*. Like Lakoff (1976), McCawley proposes to keep *Pronominalisation* as a transformation (that is, he accepts that at least some pronouns are derived transformationally in local relations between NP nodes or S nodes), but makes an important change to the way in which the transformation is conceptualised with respect to the original Lees-Klima version:

> *Pronominalization consists not of replacing repetitions of a noun phrase by pronouns, but rather of determining which occurrence of an index will have the corresponding noun phrase substituted for it. Those occurrences of indexes for which the substitution is not made are then filled by pronouns* (McCawley, 1970: 176)

As it is, McCawley's proposal rested on notational conventions, namely, indexes. But this need not be: an occurrence of an index in a structural description, under present assumptions, is simply the occurrence of a node in a trail. That is: giving up the SMC and making nodes stand for IL translations makes it unnecessary to resort to indexes, for structural descriptions and structural changes in the transformational approach simply make reference to the context of a node (that is, the set of nodes that are *directly* connected to it). This said, it is interesting to take a look at the 'deeper structure' that underlies (201) in McCawley's conception (taken from McCawley, 1970: 177)[53]:

202)

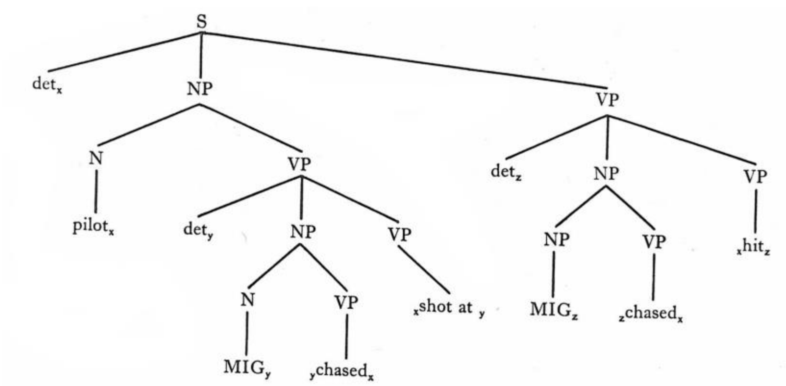


McCawley's proposal is to substitute x_1 and x_2 in Prop by the indexed NPs, leaving the indexed occurrences of variables *within* the NPs untouched; because these are not substituted, they surface as pronouns *her* and *him* respectively. Note that the cyclic domains in (200) coincide with those identified in our analysis of (191) (with the caveat that (191) has a complement clause as the object of *show*); the only formal difference being that in McCawley's approach, substitution of each variable for the relevant NP takes place sequentially and –presumably- at *shallow* or *surface* structure. The reason for this is that in his proposal (as is customary in transformational generative grammar) the phrase marker must somehow determine the morpho-phonology. Having (200) as a 'deeper' structure (in McCawley's theory, 'deeper' means 'more abstract' than the Deep Structures assumed in the Standard Theory), to which certain rules apply and derive new representations is a staple of *proof-theoretic* models of syntax, thus, our agreement with McCawley's approach can only go so far. What is relevant for our purposes, however, is that (201'a) is a possible variant of (201) but (201'b) is *not*; this depends on structural conditions over the choice of nodes to substitute for NP:

201') a. A boy who saw a girl who knew him kissed her (substitute x_2 in x_1 for NP: x_2; substitute x_2 in Prop for Prn)

---

53 The reader may find it interesting to compare McCawley's 'deeper structure' with that proposed in Altham & Tennant (1975: 55) for MIG sentences:



These authors seem unaware of the similarities between their proposal and McCawley's, but the logical properties of both (the relations they assume and how these relations are satisfied) seem to us to be equivalent.

b. *He kissed a girl who knew a boy who saw her (substitute x_1 in Prop for Prn; substitute x_1 in x_2 for NP: x_1)

The ill-formedness of (201'b) can be blamed on a violation of the so-called *Novelty Condition*, which we referred to above and whose formulation we repeat here:

An anaphorically dependent element cannot have more determinate reference than its antecedent. (Wasow, 1979: 36)

The Novelty Condition can be interpreted as pertaining to the interpretation of visits to nodes in trails: because there is a *strict order* requirement on graphs, in endophoric correference –where, recall, we have more than one visit to the same node in a trail- one visit will be ordered before the other. We have said that a node pronominalises another if they have the same IL translation and there is a specific ordering between them; the Novelty Condition essentially determines the relative ordering between the visit to a specific node that is interpreted as the antecedent and the visit to that same node that is interpreted as being referentially dependant.

To summarise: we share with McCawley a strongly cyclic approach to the structural description of Bach-Peters sentences, but our proposal differs from his in terms of the topological properties of structural descriptions and the necessity to appeal to substitution of indexed nodes. We also share aspects of Wasow's (1979) theory of anaphora, but dispensing with deletion operations (most notably, *equi*, see Wasow, 1979: Chapter 7). Thus, the operation *substitution-by-NP*, which McCawley needs to convert (202) into a *surface structure* without variables, can be eliminated if we allow nodes to be visited more than once in a trail. Moreover, there is no need to stipulate that to each index in a structural description corresponds exactly one NP (McCawley, 1970: 178), because restriction this follows from the trail-approach without further stipulations.

It does not go amiss to note that we do not need to multiply the NPs by distinguishing NPs from indexes because we define NP nodes to be IL translations of NPs. It is in this sense that the proposal made here is compatible with Jacobson's (2000). At the same time, we can capture some aspects of the informal constraints on pronoun-antecedent pairs proposed by Wasow (1979: 61):

Given an NP and a definite pronoun in the same sentence [note: 'sentence', not 'S'], *the NP may serve as the antecedent for the pronoun, unless*:

(a) the pronoun and the NP disagree in person, gender, or number;
(b) the pronoun is to the left of the NP and the pronoun is less deeply embedded than the NP
(c) the pronoun is to the left of the NP, and the NP is indefinite

What we see in (201'b) is precisely a situation like (c). But we can think of a variant with a definite NP just as easily:

201') c. *He kissed a girl who knew the boy who saw her

The argument that we have made here has some interesting consequences for the further study of English sentences displaying a combination of *crossing reference* + *embedding*. In this sense, it is relevant to point out that Karttunen (1971: 157) assumes that the paradox in Bach-Peters sentences arises because the following three assumptions are held simultaneously:

(a) There is a rule of pronominalization that operates on two identical noun phrases.

(b) The rule requires that the noun phrases in question be (i) structurally, (ii) morphemically, and (iii) referentially identical.

(c) Pronominalization is an obligatory cyclic rule

But there is a fourth assumption involved, which Karttunen overlooks (and which underlies the previous three), formulated here as (d):

(d) Pronominalisation (and, more generally, the establishment of referential dependencies) operates over distinct nodes in SMC-complying tree structures

Conditions over licensing in the cases that interest us depend on (a-c) as much as they depend on (d), and we would go as far as saying that (d) is an even deeper and more fundamental assumption for it restricts the class of possible solutions (that is: the SMC is an admissibility condition which applies to whatever 'deep' or 'deeper' structural description we propose for a certain class of sentences). Karttunen objects to McCawley's structural description (in (202) here) by saying that it is not capable of distinguishing between non-synonymous sentences like (194) above (*every pilot who shot at it hit the MIG that chased him*) and the variant in (203) below: the reason is that the same deeper structure would underlie both, and there is no distinct level of semantic representation nor is there a set of semantic interpretation rules. However, the objection is not quite fatal insofar as we can rescue the McCawlean intuition at least in its rejection of an infinite regression at deep structure (see also note *l* in the reprint of McCawley, 1970 in McCawley, 1973: 152-153). Moreover, since by definition our nodes correspond to IL translations, Karttunen's proposal, which incorporates a distinction between *individuals* and *descriptions*, can be incorporated into our theory with minimal adjustments at the level of the interplay between syntax and semantics: in Karttunen's view, the pronouns *him* and *it* in (194) do not refer to individuals (the pilot and the MIG respectively), but to the definite descriptions *every pilot who shot at x* and *the MIG that chased y*. Giving up the SMC and more fundamentally changing the way in which structural descriptions are conceived of does not affect the aspects of Karttunen's argument that we are interested in, which pertain to the nature of referential expressions in Bach-Peters paradoxes. The error in McCawley's conception (as he himself acknowledges) is to treat NPs as referential in the sense of Donnellan (1966); McCawley's revision of (1970) uses Karttunen's definite descriptions as the correct representation of what NPs stand for in deep structure (= semantic structure). We agree with this observation, and as a matter of fact we do not have 'individuals' at all, but *only descriptions*: placeholder variables in graphs do not require *individuals* to be bound by, but rather sub-graphs which correspond to *descriptions* (of individuals or events; see also Schmerling & Krivochen, 2017 for some discussion about the format of these descriptions).

We will now consider some aspects of licensing in Bach-Peters sentences more carefully. When we consider a modified version of (194), in (203) below, we may ask how it is possible to get all relevant connections to hold:

203) [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 (203) is in the object position within a restrictive relative clause (in bold): *Every MIG* [*that chased* ***a pilot who shot at it***]. Importantly, the reading notated in (203) with indexes should not be possible 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, there is *no embedding dependency* between indexes: that is, we do not have *i…j…j…i*, but rather *i…j…i…j*. It is not clear how the base component of a transformational grammar (which is strictly Context-Free) could generate such dependencies.

There is an additional problem which arises in strongly cyclic approaches to structural descriptions: *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 a local domain, in flagrant violation of Relativised Minimality (Rizzi, 1990). *Locality-as-intervenience* in a theory of syntax in which the syntactic component is an autonomous blind combinatoric engine does not have access to properties of the elements it manipulates like their denotation or other semantic properties. At most, it can have access to their category labels (NP, VP, etc.) or aspects of configuration (a given syntactic object can appear in an argumental or non argumental position –A vs. A' in MGG jargon-). Neither helps in this particular case[54].

How can we derive the correct reading and give (201) an adequate structural description? 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 realisations of just two nodes: *MIG*' and *pilot*'. As pointed out above, this avoids the infinite regression problem noted by Bach (1970). 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 the structural descriptions for (193)-(194) 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 (193) as (204a) and that of (194) is represented as (204b):

204) 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 (204a) 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 (193) given in (200), above. The same happens in (204b): *pilot*' and *MIG*' are dominated by nodes in all cycles (i.e., in all sub-graphs). The corresponding ρ-set for (203) is (205), where we have separated the ρ-sets which correspond to different cycles (cf. (204b)):

205) ρ_1 = ⟨(hit, *pilot*'); (hit, *MIG*')⟩
ρ_2 = ⟨(shot, *pilot*'); (shot, *MIG*')⟩
ρ_3 = ⟨(chase, *MIG*'); (chase, *pilot*')⟩

All three cycles are linked at *pilot* and *MIG*, which means that we are not dealing with any *self-contained* unit. The combination of sub-graphs via linking thus allows for all conditions required for licensing to hold, in both the structural descriptions of (194) and (203). We repeat the definition of *licensing* in (125) for the reader's convenience:

---

[54] Friedman et al. (2009), Rizzi (2013) and related work, propose a reformulation of Relatizived Minimality which considers the morphosyntactic composition of syntactic objects alongside their structural position. In addition to the inherent problem posed by the theory of syntactic features in the absence of a meta-theory that restricts what possible features are (see Postal, 1972), this approach has its own weaknesses when it comes to predicting intervention effects (see Villata & Franck, 2016 for some discussion), and still neglects semantics. Partly, this is a problem of granularity (i.e., at which level of 'syntactic organisation' is semantic interpretation determined?), and partly it is an architectural problem (syntax is still the only generative component, autonomous and severed from semantics, which is –still without proper justification- interpretative).

206) <u>*Licensing (final formulation)*</u>

Let G and G' be sub-graphs and v_i and v_j be nodes. Then, $v_i \in G$ may license $v_j \in G'$ iff

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

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

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

i. $\rho^*(v_j, v_i)$ holds for $v_j \in G'$ and $G' \neq G$, ***and***
ii. v_i receives a grammatical function in G

The graph-approach can provide us with an exhaustive description of relations between nodes while maintaining the advantages of a strongly cyclic approach to syntax. Specifically, we see that the cycles that form the structural description of (203) can be linked at the relevant nodes, giving us the desired lecture, without violating any of the conditions for licensing or linking.

In connection to 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 (following Ross' ground-breaking 1967 work). 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, 2016a) in terms of possible relations between elements belonging to separate sub-graphs.

11. <u>Two kinds of 'and' and the Coordinate Structure Constraint</u>

We may begin this section by recalling Fillmore's (1936) distinction between two kinds of generalised transformations: *embedding* transformations and *conjoining* transformations. The strings thereby generated involve different dependencies, and thus receive distinct structural analyses; we therefore must be careful in order to assign the appropriate structural description to natural language sentences obtained by conjoining and embedding. So far, we have been analysing 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, 2018) we argued that *true coordination*[55] is not a unified phenomenon syntactically or

[55] 'True' coordination is to be distinguished from what Krivochen & Schmerling (2016b) call 'mirage coordination' (see also De Vos, 2005; Biberauer & Vikner, 2017; Bravo, 2020, among others, who use the term 'pseudo-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; they also display strong restrictions which do not apply to garden-variety true-coordinations (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 semantically. 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 summarise the main characteristics of each:

Que-coordination:

- No internal structure that can be probed by an external element: *que-coordinated* outputs are opaque for all syntactic intents and purposes
- 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 (*infective*), 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)

208) a. The sudden rise and the equally sudden fall of the stock market have economists worried.
b. The sudden rise and equally sudden fall of the stock market has economists worried.

209) 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 (208b) and the Spanish example (209b) 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 (208a) and (209a), exhibiting *et-coordination*, are not semantically equivalent to their *que-coordinated* counterparts. Consider the *et-coordinated* example (208a): 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

'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*.

about its fall and not its rise. This interpretation is not possible for the *que-coordinated* example (208b), where economists must be worried about the combination of these phenomena.

It is important to observe that the distinction between *et*- and *que-coordination* does *not* overlap with that between *symmetric* and *asymmetric* coordination (see Schmerling, 1975 for a Gricean discussion about this latter difference). Distinguishing between *symmetric* and *asymmetric* true coordination requires us to have access to the terms of the coordination; this therefore means that the distinction between symmetric and asymmetric conjunction falls entirely within the realm of *et-coordination*. *Que-coordination* is internally opaque, thus its terms cannot be either symmetric or asymmetric. A feature of *symmetric* coordination is the possibility of inverting the terms of the coordination *salva veritate*, as in (210) (210 c-d are taken from Schmerling, 1975):

210) a. John had a beer and Mary had wine
b. Mary had wine and John had a beer
c. Rome is the capital of Italy and Paris is the capital of France
d. France is the capital of France and Rome is the capital of Italy

The conditions under which (210a) is true are the same as those under which (210b) is true; there is no implicature of temporal order between the events of John having a beer and Mary having wine. The same happens in (210 c-d). We can have the terms of the coordination in any order and the compositional interpretation is exactly the same: there are two events, both accessible and independent of each other: the dependency between these terms (which remain distinct syntactic and semantically) is *paratactic* (cf. Camacho, 2003, who argues in favour of a uniformly *asymmetric* phrase marker for coordination, ensuring generalised hypotaxis; such a view is generally endorsed by proponents of phrase structure grammars; see also Kayne, 1994, 2018; Progovac, 1998). This is not possible with *asymmetric* coordination, since there is an ordering between the terms of the coordination (often implicating temporal ordering between events denoted by those terms):

211) a. Mary went out with Peter and broke John's heart
b. Mary broke John's heart and went out with Peter

Note that changing the order of the terms changes the meaning of the sentence, and the conditions under which each is true (or, at the very least, felicitous). For (211a) to be an adequate characterisation of a sequence of events, it must be the case that John's heart was broken *after* (and most likely *because*) Mary went out with Peter. But that is not at all the case in (211b), where Mary *first* broke John's heart and *then* went out with Peter (there seems to be no cause-consequence relation between the events in this latter reading). In both cases, the terms of the coordination are accessible, such that there are *two* events, not a complex one; however, these events are not independent of each other. There is a *hypotactic* dependency between the terms of the coordination, which requires them to be separate entities such that they can be ordered.

Que-coordination is neither paratactic nor hypotactic, in the sense that the output of *que-coordination* is a single complex entity (sortal or eventive) which we cannot internally manipulate with syntactic rules or semantic interpretation principles: these must apply to the *whole* term as a computational atom. Neither *symmetric* nor *asymmetric* coordination can be outputs of *que-coordination*, in the light of the brief discussion about (210) and (211).

