ITALIAN POSTVERBAL SUBJECTS FROM A MINIMALIST PARSING PERSPECTIVE

ANIELLO DE SANTO

ABSTRACT: Stabler (2013)'s parser for Minimalist grammars has been shown to successfully predict various off-line processing preferences, by exploiting complexity metrics connecting syntactic structure to memory load. This approach provides a quantifiable way to test the effects of structural hypotheses on linguistic behavior, and thus can help bridge syntactic theory and processing phenomena. This paper extends the empirical coverage of the model by looking at the processing of Italian postverbal subjects, and it discusses the relevance of transparent computational models for existing approaches to sentence processing.

KEYWORDS: Minimalist grammars, parsing, Italian, syntax, memory.

1. INTRODUCTION[1]

An important problem at the intersection between theoretical linguistics and psycholinguistics is whether the fine-grained grammatical analyses posited by syntactitians have any relevance to the cognitive processes underlying language processing (Miller & Chomsky 1963; Bresnan 1978). This paper follows a line of research recasting such question in a computational framework, by specifying a *transparent* (i.e., interpretable) *linking hypothesis* between grammatical structure and processing complexity.

Recent work has shown that a top-down parser for Minimalist grammars (MGs; Stabler 1996, 2013) successfully models off-line sentence processing preferences across a variety of phenomena cross-linguistically (Kobele *et al.* 2013; Gerth 2015; Graf *et al.* 2017: a.o.). This model adopts a fully formalized theory of grammatical structures (MGs), an algorithm detailing how such structures are built over time, and an explicit theory of how structure-building operations affect cognitive load – as a vast set of complexity metrics measuring memory usage. While the variety of constructions modeled so far in the literature is encouraging, extending the range of phenomena that the parser correctly accounts for is still crucial to confirm the cognitive plausibility of the approach.

---

[1] The author would like to thank the two anonymous reviewers for helping improve the clarity of this paper, as well as John Baylin, Thomas Graf, Jon Sprouse, and the audience at SCiL 2019 for comments on previous versions of this work.

In this paper, I evaluate the MG parser's performance on the processing asymmetries reported for Italian postverbal subject constructions. Postverbal subjects in Romance languages have been object of extensive study both in the theoretical syntax and in the psycholinguistic literature (Cardinaletti 1998; De Vincenzi 1991; Cardinaletti 2004; Belletti & Contemori 2009; Arosio *et al.* 2017; De Santo 2019). Thus, they make for an ideal testing ground to evaluate the MG model's ability to account for processing contrasts *just* in terms of *structural complexity*. Moreover, the existing variety of competing syntactic analyses for the Italian constructions will allow for a careful evaluation of how fine-grained grammatical assumptions can affect processing predictions. By clarifying which aspects of sentence structure – as hypothesized by different grammatical theories – modulate processing difficulty, this paper aims to highlight how transparent computational models can strengthen the connection between experimental results and syntactic theory.

2. MINIMALIST GRAMMARS AND SENTENCE PROCESSING

Before approaching the Italian data, it is important to understand the core ideas behind the computational framework adopted in this paper. This section summarizes work on an interpretable computational parsing model, which can be used to explore how off-line sentence processing profiles are modulated by the rich structural hypotheses of the most recent version of Chomsky's transformational grammar. I discuss the intuitions behind the choice of grammatical representations adopted by the model, and the way the parser's tree traversal strategy affects complexity metrics indexing memory load. In doing so, I review a series of results showing the validity of the approach, in term of coverage for a variety of psycholinguistic phenomena. For a broader discussion of how this particular model situates itself in the ongoing debates about the relation of grammatical theories, parsing, and processing mechanisms, I refer the reader to (Rambow & Joshi 1994; Gerth 2015; Graf *et al.* 2017; De Santo 2020b), and references therein.

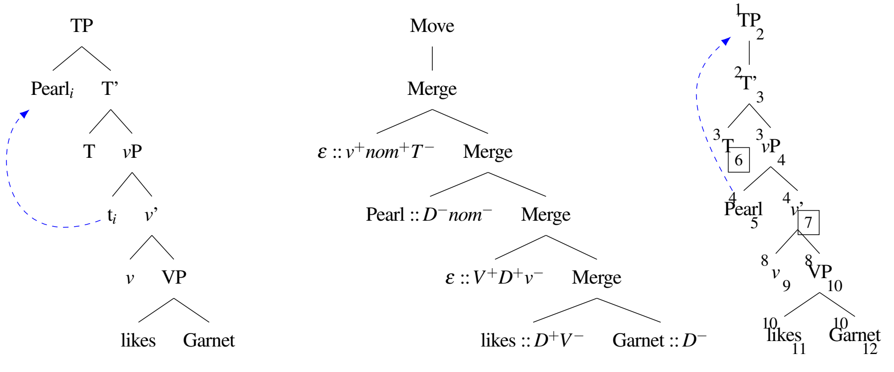
2.1 Minimalist Grammars

MGs (Stabler 1996, 2011) are a lexicalized formalism incorporating the structurally rich analyses of the earliest versions of Minimalist syntax. An MG is a set of lexical items (LIs) consisting of a phonetic form and a finite, non-empty string of features. Syntactic objects are built from LIs via two feature checking operations: *Merge* – encoding subcategorization – and *Move* – allowing for long-distance movement dependencies. The fundamental data structure in MGs is a *derivation tree*, which encodes the sequence of Merge and Move operations required to build the phrase structure tree for a given sentence (Michaelis 1998; Harkema 2001; Kobele *et al.*

2007). For instance, Figure 1a and Figure 1b compare these two kind of trees for a simplified analysis of the sentence *Pearl likes Garnet*.

In a derivation tree, all leaf nodes are labeled by LIs, while unary and binary branching nodes are labeled as Move or Merge, respectively. Crucially, the main difference between the phrase structure tree and the derivation tree is that in the latter, moving phrases remain in their base position, and their landing site can be deterministically reconstructed via the feature calculus. Thus, the final word order of a sentence is not directly reflected in the order of the leaf nodes in a derivation tree.

Importantly, MG derivation trees form a regular tree language, and thus – modulo a more complex mapping from trees to strings – allow us to exploit simple variants of established parsing algorithms for context-free grammars (CFG).



(a) (b) (c)

Figure 1: Phrase structure tree *(a)*, MG derivation tree *(b)*, and annotated derivation tree *(c)* for *Pearl likes Garnet*. Boxed nodes in *(c)* are those with tenure value greater than 2, following (Graf & Marcinek 2014).

2.2 MG Parsing

This paper adopts Stabler (2013)'s MG variant of a standard recursive-descent parser for CFGs. This parser takes as input a sentence represented as string of words, hypothesizes the structure *left-to-right*, *depth-first*, *top-