In past works (Krivochen & Schmerling, 2016a; Krivochen, 2018) we proposed that the difference between *et*- and *que-coordination* must be captured in the structural descriptions assigned to these sentences, which in turn will impact on the operations that can apply to them because the input for some transformational rule may not be met anymore in some given case. Let us be more specific. In this context, we argued that the Coordinate Structure Constraint (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. Summarising much discussion from previous work, the reason is relatively clear: the CSC is formulated as a condition over chopping transformational rules[56], 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 central place in the theory of the grammar cannot be underestimated: it is possibly the most problem-free principle ever formulated in generative grammar (Postal, 1998: 52), and its empirical applicability goes well beyond English (and even beyond Indo-European languages; see 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 insights of the CSC as a condition over well-formed expressions of the language. We must be able to filter out the graphs that would be assigned to expressions which violate the CSC.

As for the first challenge, it is useful to take a look at what other graph theory-based frameworks have proposed. APG, for instance, recognises 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 *coordinandos*, also known as the 'law of coordination of likes'). The relevant definitions, pertaining to heads and tails of *Con* arcs, are the following (Johnson & Postal, 1980: 209):

212) a. Coordinate(a) ↔ (∃A) (Con arc(A) ∧ Tail (a, A))
b. Conjunctive(a) ↔ (∃A) (Con arc(A) ∧ Head (a, A))

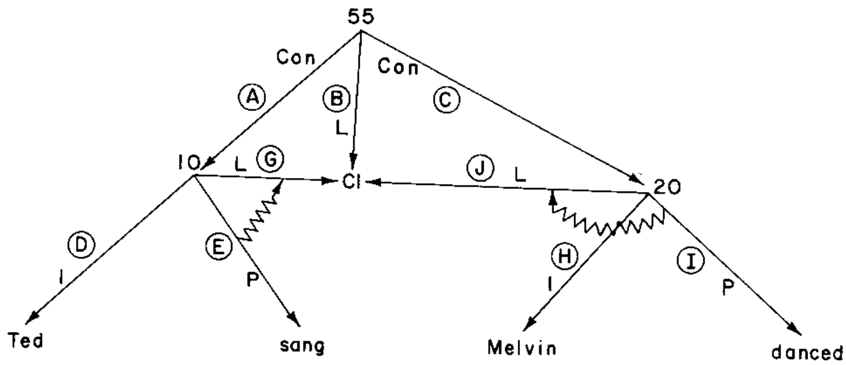
For example, in the asymmetric example *The fiend shot and knifed its victim* (example taken from Johnson & Postal, 1980: 222), the coordinated Vs are the heads of Con arcs. If we consider a

[56] 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)

symmetric coordination, like *Ted sang and Melvin danced*, the arc pair description looks like this (from Johnson & Postal, 1980: 207):

213)



In and of itself, we think, the APG treatment of coordination does not capture the difference between *et*- and *que-coordination* in its structural descriptions, which is empirically motivated at its very core. It is also unclear how (if in any way) APG could represent the difference between *symmetric* and *asymmetric* coordination; note that both V arcs E and I sponsor arcs G and L. Johnson & Postal offer no arc description of an asymmetric coordination that we can contrast (213) with. This does not mean that APG or Metagraph grammar could not be modified in order to capture these distinctions, but how to do it is not evident based on available analyses[57].

Let us continue the survey, now dealing with Dependency Grammars. 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 (i.e., an *asymmetry* between coordinated terms). 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 (2018) as part of an argument in favor of *mixed computation* (the idea that the grammar oscillates up and down the Chomsky Hierarchy when assigning adequate structural descriptions to local sub-strings):

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

Note that –as we saw in (210)– (214') below, where the order of the conjuncts has been reversed, 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):

214') Bill got the cookies and John bought the milk

[57] We must highlight that Johnson & Postal (1980: 207) themselves acknowledge that there are many questions raised by coordinate structures which they '*have not had the chance to study*', so it is possible that further research would have made it possible to distinguish between *symmetric* and *asymmetric* coordinations in terms of distinct arc pair descriptions.

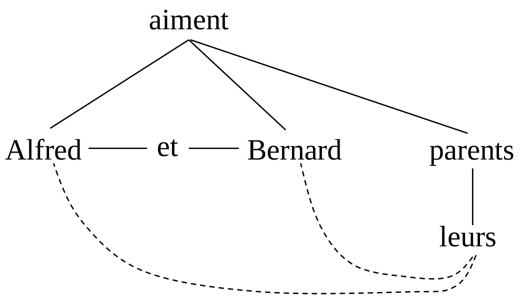
The reversibility of conjuncts -we argued in Krivochen (2018)- points towards a paratactic structural description (essentially, with both S terms being states in a Markov chain); this is consistent with the lack of order in the interpretation of events since there is no hierarchical structure. But let's go one step further, and consider (212) embedded in a bigger structure:

215) 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 (the first of which is itself a coordination), 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. Besides the fact that symmetric coordination allows for the reversibility of conjuncts *salva veritate*, we need to point out that the grammar also allows for a *combination* of hypotactic and paratactic coordinations in the same sentence: if all instances of true coordination were assigned the same structural description (a headed one, in Mel'čuk's view and MGG), among other issues, we would be faced with the problem of getting the meaning right in each specific case since we would get no aid from the phrase marker or dependency diagram. Ultimately, uniform approaches to coordination as headed structures (be these dependency based or phrase structure based)

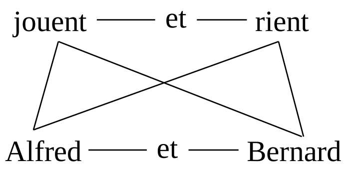
A view different from Mel'čuk's, but still within DG, is expressed by Tesnière (1959: 80, ff.). Tesnière assigns coordinating conjunctions to a category *junctives*, which are *functional elements* (or *empty words*, 'mots vides'). Junctives are structurally between the terms they conjoin (called *nuclei*), but remain outside these terms. In brief, coordinated structures are not only not headed structures, but they are not even dependency-based. His diagrams, which he calls *stemmas*, are similar to ours insofar as they can violate the SMC if necessary: any term can have more than one *governor*. However, there is no requirement of *strict ordering* in his *stemmas*, unlike our graphs. A simple coordinated structure like *Alfred et Bernard aiment leurs parents* ('Alfred and Bernard love their parents') is analyzed thus:

216)



Note that, because the sentence means that Alfred loves his parents and Bernard loves his parents (but neither Alfred nor Bernard love each other's parents), there is a dependency between the possessive *leurs* and both *Alfred* and *Bernard*; it is possible to dissociate the reference of the possessive NP without multiplying entities or indices. We will not provide a full review of the kinds of coordinated structures considered by Tesnière, which constitutes a remarkably rich landscape, but we do want to point out that his view on coordination is particularly interesting insofar as he deals with 'crossings' in coordinated structure free from the prejudices of Phrase Structure Grammars and diagrams of L-trees. For instance, the *stemma* he assigns to a sentence featuring a coordinated subject and a coordinated predicate phrase, like *Alfred et Bernard jouent et rient* ('Alfred and Bernard play and laugh') is the following (taken from Tesnière, 1959: *Stemma* 261):

217)



It is important to note that there is nothing in Tesnière's *stemmas* that specifies the linear order of the sequence for which the *stemma* provides a structural specification. In that sense, our perspective is close to his. However, there is no overt identification of grammatical functions *in the stemmas*; in a sentence like 'Alfred gives the book to Charles', *Alfred*, *the book*, and *Charles* are all equally *actants*, despite the fact that *Charles* and *the book* strictly speaking do not 'act' (Tesnière, 1959: Chapitre 48). The distinction between Subjects and Objects is, according to Tesnière, a 'remnant' from the Port Royal grammarians. His argument is rather strongly worded:

> *§5 Indeed, all arguments that can be invoked against the concept of the verbal node and in favor of the opposition between subject and predicate come a priori from formal logic, which has nothing to do with linguistics.*
>
> *§6 Concerning strictly linguistic observations about the facts of language, the conclusions drawn a posteriori are of a much different nature. There is no purely linguistic fact in any language that suggests the existence of the subject-predicate opposition.* (Tesnière, 1959: Chapitre 49. Translation by Timothy Osborne & Sylvain Kahane)

It should go without saying that we disagree with this view, given the prominent role that grammatical functions play in our proposal. Furthermore, there are well-documented asymmetries between subjects and non-subjects in English and Spanish (the languages we have drawn almost all of our examples from) in terms of extraction: filler-gap dependencies targeting objects are fine, whereas extraction from subjects is a rather problematic issue[58]. Within MGG, several conditions have been proposed to account for the relevant facts: the *Condition on Extraction Domains* (Huang, 1982), the *Subject Condition* (Chomsky, 1973), the *Left Branch Constraint* (Ross, 1967), the *Sentential Subject Constraint* (Ross, 1967), among others. While the universality of these conditions has not gone uncontested (e.g., Mayr, 2007 shows a lack of subject-object asymmetries for purposes of extraction in Bavarian), they do capture (with varying degrees of success) English phenomena. It seems to us that there is enough syntactic evidence for subject-object asymmetries to maintain the central role of grammatical functions in the theory of the grammar in order to provide fully explicit connectivity maps for English expressions.

After APG and DG, we have MGG left to review. During the Standard Theory days, it was the norm to see *n*-ary branching representations of coordinated structures, and as a matter of fact, unbounded coordination was one of the phenomena usually invoked in the justification of transformations, *singulary* or *generalised* (Lees, 1976: 33, ff.; Chomsky, 1957: 37-38 respectively). However, as we briefly saw in **Section 4**, during the GB-MP days MGG attempted to make

[58] Subject-object asymmetries have been object (no pun intended) of inquiry in psycholinguistic studies, dealing with acquisition of asymmetries and processing of extraction from both positions. We will not refer to these studies because the interest of the present work is the grammar as a formal system, without the requirement that it be implemented psychologically. 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 Categorial Grammar (see Dowty, 1978: fns. 2 and 3); see also the epigraph of this very work.

coordinated structures fit the X-bar scheme: binary-branching, headed, projecting syntactic objects. The result is that coordinated structures always display hypotaxis, with one term of the coordination c-commanding the other as specifier and complement of a phrase whose head is the coordinating conjunction (e.g., an &P or CoordP).

We argue, against MGG, that coordination is *never* an endocentric structure &P, CoordP, or anything like that: we agree with Tesnière (1959) there is no 'head' in coordinated structures in X-bar theoretic terms. Thus, there cannot be a *dependency* in all cases (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 true *et-coordination* and *que-coordination* yields a perfective (internally opaque) entity; this simply cannot be captured with an *a priori* uniform approach to coordination. That is: to a string *X and Y*, where X and Y are variables over strings, there can correspond three structural descriptions, roughly along the lines of (218):

218) a. X → (and, Y)
b. and → (X, Y)
c. Z

(218a) corresponds to the *asymmetric et-coordinated* case, in which the root of the sub-graph X dominates the root of the sub-graph Y. Therein lies the essence of *hypotaxis* in the present view. (218b) corresponds to *symmetric et-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* (the structural relation between them is strictly paratactic) and the graph is extended by having a *new root*; in the former case, the second conjunct's root is within the ρ-domain of the first conjunct's root, which means that the graph is not extended at the root, but at the frontier as one conjunct is embedded into the other. (218c) features a *single object* corresponding to a string *X and Y*, distinct from both X and Y, within which we cannot probe: this corresponds to the *que-coordinated* case. This is a way (perhaps not the best, certainly not the only one) to represent the internal accessibility of *et-coordination* and the opacity of *que-coordinated* terms.

Let us now consider the ρ-sets of the two conjuncts in (219) below, an instance of *asymmetric et-coordination*:

219) John had a beer and fell asleep
220) ρ_1 = ⟨(have, *John*'); (have, *beer'*); (have, and)⟩
ρ_2 = ⟨(and, fall); (fall, *John*'); (fall, asleep); (asleep, *John*')⟩

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 *symmetric et-coordinated* case is different, as we anticipated. We would like to propose that in this case 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*. However, because there is no dominance relation between the terms of the coordination, we predict that any linear ordering of these terms should be not only grammatical, but also preserve all grammatical relations and semantic representation (at least those aspects of semantics which can be accounted for configurationally; see Schmerling, 2018b for discussion).

Now, we can analyze a *symmetric et-coordination*. Recall that one of the features of symmetric coordination was the reversibility of conjuncts, which we saw in (214) and (214'), repeated here as (221a-b):

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

Our argument is that (221a) and (221b) are expressions whose structures display the same connectivity patterns, therefore, there is a single graph (or, equivalently, two isomorphic graphs) that describe(s) them. What we have in (221) is, we argue, something along the lines of (222):

222) ρ = ⟨(and, get); (and, buy); (get, *Bill*'); (get, *cookies*'); (buy, *John*'); (buy, *milk*')⟩

We have two lexical verbs, *buy* and *get*, each of which is the root of its own cycle. These cycles contain each verb's nominal dependants: *Bill* and *cookies* for *get*, and *John* and *milk* for *buy* (plus the temporal modifiers of these verbs, which we omit for convenience). Note that, in contrast to (220), here neither cycle dominates the other: they are *paratactically related*.

The kinds of coordination we have identified are, in principle, not limited to specific categories: we can *et*- or *que-coordinate* any category that can be coordinated. For example, let us see what happens when we *que-coordinate* Ns or NPs:

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

The *que-coordinated* reading of (223a) depends on *red beans and rice* being a single dish (there is, thus, some potential for variation depending on the reader's culinary background; this particular example was suggested to us by Susan Schmerling, p.c., as a traditional dish in Southern US). Likewise, the *que-coordinated* reading of 223b) depends on *the rise and fall of the stock market* referring to the oscillation of the stock market, where *the rise and fall* denote a single process.

But we can also have *et-coordinated* Ns. The following are examples of *symmetric et-coordination* of Ns:

224) a. John drinks beer and whisky
b. John drinks whisky and beer

Obviously, John is drinking *two* things, not one; a *que-coordinated* reading is not an option in this case. But we can give some more detail about the conditions under which In Krivochen & Schmerling (2016a) 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. With respect to this, note the following paradigm, where *both* and *either* necessarily require accessibility of two distinct terms in a true coordinated structure:

225) 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

If the distinction between *et*- and *que-coordination* impacts on the structural description assigned to expressions of the language, as we have argued here and in past works, then it should determine whether specific syntactic operations can or cannot apply. In classical terms, this is so because different structural descriptions may or may not satisfy the input conditions for specific transformations. Here, we are concerned with what initially could be thought of as a *second-order* condition, a restriction over rules (Pullum & Scholz, 2001): Ross' (1967) *Coordinate Structure Constraint*, which we repeat here:

> *In a coordinate structure, no conjunct may be chopped, nor may any element contained in a conjunct be chopped out of that conjunct.* (Ross, 1967: 428)

Note that this formulation of the CSC restrict the *rules* that can apply to a coordinate structure. However, it can be rephrased as a well-formedness condition over graphs, in which case we could keep the empirical coverage of the CSC without the issues that come with a *transformational proof-theoretic* framework. We need to identify which expressions display the kind of dependencies that we want the CSC to restrict.

Krivochen & Schmerling (2016a); Krivochen (2018) defend the idea that the internal opacity of *que-coordination* makes it impossible for any of its terms to be targeted by extraction operations: nothing can be moved outside of a *que-coordinated* structure because its internal structure is not accessible; it is an atomic object. Moreover, if the terms of the *que-coordination* are indeed interpreted as a single, internally unanalyzable unit, there is no rescuing a reordered structure with ATB (which, it is worth mentioning again, would require access to the internal structure of each conjunct). As such, the CSC applies to *et-coordinated* structures, not *que-coordinates* ones: in the former, the conjuncts are kept distinct, and thus remain in principle accessible, both syntactically and semantically. How can we formulate a relevant admissibility condition that captures the descriptive power of the CSC? The condition proposed in Krivochen (2018: 257-258) could prove useful in this respect:

226) In an *et*-coordinated structure of the form [$_K$...SO_i [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; also Goodall, 1984 for a slightly different take),
- K and L are terms, and
- $L \subset K$;

a mapped phrase marker [$SO_{(i/j)}$...[$_K$...SO_i [and [$_L$...SO_j]]]] is legitimate *iff* $i = j$

There are some things to note here. Condition (226) pertains to *asymmetric et-coordination*, and addresses the kind of data considered in Na and Huck (1992) and which led them to formulate their 'Condition on Asymmetric Conjunction':

> *Condition on Asymmetric Conjunction (CAC): In any asymmetrical conjunction, if extraction is performed on a secondary conjunct, it must be performed across-the-board.* (Na and Huck, 1992: 125)

Condition (226) straightforwardly yields the well-formedness of ATB constructions (Ross, 1967; Williams, 1978) in which a rule applies to the same element across all coordination terms, but it also predicts that canonical examples of the applicability of the CSC are indeed ill-formed expressions of the language. The requirement that $i = j$ in (226) pertained to indexed syntactic objects in the proof-theoretic framework adopted in Krivochen (2018); it can readily be replaced by the condition that there be a node v_i that is visited three times in a trail: one in a scope position, one in K, one in L. Note, incidentally, that this formulation makes *no reference* to situations of the following kind,

227) [SO$_i$...[$_K$...SO$_i$ [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: 3) –traces and indexes added for expository purposes only-

228) [[$_K$ [How many lakes]$_i$ can [$_K$ we destroy t_i]] and [$_L$ not arouse public antipathy]]?

It is crucial to point out that the nature of the term L is not relevant: Schmerling's (1983b: 17-18) argument that [and not] is actually a basic expression, and thus the possibility that [not arouse…] is not itself a constituent, makes no difference for the purposes of (226) applied to (228). 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 empirically motivated 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-α*; *Merge*, etc.) which greatly obscure grammatical description (and do not provide *explanations* in any meaningful sense; see Postal, 2004: §§ 9, 12, 13 for detailed discussion). What follows is a classification of well-known and time-tested 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 summarises 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 5.1**)

c) Raising-to-Object (**Section 5.2**)

d) Topicalisation (**Section 5**)

e) Focalisation (**Section 5.5**)

f) NPI fronting (**Section 5.5**)

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

a) RNR (see McCawley, 1988)

b) Wrap (as in Bach, 1979; **Section 3**)

c) Dative Shift (*contra* Dowty, 1979)

d) Right dislocation, including Heavy NP Shift (**Section 13.1**)

e) Location fronting

f) Though- / as- preposing (see Ross, 2012)

g) Parenthetical insertion (see McCawley, 1982 and **Sections 6 and 7**)

h) Clitic climbing (**Section 6.1**)

i) Pseudocleft formation (*as per* Ross, 2011)

j) Classical NEG raising (in the interpretation where raised and non-raised NEG yield equivalent interpretations; see Collins & Postal, 2014 for extensive discussion)

II. Relation-changing transformations:

a) Passivisation

Being one of the very few, if not the only, relation-changing transformation that we have identified in the grammar of English, *passivisation* requires some further discussion. We follow Perlmutter & Postal's (1983) characterisation of the *passive* in terms of grammatical relation changing, regardless of how that change is attained (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 *passivisation* 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* (and what was a 1 becomes a chômeur). 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 passivisation (in English) is an entirely relational process (Perlmutter & Postal, 1983; Postal, 1986). This means, importantly, that accounts of passivisation in terms of *word order, verbal morphology,* or *case* cannot form the basis of an adequate universal characterisation of the process (Perlmutter & Postal, 1983: §3). In English, there are at least two interesting examples of constructions that do not fit the mould of traditional accounts of the passive in terms of verb typology ('only transitive verbs can be passivised') or morphology ('subjects of passive clauses cannot be non-nominative'[59]): namely, passives of intransitive verbs (Postal, 1986: 30-32) and non-nominative subjects in passive clauses. Interestingly, the former cases arise when the 2-arc leaves a preposition stranding, a factor that we will not analyse here (but see Postal, 1986: Chapter 6 for discussion). The latter present a different quirk, since they are dialectally restricted. Relevant examples are like (229) and (230) –below-, respectively:

229) a. No one has ever sat on this chair
b. This chair has never been sat on (by anyone) → *sit* is an unaccusative V

The subject in (229b) is morphologically nominative, as are all subjects in what Postal (1986) names 'pseudo-passives'. We can probe this by pronominalising the subject: *They/*them have never been sat on*.

As for non-nominative subjects in passives, we can observe (230):

230) a. The Texan police always target them Latinos
b. $\begin{Bmatrix} Them\ Latinos \\ *\ They\ latinos \end{Bmatrix}$ are always targeted by the Texan police

Regardless of the arguments that can be made for *them Latinos* in (230b) being assigned some kind of abstract Nominative case (which would be mostly intra-theoretical), the point is that a view of the passive that relies on verb typology or morphology faces difficulties in the description of English facts that the RG approach easily avoids.

Transformational generative grammar, from its very early days (e.g., Harris, 1957: rule 4.11; Chomsky, 1955: 504f-g), derives a passive sentence via the application of a mapping rule to an active sentence. In the transformational view, an NP is base-generated in the object position (sister to V) at the base component, Deep Structure, or whatever level of representation or intermediate derivational step is there before the application of transformational rules, and then moved to Spec-Infl / Spec-T / Spec-AgrS (depending on the framework of choice) leaving a trace or copy behind in order to satisfy the requirement that clauses have subjects (the so-called *Extended Projection Principle*; Chomsky, 1982: 9-10). 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. It may be noted that LFG was heavily inspired by Emonds' (1970) *Structure Preservation* hypothesis, and that *passivisation* was indeed identified as a *Structure Preserving transformation*: the D-Structure NP object can move to subject position because there is an independently motivated phrase structure rule that generates the configuration [$_S$ NP VP].

[59] Importantly, here we are referring only to Nominative-Accusative languages. See Dixon (1972) and Schmerling (1979) for an analysis of Dyirbal (an Ergative-Absolutive language) which suggests that a characterisation of the Dyirbal passive strictly in RG terms may be inappropriate.

We want to emphasise that the framework we have been describing applies straightforwardly to passive clauses, with no additional stipulations (rules, levels of representation, indices) needed. Consider the following pair:

231) a. Bill killed John
b. John was killed

The ρ-sets for (231a) and (231b) are given as (232a) and (232b) below:

232) a. ρ = ⟨(kill, *Bill*'); (kill, *John*')⟩
b. ρ = ⟨(be killed, *John*')⟩

(232b) features a 'detransitivised' expression *be killed*, the GF *object* cannot be assigned in this structure. And this, following RG, is the fundamental property of the *passive*: the change of grammatical relations, which are atomic and primitive in this theory and its successors (within which we can count ourselves). Because in transformational grammar GFs are defined configurationally, a transformational account of *passivisation* says nothing about what we think is the most important property of this process: it is a Relation Changing Rule, perhaps the only one in English.

We assume two fundamental properties of passive *be* in English: (i) it has no argument structure; and (ii) it forms a derived expression with the subject of the lexical verb (see Schmerling, 1983b: 26-27). Passive *ser* is syncategorematic, therefore, it does not correspond to an independent node in the graph. We can note, on the issue of CG's flexibility (which our graphs capture), that only property (i) 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:

233) a. Mary wasn't killed, John was ~~VP~~ (= killed)
b. *María no fue asesinada, Juan fue ~~VP~~ (= asesinado)

We propose that it is the existence of a relation (be killed, *John*') that licenses examples like (233a), and presumably that relation does *not* hold in Spanish: *ser* is much more of an inflectional element (a bound morpheme) than *be* is in English, which would translate into the claim that passive *ser* actually configures a two-word expression with the participle: the ρ-set for something like *María fue asesinada* would be simply ρ = ⟨(ser asesinada, *María*')⟩, *not* ρ = ⟨(ser, *María*'); (ser, asesinada); (asesinada, *María*')⟩. There is no relation between *ser* and *María*, which readily blocks the possibility of having an expression *María fue* in so-called VP ellipsis contexts (cf. (233b), *mutatis mutandis*): passive *ser* is *not* a basic expression in Spanish, nor does it form a derived expression with the subject. Rather, in Spanish as opposed to English, it is to be grouped with the predicative elements, what in IC analyses would be called the VP.

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

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

In (234), 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 assigned to the category of expressions that must combine with an expression of category FC/NP to form an expression of category FC/NP: *ser* ∈ (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*. This is a testable prediction. A model of Spanish passives in which *ser* does not form a basic expression with the participle it selects could generate the ungrammatical (235):

235) *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*, (235) 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 in a CG can be multi-word (an idea that, as we have pointed out, has an illustrious pedigree, including Jespersen, 1985 [1937]; see also Culicover et al., 2017 for a constructional perspective), and allowing nodes in our graphs to correspond to (IL translations of) basic expressions of the language rather than to lexical items, we can capture an empirical difference which cannot be straightforwardly accounted for (or even described) in structurally uniform, *a priori*-based PSGs where terminal nodes are most often than not equated to orthographic 'words' (or, in some recent developments, smaller units like morphemes or features –e.g., Embick & Noyer, 2007; Baunaz & Lander, 2018 respectively-, but never *supra-lexical* units).

The rather surprising (well, that depends on the reader) provisional conclusion that we tried to illustrate in this brief section is that most 'transformations' identified in the generative tradition (see Ross, 2012 for a very complete overview) either create new relations *while preserving existing ones* or only change linear order, thus not impacting on node connectivity at all. In other words, an adequate grammar for English should focus on modelling RPTs rather than RCTs. 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 note on leftward and rightward 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, Right Node Raising- is heavily constrained (only changing linear order, but not constituency). Now, what is not clear is how to encode the specific filters (e.g., the 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 structured around a single lexical predicate. This view shares some aspects with TAG analyses of *Extraposition* (e.g. Kroch & Joshi, 1987), notably the way in which rightwards movement is bounded, although the format of structural descriptions changes dramatically (since elementary trees in TAGs are context-free phrase structure trees).

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, 2016a, 2018) and Krivochen & Schmerling (2016a, b), according to which coordinated structures can be hypotactic and phrase-structural (*et-coordination*) or paratactic and finite-state (*que-coordination*). Because of its relevance in the arguments in favour of discontinuous structure and multidominance, we will focus on Right Node Raising (RNR) among the processes mentioned above. As pointed out in Postal (1998: 97) RNR is a *phenomenon*, not a rule. In RNR structures, there is a constituent that is shared between coordinated terms and which surfaces in a position to the right of all these terms in an ATB manner: the rightwards constituent needs to receive an interpretation in all coordinated terms, and all gaps must be coindexed, as we can see in the examples (236) (from Postal, 1998: 97; De Vos and Vicente, 2005: 98):

236) a. Ernest suspected e_i, Louise believed e_i, and Michael proved e_i [that she was guilty]$_i$.
b. I know a man who loves e_i and a woman who hates e_i London$_i$

Grosz (2015) distinguishes three kinds of syntactic approaches to RNR: (i) backwards deletion, (ii) rightwards ATB movement, and (iii) multidominance. Baltin (2006) includes a fourth alternative, originally suggested in Kayne (1994: 78, ff.), according to which there is leftwards movement of the object modified by the extraposed term (so as to have the phrase marker comply with the apriorisms of antisymmetry; more generally, Kayne formulates a ban on right-adjunction). Deciding between those is not a straightforward matter, and we will not spend much time arguing *against* some account or another, but rather *in favour* of the kind of approach we defend.

First and foremost, we need to appropriately characterise the configurations that license RNR. Something that all accounts agree on is that in order to have RNR the terms of coordination need to be probed into, be it by a deletion rule, a movement rule, or the establishment of a direct dependency. This means that RNR arises in *et-coordinated* structures, symmetric or asymmetric (the examples in (236), for instance, are both *symmetric*). The cyclic analysis and ρ-set for a simple case like (237) (arguably, an *asymmetric et-coordination*) is given in (238):

237) Bill bought, and John washed, those china dishes
238) Cycle 1: [Bill bought *dish*']
Cycle 2: [John washed *dish*']
ρ = ⟨(buy, *Bill'*), (buy, *dish*'), (wash, *John'*) (wash, *dish*')⟩

A question that arises at this point is how RNR relates to other grammatical phenomena in English. In this respect, Postal (1998: Chapter 4 and Appendix B) argues that RNR is analogous to L-extractions (*wh*-movement, topicalisation, etc.); this approach predicts that RNR should not be clause-bound because L-extractions are not (think of long-distance *wh-movement*). 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 'successive cyclic' RNR is only to be expected. For example:

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

There are indeed some common constraints between L-extractions and RNR. An important difference is their sensitivity to island constraints. De Vos and Vicente (2005) point out that RNR, unlike L-extractions, do not display CNPC islandhood effects: compare (236b) above with (240):

240) *Which city$_i$ do you know a man who loves e_i and a woman who hates e_i?

However, it is incorrect to say that RNR is 'island insensitive' without further qualifications (De Vos & Vicente, 2005: 98): McCawley (1982: 101) and Postal (1998: 121, ff.) observe that the CSC applies to RNR cases just as much as it applies to L-extraction ((241a) and (241b) are taken from McCawley, 1982: 101):

241) a. Tom is writing an article on Aristotle and Freud, and Elaine has just published a monograph on Mesmer and Freud
b. *Tom is writing an article on Aristotle e_i, and Elaine has just published a monograph on Mesmer e_i [and Freud]$_i$
c. *Who$_i$ is Tom writing an article on Aristotle (and) e_i and Elaine has just published a monograph on Mesmer (and) e_i?

This fact must be taken carefully, for in and of itself it does not provide unambiguous support for a movement, deletion, or multidominance analysis. As far as we are concerned, Postal is right in saying that

the CSC is, of course, a fundamental condition on L-extractions (Postal, 1998: 121)

But this does not solve the problem of *what* extraction actually is: if our approach to coordinated structures in **Section 11** is on the right track, then CSC effects can be captured in a no-extraction, model-theoretic syntax as much as in a proof-theoretic model or even a model-theoretic approach with sponsor-deletion arcs.

McCawley's 'order-changing' approach to RNR can be traced back to Wexler & Culicover's (1980: 301) (also cited in Postal, 1998: 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*

In our terms, this approach translates as the raised node being dominated by a node in each of the coordinated terms: grammatical relations are, in this way, preserved under RNR. It is important to note that RNR does not seem to make the targeted object internally opaque: we are referring here to what Postal calls *interwoven dependencies* and De Vos & Vicente refer to as *CoRNR* ('Coordinated structures under RNR'). Let us illustrate the relevant structures:

242) a. I didn't say that John had talked and that Peter had replied in a loud voice and with a whisper (De Vos & Vicente, 2005: 98)
b. John loves and Mary hates oysters and clams, respectively (Postal, 1998: 134)

Postal correctly observes that L-extractions allow for interwoven dependencies (if no other constraint is violated[60]):

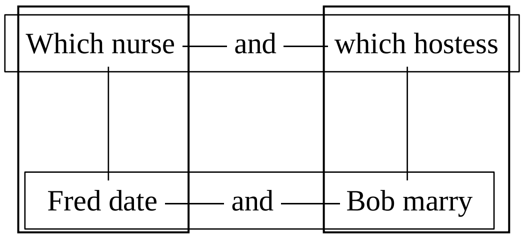
---

[60] This is an important caveat: some examples presented by De Vos & Vicente (2005) are ungrammatical not because of some problem with coordination itself, but because of an unrelated violation. For example, their (5c) (reproduced here as (i)) can be taken to illustrate a *negative island* (given the fact that the extraction sites are c-commanded by NEG; Ross, 1984; Rooryck, 1992; Abrusán, 2014; in Abrusán's terms, the condition that there

243) [[Which nurse]$_1$ and [which hostess]$_2$]$_3$ did Fred date t_1 and Bob marry t_2, respectively? (taken from Postal, 1998: 134, ex. (109b). Indices and traces are his)

In principle, this means that the relevant configurations do not involve *self-contained* objects in either case. The CSC is *not* violated in (243) because whatever rule we assume yields *wh*-fronting does apply across the board: as soon as one of the object NPs stays *in situ*, the structure becomes ungrammatical for all the usual reasons (e.g., *Which nurse did Fred date and Bob marry which hostess?). It is important to note that the S coordination in (243) is a case of *symmetric et-coordination*; neither term *Fred date NP* and *Bob marry NP* is embedded into the other. This is crucial if the well-formedness condition in (224) is on the right track: we are not dealing with a *hypotactic* dependency, therefore, the CSC does *not* apply. It may be useful for the reader if we used a *stemma*-like diagram (cf. (217)) to see what is going on in (243):

244)



(244) visually illustrates that there is *no relation* between *Bob marry* and *which nurse*, or between *Fred date* and *which hostess*. We have two cycles, given the fact that there are two lexical verbs with their respective nominal dependants, and there are no crossing dependencies from one cycle probing into the other. And, we have two symmetric coordinations, *Fred date and Bob marry* and *Which nurse and which hostess*. That means that, in total, we can identify four syntactic objects, to make things simpler (which we have boxed in (244) above). Let us provide the ρ-sets for them (we will use numerical subindices to distinguish between *which* in *which nurse* and *which hostess* for expository purposes only):

245) Object 1 = ⟨(date, *Fred*'); (date, which$_1$); (which$_1$, nurse)⟩
Object 2 = ⟨(marry, *Bob*'); (marry, which$_2$); (which$_2$, hostess)⟩
Object 3 = ⟨(and, which$_1$); (and, which$_2$)⟩
Object 4 = ⟨(and, date); (and, marry)⟩

The sets of dominance relations in (245) capture what in our opinion are the crucial features of (243): first, that there are no *crossing* relations (cf. the examples analysed in **Section 4**); second, that we are dealing with *symmetric* coordination. Indeed, as Postal claims, RNR behaves in the same way, *ceteris paribus*:

246) Fred dated and Bob married a lovely nurse and a gentle hostess, respectively

Both RNR and L-extractions are in principle unbounded processes in English. This means, in this context, that as long as the structural conditions specified above (ATB rule application in *et-coordinated* structures) hold, we can make the structure grow monotonically without violating any syntactic condition (although, needless to say, things get harder to parse). For example, we can have

be a single most informative answer is not met) rather than L-extraction being incompatible with interwoven dependencies:

i) *[How] didn't you say [[that John had talked *e*] and [that Peter had replied *e*]]?

the following rather extreme example of RNR which is by no means reader-friendly, but it is perfectly grammatical:

247) Frank didn't admit to t_1 that he could t_2 nor deny to t_1 that he should t_2 [hire (t_1), train (t_1), and deal with t_1 as an equal]$_2$ nor did Glen reach any agreement with t_1 [that angry middle-aged person]$_1$. (taken from Postal, 1998: 199. Traces between brackets are ours)

But, as we have already said, this does not provide unambiguous evidence in favour of any particular treatment. Postal (1998: 178) claims that '*the base-generation/in-situ view is of course entirely impotent for those RNR structures that involve what are called interwoven dependencies*', but this claim needs to be appropriately restricted: Postal makes this rather strong claim in the context of a discussion of Bošković (2004) (which was then a manuscript not widely available). We must emphasise that the limitations recognised by Postal hold only for base-generation hypotheses which *also* assume SMC-complying phrase structure trees (see, for instance, Kayne, 1994: 67-68, who supports Wexler & Culicover's 1980 deletion analysis and rejects the discontinuous analyses in McCawley, 1982 because the latter is not compatible with the apriorisms of the *antisymmetric* enterprise).

It seems to be the case that L-extractions and RNR behave in the same way under specific structural conditions, namely, *et-coordinated* structures to which a rule applies *across the board*. In that respect, we agree with Postal in that L-extractions and RNR are derived *by the same means*, but in the present view 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 or MG there are no constituents as such (and recall that RNR is, in McCawley's view, a rule that only changes linear order; just like *parenthetical placement*). What seems clear is that *neither RNR nor L-extractions modify existing grammatical relations*.

In this context, it would be a mistake to group RNR with other rightward displacements (there is no unified category of 'R-extractions', which is also the case for L-extractions in Postal, 1998). Specifically, Relative Clause Extraposition (RCE) and Heavy NP shift do not work in the same way RNR does (examples from McCawley, 1998: 529. Annotations are ours):

248) 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)

249) 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?

In the McCawley-Levine view, to which we adhere, RNR is an unbounded process which does not change grammatical relations (objects are still objects; only linear order is disrupted); in this sense it does side with L-extractions. In contrast, we get heavily constrained displacement in RCE and Heavy NP shift, which are both cycle-bounded operations: specifically, S is a bounding node for Extraposition from VP and VP is a bounding node for Extraposition from NP (Kroch & Joshi, 1987:

132)[61]. RNR preserves grammatical relations *across* cycles in monotonically growing structures (which results in RNR being an unbounded operation, as pointed out above), whereas RCE and Heavy NP Shift change order *within* a local cycle: we believe this to be the core observation behind the Right Roof Constraint[62]. That is, not just that RCE and HNPS are *cyclic* rules, but that they *cannot be successive-cyclic*. In proof-theoretic terms, the problem seems to be one of derivational timing (pre- vs. post- vs. last-cyclic rules; see e.g. Ross, 1967: 285, ff.) and the interaction between strictly locally cyclic rules and what presumably are successive 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:

250) 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 in the framework developed in this monograph, 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 ≠ N' and N' is the root of G',* ***and***
> *b. G' is the smallest cycle that properly contains G*

---

[61] A successive-cyclic approach to L-extractions, incidentally, runs into the problem that the set of bounding nodes for Extraposition is larger than the set of bounding nodes for L-extractions and RNR if all these are assumed to be derived by the same means, whichever these are (Kroch & Joshi, 1987: 132; Baltin, 1981).

[62] In Ross' (1967: 307) terms,

> *Any rule whose structural index is of the form ...A Y, and whose structural change specifies that A is to be adjoined to the right of Y, is upward bounded.*

II. P creates an edge between N and N' where

a. N ≠ N' and N' is the root of G', ***and***

b. There is at least one single rooted graph G" such that G ⊊ G" ⊊ 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 (250a), in which *strong cyclicity* comes into play (each cycle counts); condition II refers to processes of the kind (250b), in which only the *last cycle* is relevant; and condition III refers to processes of the kind (250c), 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 (250a) 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:

251) 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). But we have said nothing explicitly about *deletion* (but see **Section 5.3-5.4** for an account of *Equi-NP 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:

252) a. Gapping
b. Ellipsis
c. Sluicing
d. Stripping
e. Bare Argument Ellipsis
…

It is crucial to note that, just like *movement* or *extraction*, we are using *deletion* in a purely descriptive sense, attending to the long history that the term has in generative grammar. In a restrictive framework like the Standard Theory, we can define a *deletion* rule as any rule that replaces at least one non-zero index in a structural description by zero. That is, roughly:

If the structural index of a transformation has n terms a_1, a_2, a_3, …a_n it is a ***deletion*** *rule iff*

i. *its structural change has any a_i replaced by 0, and*
ii. *a_i is neither adjoined to some a_j nor is it substituted by some a_k in the structural index*

In this sense, for instance, Ross' (1969: 267) formulation of *sluicing* is a deletion transformation:

253) W – [X – [-Def]$_{NP}$ – Y]$_S$ – Z – [NP – [$_S$ X – (P) – Y]]$_S$ – R

1	2	3	4	5	6	7	8	9	10

→ OPT

1	2	3	4	5	0	0	8	0	10

Condition: 2 = 7
4 = 9
6⁀7⁀8⁀9 is an embedded question

A specific characterisation of each of the rules in (252) is beyond the scope of the present work, but a general characterisation of the class to which those rules belong is not. This general characterisation will necessarily overlook details and quirks of each rule, but we are confident it will provide some general guidelines for the interested reader to pursue a thorough study of any of the aforementioned processes.

A general requirement on deletion operations is that '*all syntactic deletion be recoverable*' (Hankamer, 1979: 2), which means simply that we *cannot* 'delete' syntactic objects such that afterwards (a) there is no indication that a rule has applied at all, and (b) we have no idea what has been deleted (see also Katz & Postal, 1964: 80). This requirement will be important, as a reality check on the structural descriptions we propose: are relations preserved after 'deletion' such that a complete propositional form can be built?

We will focus on those structures where reconstruction results in an illicit structural description: the reason is that a transformational approach to these runs into problems more obviously than in those cases where reconstruction is a viable option (see Barss, 2001 for an overview of 'reconstruction' in syntax). Another way to look at these is to say that deletion has applied to *repair* a syntactic violation. Consider, for instance, cases in which the remnant of *sluicing* does not correspond to any plausible reconstructible syntactic object:

254) a. They want to hire someone who speaks a Balkan language, but I don't know
$$\begin{Bmatrix} which \\ *\ which\ they\ do \\ *\ which\ Balkan\ language\ they\ want\ to\ hire\ someone\ who\ speaks\ t \end{Bmatrix}$$ (Wh-Island repair under sluicing; taken from Merchant, 2008: 138)

b. He is writing something about deletion rules, but I can't imagine what ~~he is writing [*t* about deletion rules]~~ (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. Taken from Culicover & Jackendoff, 2005: 268)

Recall that we are using 'deletion' as a *descriptive* term, without assigning any theoretical weight to it. It must thus not be assumed that by saying that *gapping*, *sluicing*, or *stripping* are 'deletion' operations we are implying adherence to a treatment of those constructions in which syntactic objects to be found at some level of representation are erased and thus do not appear at some subsequent level. This is, certainly, a possibility: transformational generative grammar has explored this view, and suggested different motivations to have specific syntactic objects 'deleted' from a structural description. The following quote from Merchant (2008: 133) showcases the core methodology of MGG:

> *Assume that PF deletion is triggered by the presence of a feature on a head. Let us call this triggering feature E. Ideally, E will have exactly those syntactic, phonological, and semantic effects that yield all the attested properties of the elliptical construction at hand, with nothing further needing to be said.*

This view is, to different extents, adopted in the literature that assumes that ellipsis and other deletion operations are part of the dynamics of what chomskyan MGG refers to as *phases*: syntactic objects where all featural requirements are satisfied (such that all unvalued uninterpretable features are valued and deleted at the phase level) and which attempt to capture cyclicity effects in the same way Barriers and Bounding Nodes did in previous versions of the theory (Chomsky, 2001; Müller, 2011; Richards, 2011). Phases, in the simplest versions of the theory, are *v** (the 'light verb' of causative constructions) and C, with D and P being candidates for phasehood in some languages (parametrically determined) or in certain syntactic contexts (Bošković, 2014). Wurmbrand (2017: 348), for instance, assumes that

> *phase heads trigger Spell-Out of their complements, and ellipsis is the option of not realizing a Spell-Out domain (SOD) at PF* […] *Thus, elided constituents are unpronounced SODs. I will refer to this approach as the zero Spell-Out approach.*

Phasehood is determined in the syntactic component (since phase boundaries are defined by the presence of feature-checking domains; Chomsky, 2001), and the level of representation at which certain objects are deleted is post-syntactic: the level of Phonetic Form. We will not comment on the MGG approach, which has paid quite a bit of attention to ellipsis and related phenomena, because the theory presented in this work is orthogonal to some of MGG's most basic assumptions and axioms. May it suffice to say that assigning arbitrary features to certain syntactic objects and saying that such assignment constitutes anything resembling 'explanation' is a position that we do not share at all.

If reconstruction is so problematic in cases where deletion 'repairs' structures that would otherwise constitute violations of constraints of the grammar, perhaps a way to go is not to have reconstruction at all. Lechner (2013: 242, ff.) presents the possible transformational scenarios in a very clear way:

> *In configuration [α… [β… , nodes reflexively dominated by α are interpreted*
>
> *a. above β or below β (optional reconstruction)*
>
> *b. only above β (anti-reconstruction)*
>
> *c. only below β (obligatory reconstruction)*

And goes on to say that '*Detecting any of these constellations indicates that α has reached its position by movement*' (2013: 242), where 'movement' is defined in terms of *occurrences* of subtrees in distinct contexts in subsequent linguistic representations (2013: 235). In principle, these possibilities could be used to build a classification of syntactic mapping processes in specific languages: there is no reason to assume that because two languages L and L' feature a process that superficially looks the same it has to follow the same reconstruction typology.

In any case, deletion transformations have to be handled with care: Berwick (1984) makes a strong case for blaming the computational problems in the early Standard Theory on unbounded deletion, which in turn leads to the necessity to reconstruct a pre-transformational phrase marker (call it Deep Structure) in order to build a semantic representation (see also Peters & Ritchie, 1973). Unbounded deletion, in essence, can make a surface string arbitrarily shorter than its corresponding Deep Structure, which makes recoverability for purposes of semantic interpretation potentially intractable. This point had already been made –with an empirical rather than purely computational motivation- in Ross (1970: 249):

> *Deep structures can contain elements or even whole clauses which do not appear in surface structure, and the order in deep structure of elements which appear in both levels of representation may be far different from the surface structure order of the same elements. Furthermore, it seems to be the case that eyen in apparently simple sentences, the transformational mapping between deep and surface structure is extremely complex - far more so, in fact, than has previously been thought. These facts make it extremely difficult to ascertain the nature of deep structure, and necessitate the use of long chains of inference to this end*

If we follow this line of reasoning, we can plausibly think that the problem is not *recoverability* per se (which may be related to *reconstruction* in movement terms), but rather *recoverability of deletion*. In this context, we can kill two birds with one stone: there is no need to worry about *recoverability* if there is no *deletion* as a class of syntactic operations in the way we have characterised it (just like there is no *reconstruction* if there is no movement to begin with). Much research pending, let us briefly explore what this perspective can give us. Consider to this end the following example of *sluicing*:

255) They want to invite a linguist who knows $\begin{Bmatrix} a \\ * \ every \end{Bmatrix}$ version of Dependency Grammar, but I don't remember which (= which version of DG; ≠ which linguist)

In (255) the *wh*-word has an antecedent in the local structure, but this is not a *sine qua non* condition. For instance,

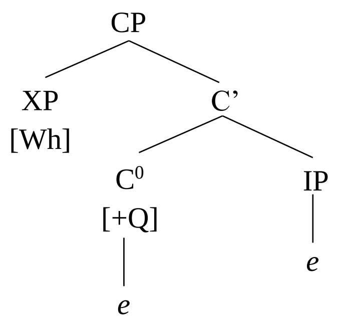
256) He's writing (something), but you can't imagine $\left\{ \begin{matrix} what \\ where \\ why \\ how\ (fast) \\ \begin{Bmatrix} to \\ with \\ for \end{Bmatrix} \ whom \end{matrix} \right\}$ (Ross, 1969: 252)

In cases like (256) the *wh*-word has no antecedent, which simply means that there is no constituent in the sentence with which it would be 'coindexed'. For our present purposes, both (255) and (256) present the same property: as noted in Chung et al. (1995: 241) and Reinhart (2006: 67): the *wh*-word in sluicing can only 'correlate' with an existentially quantified phrase (Chung et al. refer to '*an indefinite or other weak DP*').

We will focus on Recall that in **Section 9** we followed Reinhart (1998) in saying that *wh*-words grammatically instantiate a *choice function* CH: *wh*-interrogatives require the selection of a member of a set. And, following Montague (1973), **?** has scope over an indexed NP. We are interested in the structural description assigned to (255), and the reason why the choice of quantifier matters. In syntactic terms, it is not particularly difficult to account for the interpretation of *which*: if there is no deletion, we can have a multidominated object *a version of dependency grammar*, by **?** and V. But the difference in grammaticality given the desired interpretation remains unexplained: we need to be able to say something more.

In this context, it seems rather natural to assume that an existential quantifier can appear in the scope of CH, but a universal quantifier cannot: indefinites can be viewed as 'restricted free variables' (Reinhart, 2006: 68; also Heim, 1982), which means in this context that they can unselectively be bound by an operator -including **?**. We will not get too deep into the semantic consequences of this assumption, but rather analyse the impact that it has on the structure. With Chung et al., we assume that there is no deletion, but unlike them we do not assume a separate level of Logical Form with its own 'covert' syntactic operations. This assumption is only required, we argue, if structural descriptions are restricted to paths, not trails: it is relevant to point out that one of the basic assumptions that Chung et al. make is that phrase markers have the format of X-bar trees, and sluicing is a process whereby there is *wh*-movement to Spec-CP and C^0 and IP are phonologically null ((257) is taken from Chung et al., 1995: 242):

257)



If X-bar theoretic assumptions about the format of phrase markers are dropped, the problem needs to be rephrased: it is not about constructing a Logical Form for the 'defective' structure in (257) from which the question type meaning can be determined; rather, we need to ensure that there is a well-formed graph that serves as the structural description to sluiced expressions of the language. In our view, there is no need to construct an LF representation from the surface structure of a sluiced sentence: if there is a legitimate graph that corresponds to a sluiced expression of the language, we will have everything we need. We may note that the approach to *wh*-movement developed here satisfies the specifications in Chung et al. (1995: 244):

> *On the syntactic side, the displaced constituent must syntactically bind a position within the IP complement of C^0. On the semantic side, the displaced constituent must contain a Wh-indefinite that is interpreted as a variable semantically bound by an interrogative operator. Third and finally, the displaced constituent must contribute to semantic interpretation just as if it were sitting in the syntactically bound position.*

The 'binding' condition is satisfied by the **?**-replacement operation. The semantic condition is not a condition over structural descriptions, but rather over the expression itself (if indefiniteness is at least partly a matter of form). The final condition is satisfied by the specification of dominance relations in the graph: because what could amount to 'interpretation' is walking a *trail* and not a *path*, the ‘displaced constituent’ *is indeed sitting in the syntactically bound position*. The difference between *binding* and *bound* is simply one of structural context.

If we are going to propose that sluicing involves multidominance, it is crucial to determine exactly *what* is being multidominated. In other words: if we have *some version of Dependency Grammar* as the local antecedent for *which*, what is the syntactic object that appears in the contexts *e*<know, __> and *e*<remember, __>? Specifically, is it the quantified NP *some version of dependency grammar* or only the lexical NP *version of dependency grammar*?

Let us assume that what gets multidominated is only the lexical NP, that is, that the representation ‘bifurcates’ (as it were) *at the cycle* leaving the Q/D layer untouched. This assumption would have undesirable consequences: if the multidominated object excluded the quantifier, then both the existential quantifier and the universal quantifier should be grammatical since the *wh*-word would only have scope over the lexical N. In this context, it is useful to remember the translation rules for quantified NPs *every N* and *a/an N* in Montague (1973):

> *T2. If* $\zeta \in P_{CN}$ [a phrase of category Common Noun] *and* ζ *translates into* ζ', *then* **every** ζ *translates into* $\hat{P} \bigwedge x[\zeta'(x) \rightarrow P\{x\}]$, **the** ζ *translates into* $\hat{P} \bigvee y[\bigwedge x\, [\zeta'(x) \leftrightarrow x = y] \wedge P\{y\}]$, $F_2(\zeta)$ [i.e., **a**/**an** ζ] *translates into* $\hat{P} \bigvee x[\zeta'(x) \wedge P\{x\}]$. (Montague, 1973: 233)

We would like to propose that the Montague-Karttunen representation of (255) goes along the (informal) lines of (258):

258) A version of dependency grammar$_0$ they want to invite a linguist who knows it$_0$ but I don’t know **?** it$_0$

In other words, the indexed pronoun *it$_0$*, as in Montague (1973), stands for the quantified NP, not just for the lexical content. This is consistent with the translation of *wh*-words that we assumed in **Section 9**:

259) *which* N $\equiv \hat{P} \bigvee x[N'(x) \wedge P\{x\}]$

This approach is consistent with the requirements in Chung et al. (1995) in both syntactic and semantic aspects (see the quotation above); it is the purported *goal of the theory of the grammar* that the present approach modifies. Note also that the graph-theoretic approach eliminates the need to specify ‘chains of inference’ (i.e., reconstruction procedures), as a side-effect of eliminating *deletion* altogether as a class of operations over structural descriptions.

Needless to say, the topic of *deletion* is way too vast and complex for us to claim that the present approach works for all cases (particularly, phenomena like Bare Argument Ellipsis, as analysed in Culicover & Jackendoff, 2005: Chapter 7). However, we expect to have provided enough arguments to consider a *trail*-based graph-theoretic approach as an empirically viable and theoretically consistent alternative to transformational accounts of sentence-bound deletion.

13.3 Long distance dependencies and resumptive pronouns

The model presented in this paper requires some revision of the formulation of mechanisms and constraints involved in filler-gap constructions in general, and long-distance dependencies more in particular. We have given a programmatic introduction to the dynamics of *wh*-movement in **Section 9**, which is of course far from comprehensive. The most general condition for filler-gap dependencies is the satisfaction of the conditions imposed in the definition of *licensing* (see (206) above):

informally, *licensing* aims at ensuring *accessibility* between the objects involved in a referential dependency. In this context, consider a sentence like (260):

260) Who$_i$ did John want to tell Mary that he had talked to _$_i$?

The structure of (260) is -essentially- monotonically recursive, with subordinate clauses being arguments of *want* and *tell*:

260') [Who$_i$ did John want [to tell Mary [that he had talked to _$_i$]]]

Thus, there is no post-cyclic adjunction or anything of the sort that we must be on the lookout for: conditions *i* (the existence of a trail in which the relevant node transitively dominates itself) and *ii* (which pertains to root adjunction) are thus met. Now, what about condition *iii*, which filters out dependencies into self-contained graphs? If the structural description indeed features no adjunction, then every node in the ρ-set of (260) is in the ρ-domain of *who* (including itself, transitively). This means that there is no self-contained domain in (260), and therefore licensing can take place without problems. So far, so good.

But let us now consider a different kind of example, slightly more complex:

261) Which picture$_i$ did they all blush when John saw *(it$_i$)? (adapted from Kaplan & Zaenen, 1995: 139)

There is a fundamental difference between (260) and (261): only (261) features an adjunct, the *when* clause. This contrasts with the monotonic nature of (260), in which –in phrase structural terms- *want* takes as a complement the clause headed by *tell* and *tell* takes as a complement the clause headed by *talk*. The * indicates that if the pronoun *it* is omitted, the sentence is ungrammatical. We can frame this more technically: *chopping* from the adjunct is not possible, but *copying* yields an acceptable output (we use these terms in the sense that they have in Ross, 1967: 427).

Following Ross, we assume that *chopping* and *copying* reordering transformations must be distinguished formally: here, we want to defend the idea that only if reordering of *which picture* is the product of *copying* (thus, if it leaves a 'copy' behind, in a sense that we will refine shortly) will we obtain a grammatical output. This is relevant because, in Ross' own words, '*chopping rules are subject to the constraints in Chapter 4* [the Complex NP Constraint, the Coordinate Structure Constraint, the Sentential Subject Constraint, etc.], *copying rules are not*' (1967: 428). In the case of (261), it is usual to invoke a constraint along the lines of an *adjunct condition* which bans extraction from a syntactic object SO if there is no subcategorisation involved in SO (e.g., Huang, 1982: 503; Johnson, 2003); the Empty Category Principle (the requirement that a trace be properly governed; see Chomsky, 1981: 250) goes along the same lines. Transformational approaches define restrictions in terms of the relations between operators and variables, and specify the conditions under which these relations can hold. In this sense, they are in fact establishing conditions over the well-formedness of phrase markers, such that a phrase marker will be well-formed if and only if it violates no principle of the grammar. Nothing is said, however, about expressions of the language.

The theory of locality as conceived of in this work is based on two main concepts: *accessibility* and *cyclicity*. *Accessibility,* as per the definition of *licensing* and *self-containment* aims at specifying the conditions under which scope and bound occurrences of a syntactic object are part of a well-formed *trail*. Conditions on *cyclicity* pertain to those structures in which we have several lexical predicates with their corresponding functional modifiers and nominal dependants: more specifically, we are interested in defining the configurations in which we can *link* cycles (which in turn depend on *accessibility*). In this context, the question is: what is the role of resumptive pronouns in expressions like (261)? First, we should define what we understand by *resumptive pronoun*. Descriptively, resumptive pronouns are pronouns that '*appear in positions where one would, in a certain sense, have expected to find a gap*' (McCloskey, 2006: 95). This is important, theoretically, because it implies that the dynamics of resumptive pronouns are linked to whatever mechanism is assumed to account for displacement. Resumptive pronouns are obligatorily bound, unlike garden-variety pronouns (which can have exophoric reference). Resumption is a wider phenomenon than just resumptive pronouns, including epithet NPs (like *that poor idiot* in *He's a guy who we couldn't understand how that poor idiot could get a job*); here we are only concerned with pronouns since we have given a programmatic account of *pronominalisation* in graph-theoretic terms: we can therefore ask whether resumptive pronouns constitute an exceptional phenomenon or not from the perspective of how these pronouns come to be. The null hypothesis is that resumptive pronouns are the result of *node pronominalisation*, like any other *bound* pronoun in the grammar (see Postal, 1969). However, *why* resumptive pronouns appear in specific contexts is far from being a straightforward queston.

It is a usual claim that resumptive pronouns are somewhat of a *last resort* to rescue an otherwise ungrammatical structure: for some reason, an operator-variable relation that would have violated a constraint is saved from ungrammaticality if the variable is spelled out as a pronoun (e.g., Ross, 1967; Kroch, 1981; Shlonsky, 1992: 443; Boeckx, 2003; also Alexopoulou, 2010 from a processing perspective). It is important to note that the various perspectives that have arisen since Ross (1967) about resumption as a last resort (transformational vs. base-generation origin of the pronoun, mainly) present similar problems when viewed from our perspective: in both cases it is necessary to specify the conditions under which a relation is legitimate that would otherwise violate a constraint; this relation may be operator-variable (if resumption is a form of *stranding*, as in Boeckx, 2003: 25, ff.) or co-indexing (if there is no movement rule involved in resumption, as assumed in Chomsky, 1977: 80-81). In either case there is a relation that needs to be appropriately characterised, including a specification of the structural descriptions where that relation can hold without violating a constraint of the grammar. Kroch (1981) correctly identifies problems with both a movement and a base-generation analysis of resumption; essentially, the rules that can be invoked in the grammar are either too restrictive or not restrictive enough. Note that these problems pertain to the *generative power* of the grammar, its capacity to recursively enumerate structural descriptions (Chomsky, 1965): things look very different if we consider *expressions* of the language instead of *generative functions*. The reason is that there is no 'overgeneration' to be worried about, as the goal of grammatical theory is quite different in model-theoretic syntax: we aim at proving (by means of fully explicit dominance sets which exhaustively characterise graphs) that specific sentences are well-formed expressions of the language; in this sense our goals are those of IP grammars (Hockett, 1954; Schmerling, 1983a). For example, in a 'pure' CG, a proof that an expression is (or isn't) a well-formed expression of category C defines the relations that basic and derived expressions establish within the algebra of the language. The theory developed here does not make use of indexed categories in the specification of connectivity relations (although it could, at no cost: ρ-sets could be generalised by making reference to indexed categories instead of the basic expressions that belong to those categories), but other than that, we share pure CG's goals. This is important because –we repeat- by doing this some problems pertaining to the adequacy of rules and constraints simply do not arise.

After a brief methodological parenthesis, let us go back to the analysis. Above, we asked 'what is the role of resumptive pronouns in structures like (261)?' Now, we have the elements to formulate an answer to that question: the need for a resumptive pronoun in (261); more specifically, the need for an occurrence of *picture*' in the *when*-adjunct in the expression of the language, arises so that the adjunct clause is *not self-contained*, and therefore the conditions for *licensing* hold. If the adjunct clause was *self-contained*, then the embedded verb *saw* would be left with no object, and *who* would have no thematic role (recall that scope requirements are divorced from thematic requirements in the present framework, and 'reconstruction' in VP internal positions obeys the latter, not the former). In the view presented here, resumptive pronouns are mechanisms to *join* cycles, in the example that interests us: *it* corresponds to the same node as *which pictures*, getting pronominalised as it is visited for the second time in the trail. Let us provide the ρ-set for (261):

262) Cycle 1: ρ = ⟨(*picture*', blush), (blush, *they*'), (blush, when)⟩
Cycle 2: ρ = ⟨(when, saw), (saw, *John*'), (saw, *picture*')⟩

We see that *picture*' belongs to both cycles, and, what is more, (*picture*', *picture*') ∈ ρ* by virtue of the V *blush* dominating the root of Cycle 2 (which intends to capture, much discussion pending, that the *when*-clause is a VP adjunct).

The problem of resumption is much more complex than we can cover in a simple note, but it seems that we can capture the insight in Zaenen et al. (1983: 679) that

> *the binding relation between a wh-element and a "resumptive" pronoun is, at least in some languages, of the same nature as the binding relation between a wh-element and a trace*

Because in both cases we are dealing with the same kind of phenomenon: a trail visiting a node more than once. Note that if the *when* clause was self-contained, it would be opaque for purposes of operations at the matrix clause, because the output of any operation purporting to involve elements in the adjunct and the matrix clause would violate the licensing conditions. We also capture the resistance of monotonic structures to resumptive pronouns: they are simply not needed to comply with the licensing conditions in monotonically growing structures regardless of structural distance if no constraint is violated. The present view thus aligns itself with *resumption-as-last-resort* approaches: resumption is a mechanism that the grammar appeals to when there is no other option. No other option for what? Here we proposed that resumption is a mechanism by means of which the structural conditions for *licensing* can be satisfied and therefore cycles can be appropriately *linked*. Why, though? Because the sentences we are considering, with resumptive pronouns, are well-formed expressions of the language. We agree with Alexopoulou (2010: 487) in that '*it is not the case that grammars license* [intrusive] *resumption*'; rather, grammars license a range of structural configurations and resumption appears as a last resort in a limited number of those configurations in the place of an 'illicit' gap (in English, relevant configurations promimently feature relative clauses and *wh*-interrogatives). It is consistent to say that giving a node with at least two occurrences (one in a scope position, one in a bound position) a pronominal exponent if otherwise the relation between *operator* and *variable* occurrences of a node (or *scope* and *restrictor* occurrences) could not be unambiguously determined. As McCloskey (2006: 19) correctly notes,

> *It is known that resumptive elements may serve the purpose of marking variable positions in unbounded dependency constructions. It is known that resumptive elements may occur in positions from which movement is impossible (hence apparently allowing greater expressive power than is permitted by movement alone). It is also known that resumption imposes a considerably lighter burden on the human sentence processor than does the production and resolution of syntactic movement configurations. Why, then, is movement used at all in the creation of these structures?*

McCloskey's paradox emerges only if resumption is, as he puts it, *used in the creation* of structural descriptions under a *displacement-as-movement* view (which would lead to overgeneration under an economy-based approach to language like early Minimalism; Chomsky, 1995); it does not arise, however, if resumption is viewed as a *repair* strategy related to the conditions of *licensing* rather than a *generative* one related to the dynamics of Move-α (cf. McCloskey, 2002).

The analysis proposed here seems to have some advantages when looked at in interaction with other syntactic phenomena. Here we will simply sketch an instance of such interaction. Going back to our discussion about *topicalisation* and *focalisation*, if adjunction creates the kind of structural configuration in which resumptive pronouns can be called upon to link cycles, and if *topicalisation* (but not *focalisation*) is indeed an instance of *adjunction* (as we proposed in **Section 5.5**), then we have a possible explanation for the following contrast:

263) a. Syntax, Mary loves (it)
b. It is syntax that Mary loves (*it)

In (263a) we are in presence of topicalisation of *syntax*; if that topicalisation is indeed adjunction of *syntax* to the root, it is to be expected that a resumptive pronoun *can* be called upon to link the main clause and the adjoined element, but that resumptive pronoun is optional because we are dealing with a monotonic structure, regardless of the structural distance between gap and filler.

However, in (263b) the resumptive pronoun is banned, because it is not required. This is another way of expressing the *last resort* character of resumption as a grammatical procedure: if it is not *needed*, it is strongly dispreferred.

Note that if we combine *topicalisation* with extraction from an adjunct (in this case, the source of extraction is a relative clause, which is adjoined to NP), things get much worse really quick, as we can see in (264b):

264) a. Syntax, John thinks that Mary said that she loves (it)
b. Syntax, John knows a girl who likes *(it)

(264a) keeps things monotonic: we have a sequence of verbs taking finite clauses as complements. We have three cycles, structured around *thinks*, *said*, and *loves* but none of those is *self-contained* and the cycles of *said* and *loves* are the 2 of their respective selecting predicates. In this context, the relation between *scope* and *restrictor* occurrences in (264b) can be unambiguously determined, and since the *filler-gap* relation would violate the Complex NP Constraint, it is possible to resort to *resumption* as a way to generate a licit configuration in terms of locality (and perhaps also accessibility; see Keenan & Comrie, 1977).

A final note is in order. For our hypothesis about what is causing the contrast between (263a) and (264a), and (263b) and (264b) to work, we would need to commit to the claim that clefting essentially occurs *within a single cycle.* In other words, that it is structurally akin to *focus*: recall that in **Section 5.5** we argued that *focalisation* involved monotonicity and thus NPIs could be licensed in focus constructions. If, as argued e.g. in Uriagereka (2002) monotonicity correlates with syntactic accessibility (and thus, with locality), the resistance of focus constructions to displaying resumption would be at least partially accounted for: there is no need to resort to resumption in sentences like (263b) because clefting, being parallel to focus, involves local enough relations. This is a working hypothesis, however. Whether this is actually the case is far from clear, although the semantic / pragmatic relation between focus and clefting has been noted as early as Jespersen (1985 [1937]), and even MGG-inspired research within the so-called *cartographic* approach has proposed that clefting in fact makes use of a functional Focus projection (e.g., Kiss, 1999; Belletti, 2008).

14. (Some) conclusions

We began this exploration by asking a simple question: what if the grammar attempted to minimise the number of nodes and maximise connectivity between those nodes, instead of maximising 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, 2018), 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 want to- 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 (2016a, b, 2018), Saddy's (2018), and Saddy & Krivochen's (2017) oscillatory dynamical computational system, that is *just one of several internally consistent options* (for instance, Gradient Symbolic Computation –Smolensky et al., 2014- 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

Abels, Klaus & Kristine Bentzen (2012) Are Movement Paths Punctuated or Uniform? In Artemis Alexiadou, Tibor Kiss & Gereon Müller (Eds.) *Local Modelling of Non-Local Dependencies in Syntax.* Berlin: Walter de Gruyter. 431-452.

Abney, Steven Paul (1987) *The English Noun Phrase in its Sentential Aspect*. PhD Thesis, MIT.

Abrusán, Marta (2014) *Weak island semantics*. Oxfrod: OUP.

Aikhenvald, Alexandra (2004) *Evidentiality*. Oxford: OUP.

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.

Alexopoulou, Theodora (2010) Truly intrusive: Resumptive pronominals in questions and relative clauses. *Lingua*, 120. 485-505.

Altham, J. E. J. & Neil W. Tennant (1975) Sortal quantification. In Edward Keenan (ed.) *Formal semantics of natural language*. Cambridge: CUP. 46-58.

Asudeh, Ash (2005) Control and semantic resource sensitivity. *Journal of Linguistics* 41(3). 465–511.

Asudeh, Ash & Ida Toivonen (2012) Copy raising and perception. *Natural Language and Linguistic Theory*, 30(2). 321-380.

(2017) A modular approach to evidentiality. In Miriam Butt, Tracy Holloway King (eds.) *Proceedings of the LFG'17 Conference*. Stanford: CSLI Publications. 45–65.

Bach, Emmon (1964) *An Introduction to Transformational Grammars*. New York: Holt, Rinehart & Winston.

(1970) Problominalization. *Linguistic Inquiry*, 1. 121-122.

(1979) Control in Montague Grammar. *Linguistic Inquiry*, 10(4). 515-531.

(1983) Generalized Categorial Grammars and the English Auxiliary. In Frank Heny & Barry Richards (eds.), *Linguistic categories: auxiliaries and related puzzles*, Vol 2. Dordrecht: Reidel. 101–120.

Bach, Emmon & Robin Cooper (1987) The NP-S Analysis of Relative Clauses and Compositional Semantics. *Linguistics and Philosophy*, 2(1). 145-150.

Bach, Emmon & George Horn (1976) Remarks on "Conditions on Transformations". *Linguistic Inquiry* 7(2). 265-299.

Baltin, Mark (1981) Strict Bounding. In Carl L. Baker & John McCarthy (eds.) *The Logical Problem of Language Acquisition*. Cambridge: MIT Press. 257 295.

(1982) A Landing Site Theory of Movement Rules. *Linguistic Inquiry* 13(1). 1-38.

(2006) Extraposition. In Martin Everaert and Henk van Riemsdijk (eds.) *The Blackwell Companion to Syntax*. Oxford: Blackwell. 237-271.

Barss, Andrew (1986) *Chains and Anaphoric Dependence*. PhD Thesis, MIT.

(2001) Syntactic reconstruction effects. In Mark Baltin & Chris Collins (eds.) *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell. 670-696.

Baunaz, Lena & Eric Lander (2018) Nanosyntax: the basics. In Lena Baunaz, Liliane Haegeman, Karen De Clercq, and Eric Lander (eds.) *Exploring Nanosyntax*. Oxford: OUP. 3-56.

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.

Belletti, Adriana (1988) The Case of Unaccusatives. *Linguistic Inquiry* 19(1). 1-35.

(2008) The CP of Clefts. *Rivista di Grammatica Generativa* 33. 191-204.

Van Benthem, Johan (1988) The Lambek Calculus. In Richard Oherle, Emmon Bach, and Deirdre Wheeler (eds.) *Categorial 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.

Boeckx, Cedric (2003) *Islands and chains: resumption as derivational residue*. Amsterdam: John Benjamins.

(2012) Phases beyond explanatory adequacy. In Angel Gallego (ed.) *Phases: Developing the Framework*. Berlin: Mouton de Gruyter. 45-66.

Boeckx, Cedric & Kleanthes Grohmann (2004) Barriers and Phases: Forward to the Past? Presented in *Tools in Linguistic Theory* 2004 (TiLT) Budapest (May 16-18, 2004). http://people.umass.edu/roeper/711-05/Boexce%20barriers+phases%20tilt_bg_ho.pdf

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.

Bošković, Željko (2014) Now I'm a Phase, Now I'm Not a Phase: On the Variability of Phases with Extraction and Ellipsis. *Linguistic Inquiry*, 45(1). 27-89.

Bouchard, Denis (1982) *On the Content of Empty Categories*. PhD Thesis, MIT.

Brame, Michael (1978) *Base Generated Syntax*. Seattle: Noit Amrofer.

Bravo, Ana (2016) Verbos auxiliares. In Javier Gutiérrez-Rexach (ed.), *Enciclopedia de Lingüística Hispánica* Vol. 2. London: Routledge. 152-162.

(2020) On pseudo-coordination in Spanish. *Borealis* 9(1). 125-180.

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, Ash Asudeh, Ida Toivonen & Stephen Wechsler (2016) *Lexical Functional Syntax* [2nd Edition]. Oxford: Wiley Blackwell.

Bresnan, Joan & Jane Grimshaw (1978) The Syntax of Free Relatives in English. *Linguistic Inquiry*, 9(3). 331-391.

Bresnan, Joan, Ronald M. Kaplan, Stanley Peters & Annie Zaenen (1982) Cross-Serial Dependencies in Dutch. *Linguistic Inquiry*, 13(4). 613-635.

Buszkowski, Wojciech (1988) Generative Power of Categorial Grammars. In Richard Oherle, Emmon Bach, and Deirdre Wheeler (eds.) *Categorial Grammars and Natural Language Structures*. Dordrecht: Reidel. 69-94.

Camacho, José (2003) *The Structure of Coordination*. Dordrecht: Kluwer.

Carrasco Gutiérrez, Angeles & Luis García Fernández (1994) Sequence of Tenses in Spanish. *University of Venice Working Papers in Linguistics* 4(1). 45-70.

Čech, Radek, Ján Mačutek & Zdeněk Žabokrtský (2011) The role of syntax in complex networks: Local and global importance of verbs in a syntactic dependency network. *Physica A*, 390. 3614-3623.

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.

(1970a) Deep structure, surface structure, and semantic interpretation. In Roman Jakobson & Shigeo Kawamoto (eds.), *Studies in general and oriental linguistics presented to Shirô Hattori on the occasion of his sixtieth birthday*. Tokyo: TEC Corporation for Language and Education Research. 52–91.

(1970b) Remarks on nominalization. In Roderick Jacobs & Peter Rosenbaum (eds.) *Readings in English Transformational Grammar*. Waltham: Ginn & Co. 184-221.

(1973) Conditions on transformations. In Stephen Anderson & Paul Kiparsky (eds.) *A festschrift for Morris Halle*. New York: Holt, Rinehart, and Winston. 232–286. Reprinted in Noam Chomsky, *Essays on Form and Interpretation*. New York: North Holland. 81-160.

(1980) On Binding. *Linguistic Inquiry* 11(1).1-46.

(1981) *Lectures on Government and Binding.* Dordrecht: Foris.

(1982) *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, Mass.: MIT Press.

(1986) *Barriers*. Cambridge, Mass.: MIT Press.

(1995) *The Minimalist Program*. Cambridge, Mass.: MIT Press.

(2000) Minimalist Inquiries: the framework. Minimalist Inquiries: The Framework. In Roger Martin, David Michaels and Juan 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 & Howard Lasnik (1993) The theory of Principles and Parameters. In Joachim Jacobs, Arnimvon Stechow, Wolfgang Sternefeld, and Theo Vennemann (eds.) *Syntax: An international handbook of contemporary research*, vol. 1. Berlin: Walter de Gruyter. 506–569.

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.

Chung, Sandra & William Ladusaw (2003) *Restriction and Saturation*. Cambridge, Mass.: MIT Press.

Chung, Sandra, William Ladusaw & James McCloskey (1995) Sluicing and Logical Form. *Natural Language Semantics* 3(3). 239-282.

Church, Alonso (1936) An unsolvable problem of elementary number theory. *American Journal of Mathematics* 58, 354–363.

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 & Paul Postal (2014) *Classical NEG Raising*. Cambridge, Mass.: MIT Press.

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.

Culicover, Peter, Ray Jackendoff & Jenny Audring (2017) Multi-word constructions in the grammar. *Topics in Cognitive Science* 9(3). 552-568.

Culicover, Peter & Ray Jackendoff (2005) *Simpler Syntax*. Oxford: OUP.

Dalrymple, Mary (2001) *Lexical Functional Grammar* (Syntax and Semantics 34). New York: Academic Press.

(2007) Introduction to LFG. *LSA Linguistic Institute* 2007, Day 3. Slides available online at https://www1.essex.ac.uk/linguistics/external/LFG/www-lfg.stanford.edu/pubs/lfg-presentations/institute07/lecture3.pdf

Dalrymple, Mary, Ronald Kaplan & Tracy Holloway King (2015) Economy of expression as a principle of syntax. *Journal of language modelling*, 3(2). 377-412. http://dx.doi.org/10.15398/jlm.v3i2.82

Davidson, Donald (1967) The logical form of action sentences. In Nicholas Resher (ed.). *The Logic of Decision and Action.* Pittsburgh, PA: University of Pittsburgh Press. 81–95.

(1969) The individuation of events. In Nicholas Resher (ed.). *Essays in Honor of Carl G. Hempel.* Dordrecht: Reidel. 216–234.

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.

Del Gobbo, Francesca (2007) On the syntax and semantics of appositive relative clauses. In Nicole Dehé & Yordanka Kavalova (eds.) *Parentheticals*. Amsterdam: John Benjamins.

De Vries, Mark (2007) Invisible Constituents? Parentheses as B-Merged Adverbial Phrases. In Nicole Dehé & Yordanka Kavalova (eds.) *Parentheticals*. Amsterdam: John Benjamins. 203–234.

Di Sciullo, Anna-Maria (2005) *The Asymmetry of Morphology*. Cambridge (Mass.): MIT Press.

(2014) Minimalism and I-Morphology. In Peter Kosta, Steven L. Franks, Teodora Radeva-Bork & Lilia Schürcks (eds.), *Minimalism and Beyond: Radicalizing the interfaces*. Amsterdam: John Benjamins. 267-286.

Dixon, Robert M. (1972) *The Dyirbal language of North Queensland*. Cambridge: CUP.

Donnellan, Keith S. (1966) Reference and Definite Descriptions. *The Philosophical Review* 75(3). 281–304.

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.

(2003) The dual analysis of adjuncts/complements in Categorial Grammar. In Ewald Lang, Claudia Maienborn and Cathrine Fabricius-Hansen (eds.) *Modifying Adjuncts*. Berlin: Mouton de Gruyter. 33-66.

Embick, David & Rolf Noyer (2007) Distributed Morphology and the Syntax-Morphology Interface. In Gillian Ramchand and Charles Reiss (eds.) *The Oxford Handbook of Linguistic Interfaces.* Oxford: OUP. 289-324.

Emonds, Joseph (1970) *Root and structure preserving transformations*. PhD Dissertation, MIT.

(1979) Appositive relatives have no properties. *Linguistic Inquiry* 10. 211-243.

(2007) *Discovering Syntax: Clause Structures of English, German and Romance*. Berlin: Mouton de Gruyter.

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.

(2015) On I(nternalist)-Functional Explanation in Minimalism. In Samuel Epstein, Husatsugu Kitahara, & T. Daniel Seely (comp.) *Explorations in Maximizing Syntactic Minimization*. London: Routledge. 71-97.

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 Samuel Epstein, Husatsugu Kitahara, & T. Daniel Seely (comp.) *Explorations in Maximizing Syntactic Minimization*. London: Routledge. 175-194.

Epstein, Samuel D. & T. Daniel Seely (2002) Rule Applications as Cycles in a Level Free Syntax. In Samuel D. Epstein & T. Daniel Seely (eds.) *Derivation and Explanation in the Minimalist Program*. Oxford: Blackwell. 65-89.

Fang, Yu & Haitao Liu (2018) What factors are associated with dependency distances to ensure easy comprehension? A case study of *ba* sentences in Mandarin Chinese. *Language Sciences* 67. 33-45.

Falk, Yehuda N. (2001) *Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax*. Stanford, CA: CSLI Publications.

Ferrer-i-Cancho, Ramón (2014) Beyond description. Comment on "Approaching human language with complex networks" by Cong & Liu. *Physics of Life Reviews* 11(4). 621–623.

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.

(2002) *Phrase Structure Composition and Syntactic Dependencies*. Cambridge, Mass.: MIT Press.

(2006) Phase theory and Tree Adjoining Grammar. *Lingua*, 116(2). 145–202.

Friedmann, Naama, Adriana Belletti & Luigi Rizzi (2009) Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, 119(1). 67-88.

García Fernández, Luis (2006) Perífrasis verbales en español. In Luis García Fernández (dir.) *Diccionario de Perífrasis Verbales*. Madrid: Gredos. 9-58.

García Fernández, Luis & Diego Krivochen (2019) Dependencias no locales y cadenas de verbos auxiliares. To appear in *Verba*, 46. Handout version available: https://www.academia.edu/36611158/Dependencias_no_locales_y_cadenas_de_verbos_auxiliares

García Fernández, Luis, Diego Krivochen & Ana Bravo (2017) Aspectos de la semántica y sintaxis de las cadenas de verbos auxiliares en español. *Moenia* 23. 1-28.

Gazdar, Gerald (1981) Unbounded Dependencies and Coordinate Structure. *Linguistic Inquiry*, 12(2). 155-184.

(1982) Phrase Structure Grammar. In Pauline Jacobson & Geoffrey Pullum (eds.) *The Nature of Syntactic Representation*. Dordrecht: Reidel. 131-186.

Gorn, Saul (1967) Handling the growth by definition of mechanical languages. *Proceedings of the April 18-20, 1967, spring joint computer conference*. New York: Association for Computing Machinery. 213–224.

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, & James Douglas Saddy (2014) Resonant Behaviour of Irreducible Networks of Neurons with Delayed Coupling. Ms.

Gruber, Jeffrey (1965) *Studies in Lexical Relations*. PhD Dissertation, MIT.

Haegeman, Liliane (2009) Parenthetical adverbials: the radical orphanage approach. In Benjamin Shaer, Philippa Cook, Werner Frey & Maienborn Claudia (eds.) *Dislocated elements in discourse: syntactic, semantic and pragmatic perspectives*. London: Routledge. 331-347.

Haider, Hubert (forthcoming) Grammatical rules are discrete, not weighted, and not vulnerable. Ms., Univ. Salzburg, FB Linguistik & Centre for Neurocognitive Research. https://www.academia.edu/37334786/Grammatical_rules_are_discrete_not_weighted_and_not_vulnerable

Hankamer, Jorge (1979) *Deletion in Coordinate Structures*. New York: Garland.

Harary, Frank, R. Z. Norman & Dorwin Cartwright (1965) *Structural models: An Introduction to the Theory of Directed Graphs*. New York: John Wiley and Sons.

Harris, Zellig (1957) Co-Occurrence and Transformation in Linguistic Structure. *Language*, 33(3). 283-340.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298: 1569‑1579. DOI: 10.1126/science.298.5598.1569

Hefferon, Jim (2014) *Linear Algebra*. Free access book. Available online at: http://joshua.smcvt.edu/linearalgebra

Higginbotham, James (2000) On events in linguistic semantics. In James Higginbotham, Fabio Pianesi and Achille Varzi (eds.) *Speaking of events*. New York: Oxford University Press. 49–79.

Hockett, Charles (1954) Two models of grammatical description. *Word*, 10. 210–234.

(1958) *A Course in Modern Linguistics*. The Macmillan Company: New York.

Hoffmann, Ludger (1998) Parenthesen. *Linguistische Berichte* 175. 299‑328.

Horn, Laurence (1989) *A Natural History of Negation*. Chicago: Chicago University Press.

Hornstein, Norbert (1999) Movement and control. *Linguistic Inquiry* 30. 69-96.

Hornstein, Norbert & William Idsardi (2014) A Program for the Minimalist Program. In Peter Kosta, Steven L. Franks, Teodora Radeva-Bork & Lilia Schürcks (eds.), *Minimalism and Beyond: Radicalizing the interfaces*. Amsterdam: John Benjamins. 9-36.

Huang, C. T. James (1982) *Logical Relations in Chinese and the theory of Grammar*. PhD Thesis, MIT.

Jackendoff, Ray (1977) *X′ Syntax: A Study of Phrase Structure*. Cambridge, Mass: MIT Press.

(1987) The Status of Thematic Relations in Linguistic Theory. *Linguistic Inquiry* 18(3). 369-411.

(2011) Alternative minimalist visions of language. In Robert D. Borsley and Kersti Börjars (eds.) *Nontransformational Syntax*. Oxford: Wiley-Blackwell. 268-296.

Jacobson, Pauline (1977) *The Syntax of Crossing Correference 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.

(2000) Paycheck pronouns, Bach-Peters sentences, and variable-free semantics. *Natural Language Semantics* 8. 77–155.

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, Kyle (2003) Towards an Etiology of Adjunct Islands. *Nordlyd* 31(1). 187–215.

(2014) Recoverability of deletion. In Kuniya Nasukawa and Henk van Riemsdijk (eds.) *Identity relations in the Grammar*. Berlin: De Gruyter. 255-288.

(2016) Toward a Multidominant Theory of Movement. Lectures presented at ACTL, University College, June 2016. Available at https://people.umass.edu/kbj/homepage/Content/Multi_Movement.pdf

(forthcoming) Rethinking linearization. In András Bárány, Theresa Biberauer, Jamie Douglas & Sten Vikner (eds.) *Syntactic architecture and its consequences II: Between syntax and morphology*. Berlin: Language Science Press.

Johnson, David & Paul M. Postal (1980) *Arc Pair Grammar*. Princeton, NJ.: Princeton University Press.

Joshi, Aravind K. (1985) Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions? In David Dowty, Lauri Karttunen, and Arnold Zwicky (eds.) *Natural Language Parsing*. Cambridge, Mass.: CUP. 206-250.

Joshi, Aravind K. & Yves Schabes (1991) Tree-Adjoining Grammars and Lexicalized Grammars. *Technical Reports (CIS).* Paper 445. http://repository.upenn.edu/cis_reports/445

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

Kahane, Sylvain & Nicolas Mazziotta (2015) Syntactic Polygraphs: A Formalism Extending Both Constituency and Dependency. *Proceedings of the 14th Meeting on the Mathematics of Language*. Association for Computational Linguistics. 152–164.

Kaplan, Ronald & Annie Zaennen (1995) Long-distance Dependencies, Constituent Structure, and Functional Uncertainty. In Mary Dalrymple, Ronald Kaplan, John T. Maxwell III & Annie Zaenen (eds.) *Formal Issues in Lexical Functional Grammar*. Stanford: CSLI. 137-166.

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 (1981) On certain differences between French and English. *Linguistic Inquiry* 12. 349–371.

(1983) Connectedness. *Linguistic Inquiry*, 14(2). 223-249.

(1984) *Connectedness and Binary Branching*. Dordretch: Foris.

(1994) *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.

(2013) Why are there no directionality parameters? In Theresa Biberauer and Michelle Sheehan (eds.) *Theoretical Approaches to Disharmonic Word Order*. Oxford: OUP. 219-244.

(2018) The place of linear order in the language faculty. Talk delivered at the University of Venice, June 16th, 2018. Available online at: https://as.nyu.edu/content/dam/nyu-as/linguistics/documents/Kayne%200118%20Venice%20The%20Place%20of%20Linear%20Order%20in%20the%20Language%20Faculty.pdf

Keenan, Edward & Bernard Comrie (1977) Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8. 63–99.

Keshet, Ezra (2010) Split intensionality: a new scope theory of de re and de dicto. *Linguistics and Philosophy* 33(4). 251-283.

É. Kiss, Katalin (1999) The English Cleft Construction as a Focus Phrase. In Lunella Mereu (ed.) *Boundaries of Morphology and Syntax*. Amsterdam: John Benjamins. 217- 229.

Kluck, Marlies, Dennis Ott & Mark de Vries (2014) Incomplete parenthesis: An overview. In Marlies Kluck, Dennis Ott & Mark de Vries (eds.) *Parenthesis and Ellipsis*. Berlin: Mouton de Gruyter. 1-22.

Koopman, Hilda & Dominique Sportiche (2000) Variables and the Bijection Principle. In Koopman, Hilda: *The Syntax of Specifiers and Heads*. London: Routledge. 16-36.

Koutsoudas, Andreas (1972) The Strict Order Fallacy. *Language*, 48(1). 88-96.

Koyama, Masanori, Michael Orrison & 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

(2016a) Divide and…conquer? On the limits of algorithmic approaches to syntactic structure. *Czech and Slovak Linguistic Review* 1(2016). 15-38.

(2016b) Language, Computation, and Dynamical Systems. Invited lecture, Palacký University, Olomouc. Part of the *International Guest Lectures Repository* program. https://www.academia.edu/37174821/_2016_Language_Computation_and_Dynamical_Systems

(2018) *Aspects of Emergent Cyclicity in Language and Computation*. PhD Thesis. University of Reading. https://www.academia.edu/19959486/_2017_Aspects_of_Emergent_Cyclicity_in_Language_and_Computation_I_cant_believe_there_are_still_typos_version._Also_other_corrections_and_added_references_

(2019a) On trans-derivational operations: Generative Semantics and Tree Adjoining Grammar. *Language Sciences* 74. 47-76. https://doi.org/10.1016/j.langsci.2019.04.002

(2019b) On intensive endophoric devices in English. To appear in *Studia Anglica Posnaniesia*.

(forthcoming) Some notes on the locally variable complexity of natural language strings. *Evolutionary Linguistic Theory*.

Krivochen, Diego & Luis García Fernández (2018) On Coordination and Clitic Climbing in Spanish Auxiliary Verb Constructions. Ms. Under review.

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. Under review. https://www.academia.edu/25110015/_Squib_Mirage_Coordinations_in_English

(2018) A Categorial Grammar for semi-lexical categories: The case of Spanish auxiliary chains. Ms. Under review.

Kroch, Anthony (1981) On the role of resumptive pronouns in amnestying island violations. In *Proceedings of the Seventeenth Regional Meeting of the Chicago Linguistic Society*. Chicago Linguistic Society. 125-135.

(2001) Asymmetries in Long-Distance Extraction in a Tree-Adjoining Grammar. Ms. Available online at https://www.ling.upenn.edu/~kroch/online.html

Kroch, Anthony & Aravind K. Joshi (1985) The linguistic relevance of a Tree Adjoining Grammar. Available online at http://babel.ling.upenn.edu/papers/faculty/tony_kroch/papers/relevance3.pdf [Accessed on 22/03/2016].

(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.

Kural, Murat (2005) Tree Traversal and Word Order. *Linguistic Inquiry*, 36(3). 367-387.

Laca, Brenda (2004) Romance ‘Aspectual’ Periphrases: Eventuality Modification versus ‘Syntactic’ Aspect. In Jacquline Guéron and Jacqueline Lecarme (eds.) *The Syntax of Time*. Cambridge (Mass.): MIT Press. 425-440.

Ladusaw, William (1980) *Polarity sensitivity as inherent scope relations*. Bloomington, Indiana: University of Iowa, Indiana University Linguistics Club.

Lakoff, George (1965) *On the Nature of Syntactic Irregularity*. PhD Thesis, Indiana University.

(1976) Pronouns and Reference. In James D. McCawley (1976) (ed.). *Notes from the linguistic underground* (Syntax and Semantics 7). New York: Academic Press. 275-336.

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. Terence (1975) Finite-State Parsing of Phrase-Structure Languages and the Status of Readjustment Rules in Grammar. *Linguistic Inquiry* 6(4). 533-554.

Langendoen, D. Terence & Paul M. Postal (1984) *The Vastness of Natural Languages*. Oxford: Blackwell.

Larson, Richard (1987) 'Missing prepositions' and the analysis of English free relative clauses. *Linguistic Inquiry* 18(2). 239-266.

Lasnik, Howard (2011) What Kind of Computing Device is the Human Language Faculty?. In Di Sciullo, A-M. & C. Boeckx (Eds.) *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*. Oxford: OUP. 354-65.

Lasnik, Howard & Joseph Kuppin (1977) A restrictive theory of transformational grammar. *Theoretical Linguistics*, 4. 173–196.

Lasnik, Howard & Mamoru Saito (1991) On the subject of infinitives. *Proceedings of the Chicago Linguistic Society (CLS)* 27. 324–343.

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.

Lasnik, Howard & Juan Uriagereka (forthcoming) *Structure*. Cambridge, Mass.: MIT Press.

Lebeaux, David (1994) Where does Binding Theory apply? Ms. MIT.

(2009) *Where does Binding Theory Apply?* Cambridge, Mass.: MIT Press.

Lechner, Winifred (2013) Diagnosing XP movement. In Lisa Lai-Shen Cheng & Norbert Corver (eds.) *Diagnosing Syntax*. Oxford: OUP. 235-248.

Lees, Robert (1976) What are transformations? In James McCawley (ed.) *Syntax and Semantics 7: Notes from the Linguistic Underground*. New York: Academic Press. 27-41.

Lees, Robert & Edward Klima (1963) Rules for English Pronominalization. *Language*, 39(1). 17-28.

Levine, Robert D. (1985) Right node (non-)raising. *Linguistic Inquiry*, 16(3). 492-497.

Lewis, David (1972) General Semantics. In Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language*. Reidel, Dordrecht.169-218.

Liang, Junying, Yuanyuan Fang, Qianxi Lv & Haitao Liu (2017) Dependency Distance Differences across Interpreting Types: Implications for Cognitive Demand. *Frontiers in Psychology*, 8. DOI: 10.3389/fpsyg.2017.02132

Liu, Haitao, Chunshan Xu & Junying Liang (2017) Dependency distance: a new perspective on syntactic patterns in natural languages. *Physics of Life Reviews* 21. 171-193.

Lyons, John (1968) *Introduction to theoretical linguistics*. Cambridge: Academic Press.

Maienborn, Claudia (2011) Event semantics. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.) *Semantics. An international handbook of natural language meaning*, Vol. 1. Berlin, New York: Mouton de Gruyter. 802-829.

Marcolli, Matilde (2014) Principles and Parameters: a coding theory perspective. arXiv:1407.7169

(2016) Syntactic Parameters and a Coding Theory Perspective on Entropy and Complexity of Language Families. *Entropy* 18, 110. doi:10.3390/e18040110

May, Robert (1985) *Logical Form: Its Structure and Derivation*. Cambridge, Mass.: MIT Press.

Mayr, Clemens (2007) On the Lack of Subject-Object Asymmetries. *UPenn Working Papers in Linguistics* 10(1). 1-14.

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.

(1973) *Grammar and Meaning*. New York: Academic Press.

(1982) Parentheticals and Discontinuous Constituent Structure. *Linguistic Inquiry* 13(1). 91-106.

(1987) Some additional evidence for discontinuity. In Geoffrey Huck & Almerindo Ojeda (eds.) *Syntax and Semantics 20: Discontinuous Constituency*. New York: Academic Press. 185-202.

(1998) *The Syntactic Phenomena of English*. [2nd Edition]. Chicago: University of Chicago Press.

McCloskey, James (2002) Resumption, successive cyclicity, and the locality of operations. In Samuel D. Epstein & T. Daniel Seely (eds.) *Derivation and Explanation in the Minimalist Program*. Oxford: Blackwell. 184–226.

(2006) Resumption. In Martin Everaert and Henk van Riemsdijk (eds.) *The Blackwell Companion to Syntax*. Oxford: Blackwell. 94–117.

McKinney-Bock, Katherine (2013) *Building Phrase Structure from Items and Contexts*. PhD thesis, University of Southern California.

McKinney-Bock, Katherine & Jean-Roger Vergnaud (2014) Grafts and beyond: Graph-theoretic syntax. In Katherine McKinney-Bock & María Luisa Zubizarreta (eds.) *Primitive Elements of Grammatical Theory*. London: Routledge. 207-236.

Medeiros, David P. (2018) ULTRA: Universal Grammar as a Universal Parser. Invited paper for *Frontiers in Psychology: Language Sciences* 9. Research Topic Approaches to Language: Data, Theory, and Explanation. https://doi.org/10.3389/fpsyg.2018.00155

Mel'čuk, Igor (1988) *Dependency Syntax: Theory and Practice*. Albany: State University of New York Press.

Montague, Richard (1970) Universal Grammar. *Theoria* 36: 373– 98.

(1973) The proper treatment of quantification in ordinary English. In Jaakko Hintikka, Julius Moravcsik & Patrick Suppes (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 & Michael O'Malley (1969) Multi-rooted vines in semantic representation. In Robert Binnick et al. (eds.) *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, Gereon (2011) *Constraints on Displacement: a phase-based approach*. Amsterdam: John Benjamins.

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 Sidewards 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.

Ore, Oysten (1990) *Graphs and their uses* [Edition revised and updated by Robin Wilson]. Providence, Rhode Island: the Mathematical Association of America.

Osborne, Timothy (2005) Beyond the Constituent: A Dependency Grammar Analysis of Chains. *Folia Linguistica XXXIX/3-4*. 251-297.

(2008) Major constituents: and two dependency grammar constraints on sharing in coordination. *Linguistics* 46(6). 1109‐1165.

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 & 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.

Panagiotidis, Phoevos (2001) *Pronouns, Clitics and Empty Nouns: 'Pronominality' and licensing in syntax*. Amsterdam: John Benjamins.

Parsons, Terence (1990) *Events in the Semantics of English. A Study in Subatomic Semantics.* Cambridge, Mass.: MIT Press.

Partee, Barbara (1974) Some transformational extensions of Montague grammar. Ms, MIT.

(1975) Montague Grammar and Transformational Grammar. *Linguistic Inquiry* 6(2). 203-300.

Perlmutter, David (1968) Deep and Surface Structure Constraints in Syntax. PhD Thesis, MIT.

(1978) Impersonal Passives and the Unaccusative Hypothesis. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society 38*. 157-189.

(1982) Syntactic representation, syntactic levels, and the notion of Subject. In Pauline Jacobson & Geoffrey Pullum (eds.) *The Nature of Syntactic Representation*. Dordrecht: Reidel. 283-340.

(1983) Introduction. In David Perlmutter (ed.) *Studies in Relational Grammar I.* Chicago: University of Chicago Press. ix-xv.

Perlmutter, David & 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 & 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 (1987) *Information-Based Syntax and Semantics, Volume 1: Fundamentals.* CSLI Lecture Notes no. 13. Stanford: CSLI.

(1994) *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press and Stanford: CSLI Publications.

Post, Emil (1943) Formal Reductions of the General Combinatorial Decision Problem. *American Journal of Mathematics*, 65(2). 197-215.

(1944) Recursively enumerable sets of positive integers and their decision problems. *Bulletin of the American Mathematical Society* 50. 284–316.

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.

(1971) *Crossover Phenomena*. New York: Holt, Reinhart & Winston.

(1972) On some rules that are not successive-cyclic. *Linguistic Inquiry*, 3(2). 211-222.

(1974) *On Rising*. Cambridge, Mass.: MIT Press.

(1982) Some Arc Pair Grammar descriptions. In Pauline Jacobson & Geoffrey Pullum (eds.) *The Nature of Syntactic Representation*. Dordrecht: Reidel. 341-426.

(1986) *Studies of Passive Clauses*. New York: SUNY Press.

(1998) *Three Investigations on Extraction*. Cambridge, Mass.: MIT Press.

(2004) *Skeptical Linguistic Essays*. Oxford: OUP.

(2010) *Edge-Based Clausal Syntax*. Cambridge, Mass.: MIT Press.

Potsdam, Eric & Jeffrey T. Runner (2001) Richard returns: Copy Raising and its implications. In *CLS 37: The main session*, eds. Mary Andronis, Chris Ball, Heidi Elston, and Sylvain Neuvel, Vol. 1. Chicago: Chicago Linguistic Society. 453-468.

Progovac, Ljiljana (1998) Structure for coordination. *Glot International* 3(7). 3–6.

Pullum, Geoffrey K. (2019) What grammars are, or ought to be. In Stefan Müller, Petya Osenova (eds.) *Proceedings of the 26th International Conference on Head-Driven Phrase Structure Grammar*. Stanford: CSLI Publications. 58–78.

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. & 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, 2099). Berlin: Springer Verlag. 17–43.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik (1985) *A Comprehensive Grammar of the English Language*. London: Longman.

Radford, Andrew (1997) *Syntax: A Minimalist Introduction*. Cambridge: CUP.

Ramchand, Gillian (2008) *Verb Meaning and the Lexicon*. Cambridge University Press.

Reinhart, Tanya (1976) *The Syntactic Domain of Anaphora*. PhD Dissertation, MIT.

(2006) *Interface Strategies: Optimal and Costly Computations*. Cambridge, Mass.: MIT Press.

Reuland, Eric & Tanya Reinhart (1993) Reflexivity. *Linguistic Inquiry*, 24(4). 657-720.

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.

Rivas, Alberto (1974) *A Theory of clitics*. PhD Dissertation, MIT.

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.

(2013) Locality. *Lingua* 130. 169-186.

Rogers, Andy (1971) Three kinds of physical perception Vs. *CLS* 7, 206-22.

(1972) Another look at Flip perception Vs. *CLS* 8, 303-15.

(1974) A Transderivational Constraint on Richard? *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*. 551-558.

Rogers, James (2003) Syntactic structures as multidimensional trees. *Research on Language and Computation*, 1(4). 265–305.

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) *Contraints on Variables in Syntax*. PhD Thesis, MIT.

(1968) Guess Who? In Robert I. Binnick, Alice Davison, Georgia M. Green, and Jerry L. Morgan (eds.) *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*. Chicago: University of Chicago. 252-286.

(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.

(1973) Slifting. In Maurice Gross, Morris Halle & Marcel Schützenberger (eds.) *The formal analysis of natural languages*. The Hague: Mouton. 133-169.

(1984) Inner islands. In Claudia Brugman, 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*. In Tista Bagchi, Katharine 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 (2018) Syntax and Uncertainty. In Angel Gallego & Roger Martin (eds.) *Language, Syntax, and the Natural Sciences*. Cambridge: CUP. 316-332.

Saddy, Douglas & Diego Gabriel Krivochen (2017) Classical Problems, non-Classical proposals. Ms., University of Reading.

Saddy, Douglas & 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.

Sarkar, Anoop & Aravind K. Joshi (1997) Handling coordination in a tree adjoining grammar. Technical report, University of Pennsylvania. https://www2.cs.sfu.ca/~anoop/papers/pdf/tag-coordination.pdf

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.

(1979) A Categorial analysis of Dyirbal ergativity. *Texas Linguistic Forum* 13. 96-112.

(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.

(2018a) *Sound and Grammar: Towards a Neo-Sapirian Theory of Language*. Leiden: Brill.

(2018b) Rhetorical meaning. *Linguistic Frontiers* 1(1). https://doi.org/10.2478/lf-2018-0001

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.

Shlonsky, Ur (1992) Resumptive pronouns as a last resort. *Linguistic Inquiry* 23. 443–468.

Shu, Kevin, Sharjeel Aziz, Vy-Luan Huynh, David Warrick & Matilde Marcolli (2018) Syntactic Phylogenetic Trees. In Joseph Kouneiher (ed.) *Foundations of Mathematics and Physics One Century After Hilbert*. Springer, Cham. 417-441.

Siloussar, Natalia (2014) What is and what is not problematic about the T-model. In Peter Kosta, Steven Franks, Teodora Radeva-Bork, & Lilia Schürcks (eds.) *Minimalism and Beyond: Radicalizing the Interfaces*. Amsterdam: John Benjamins. 350-362.

Siva, Karthik, Jim Tao & Matilde Marcolli (2017) Syntactic Parameters and Spin Glass Models Of Language Change. *Linguistic Analysis* 41(3-4).

Sloan, Kelly (1991) Quantifier-Wh Interaction. In Lisa Cheng and Hamida Demirdache (eds.) *More Papers on Wh- movement*, MITWPL 15. 219-237.

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.

Smolensky, Paul, Matthew Goldrick & Donald Mathis (2014) Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science*, 38. 1102–1138.

Sportiche, Dominique (1985) Remarks on Crossover. *Linguistic Inquiry*, 16(3). 460-469.

Stanley, Richard (1967) Phonological Redundancy Rules. *Language*, 43. 393-436.

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.

Stowell, Tim (1981) *Origins of Phrase Structure*. PhD dissertation, MIT.

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. [There is an English translation: *Elements of structural syntax*. Translated by Timothy Osborne and Sylvain Kahane. Amsterdam: John Benjamins, 2015.]

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.

Truswell, Robert (2014) Binding Theory. In Andrew Carnie, Yosuke Sato & Daniel Siddiqi (eds.) *The Routledge Handbook of Syntax*. Vol. 1. London: Routledge. 214-238.

Ura, Hiroyuki (2000) *Checking Theory and Grammatical Functions in Universal Grammar*. Oxford: OUP.

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.

(2011) A sketch of the grammar in non-classical conditions. Ms. UMD.

(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 & Julie Franck (2016) Semantic similarity effects on weak islands acceptability. *Research in Generative Grammar*, 38. 269-285.

Wall, Robert (1972) *Introduction to Mathematical Linguistics*. New York: Prentice Hall.

Wasow, Thomas (1979) *Anaphora in Generative Grammar.* Gent: E. Story-Scientia.

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

Wexler, Kenneth & Peter Culicover (1980) *Formal Principles of Language Acquisition*. Cambridge, Mass.: MIT Press.

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: Adison Wesley.

Wurmbrand, Susi (2011) Reverse Agree. Course notes from *Problems in Syntax* (Spring 2011) http://wurmbrand.uconn.edu/Papers/Agree-and-Merge.pdf

(2017) Stripping and topless complements. *Linguistic Inquiry*, 48(2). 341-366.

Zaenen, Annie (1983) On Syntactic Binding. *Linguistic Inquiry*, 14(3). 469-504.

Zaenen, Annie, Elisabet Engdahl & Joan M. Maling (1981) Resumptive pronouns can be syntactically bound. *Linguistic Inquiry*, 12(4). 679-682.

Zagona, Karen (2003) *The Syntax of Spanish*. Cambridge: CUP.

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

Zwicky, Arnold & Stephen Isard (1963) Some aspects of tree theory. Working Paper W-6674, The MITRE Corporation, Bedford, Mass. Available at: https://web.stanford.edu/~zwicky/some-aspects-of-tree-theory.pdf

Appendix A: On some inadequacies of phrase structure trees for representing *Merge*

In this appendix we will look at binary-branching, single-rooted, oriented, endocentric tree structures and their relation to the derivational operation that is supposed to have been responsible for generating them: Merge. In this respect, a rather surprising decision of MGG is to use trees of the kind specified above to represent bottom-up derivations built stepwise via Merge, and it is the purpose of this section to shed some light on why this choice is surprising. Of course, this is not so for most MGGers, and even recent revisions of some aspects of MGG syntax in graph-theoretic terms (like McKinney-Bock & Vergnaud, 2014) still assume that:

> *classical Phrase-markers seem to be the right objects to describe interpretive properties of expressions at the interface levels. At least, we shall assume as much.* (McKinney-Bock & Vergnaud, 2014: 218)

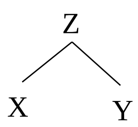
These authors prefer to multiply representations (*à la* Williams, 2003) rather than abandon binarity (i.e., combinatorics applies to 'pairs of formatives') and classical P-markers altogether (as we have done here) in order to circumvent some of the problems that Merge-based syntax has. In their view, '*Narrow syntax will be formalized as a graph in the general sense* [read: connected and directed, plus labelled edges indicating the kind of syntactic relation holding between nodes]. *Phrase-markers will be read from that graph, subject to various conditions*' (McKinney-Bock & Vergnaud, 2014: 218). The difference with our view should be evident without the need to introduce further details about their perspective: for us, there is a single kind of descriptive representation, maximally connected graphs. Furthermore, where McKinney-Bock & Vergnaud (2014: 220) '*consistently disregard the directed character of the graph*' (because it only serves the purpose to indicate projection), we use the directed character of graphs to indicate dominance relations, which are essential in our structural descriptions.

The previous paragraph was somewhat of a digression. The reason why we think that using binary branching diagrams to represent Merge is surprising is that in the traditional tree corresponding to the result of

265) Merge(X, Y)

Which is

265')



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 (265') is indeed suitable. Thus,

265") $Z \rightarrow X, Y$

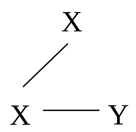
Is strongly equivalent to (265'): the PSR goes *from* higher nodes *to* lower nodes, and no direct connection between X and Y is ever suggested. But Merge is different: derivations are built from the bottom upwards, and the rule only makes direct reference to X and Y (Chomsky, 2015; Collins, 2017). Thus, we should have (266):

266)

X ——— Y

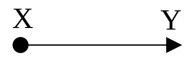
The structure generated by Merge is then labelled (either by ‘Minimal Search’, whatever that is, or via more classical procedures), with the label being either X or Y (Chomsky, 1994): Merge can *only* generate endocentric binary branching structures (Kayne, 1994 and much subsequent work). Take (168) as the input for a further application of Merge: how does the system proceed? If (266) needs to be labelled before entering further syntactic relations, the intermediate step between ‘first-Merge’ and ‘second-Merge’ (see, e.g., Zwart, 2009) is (266’):

266’)



McKinney-Bock & Vergnaud (2014: 221) propose a different alternative:

266”)



Which means ‘X merges with Y, and Y projects’. But of course, if a binary branching {Y {X, Y}} P-marker is to be read from (266”), an additional condition is needed. That additional condition is the following:

> ***Condition on Phrase-markers***
> *Let P be some classical Phrase-marker and let* $(\boldsymbol{f}_i, f_j)$, $(\boldsymbol{f}_i, f_k)$, $\boldsymbol{f}_i$, f_j, f_k *distinct formatives in P, be a pair of grammatical relations in P which share the formative* $\boldsymbol{f}_i$ *. At least one of the two relations is labeled/headed by* $\boldsymbol{f}_i$*.* (McKinney-Bock & Vergnaud, 2014: 215)

Classical P-markers are acyclic graphs which comply with the condition above. All in all, the graphs in McKinney-Bock & Vergnaud (2014) (see also McKinney-Bock, 2013) are *isomorphic* to orthodox Minimalist binary-branching trees, in the technical sense: relations (adjacency and labelling) are preserved within selected structure, and it is possible to go from the graph-level to the ‘classical P-marker’ level and back; the function relating these admits an inverse (see Gould, 1988: 5; Wilson, 1996: 9). In what follows, we will deal only with what McKinney-Bock & Vergnaud refer to as ‘classical phrase markers’.

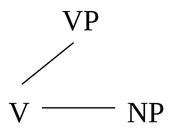
Let us go back to the abstract structure (266’). Concretely, assume that X = V and Y = NP (given that Merge of {H, XP} is ‘virtually everything’[63], according to Chomsky, 2009: 52; see also Chomsky, 2013, 2015). Then, what we have is (267):

[63] 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 the expression ‘virtually everything’ in Chomsky’s claims (see, e.g., Chomsky, 2009: 52 for two uses of this expression in the same page). 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

267)



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/*v*Ps) 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 summarised (but not really made justice to) in the following equivalence in both strong and weak generative capacity:

268)

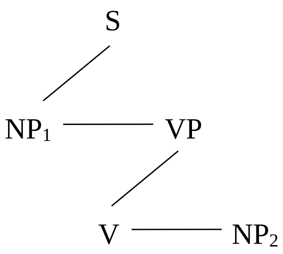
Bare Phrase Structure   Dependency tree

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 (268) 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:

of person argues against something that is 'necessary' or 'everything' (regardless of the 'virtual' character of that necessity)?

269)

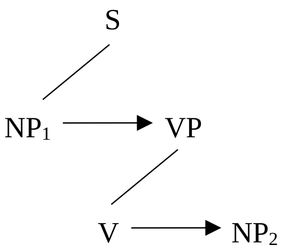


Note that walking the path from NP_2 to S gives us all the Merge relations: horizontal lines translate as ‘merges to’, whereas diagonal lines translate into ‘projects as’. Regardless of how we diagram it, be it horizontal and diagonal lines, or solid and dotted lines, or whatever else, there is a fundamental difference between ‘merges with’ and ‘projects as’. Thus, (269) reads as:

269’) V merges to NP_2 and they project a VP. NP_1 merges to VP and they project an S.

But the graphs in (267’ - 269) 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 (269) becomes (270) (a P-marker that has quite a lot in common –in terms of *strong generative power*- with those proposed in McKinney-Bock & Vergnaud, 2013):

270)



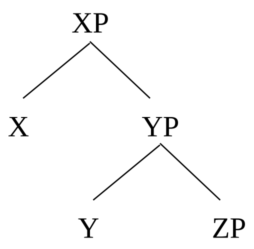
Walking from S to NP_2 follows a unique, unambiguous path in which the relevant relations are indeed preserved. In this context, some more graph theory is useful. A connected graph *G* is said to contain a *Hamilton* (or *Hamiltonian*) *path* if there is a path that visits every vertex of *G* exactly once (Van Steen, 2010: 92). In this sense, (270) 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 (271):

271)

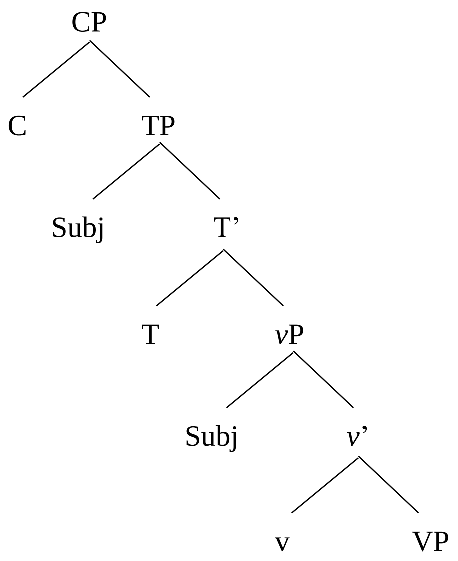


As it is, (271) contains *two* local Hamilton paths:

272) 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 (272) arises, with irrelevant differences, in the C-T-*v* skeleton assumed in MGG (but *not* in its radically label-less variants, we hasten to add):

272')



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 unwarranted *ad hoc* stipulations): if probing is some kind of searching mechanism, it will follow the walk:

273) 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

273') 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:

274) C → CP → TP → $\text{Subj}_{[NO]}$ → TP → T' → $\text{T}_{[NO]}$ → …

We have indicated the nodes at which failed attempts are recognised as such by means of the not-so-subtle diacritic [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

275) X → Y → Z

Holds, and *not any other walk*: linearisation can proceed only if the sequence is *unambiguous*.

Under a Kaynean/Chomskyan view of phrase structure (also assumed in McKinney-Bock & Vergnaud, 2014: 219, which artificially constrains the class of possible phrase markers), in a phrase marker like (170), ZP *must* be an empty category (i.e., a category with no phonological content) for linearization purposes. This is so because the Linear Correspondence Axiom maps asymmetric c-command into precedence. Wherever two nodes (both terminals or both nonterminals) c-command each other, that is called a *symmetry point*, and must be broken via movement of either ('either' in principle; the rightmost in practice, as in Moro, 2000). Thus, at the relevant level of representation or derivational point, ZP is not there to be walked on. In Phase Theory, given a derivational rhythm Phase-Non Phase-Phase… (e.g., Richards, 2011; Boeckx, 2012: 56) -which is claimed to be a 'natural' emergent in the 'narrow syntax', in line with the rhetoric of 'design perfection' of the Language Faculty pushed by Chomsky-, ZP is not accessible by operations at X, which -relevantly- we take to mean that there is no walk 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(ly governed) 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
(descriptive) transformations	Agree (feature checking / valuation / donation / sharing…)
the Complex NP Constraint	phrase structure rules
	(Relativised) 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 as in Chomsky, 1965). 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, the CNPC, the 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, in the present theory, 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 Relativised Minimality if conceptualised 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 Relativised 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* (see, e.g., Bresnan, 2001: 91; also Dalrymple et al., 2015 for detailed discussion) further restricts the possibility of having base-generated empty nodes (essentially, any node which dominates only the null expression Ø can -and *must*- be *pruned* in the sense of Ross, 1969b). 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 the model or not.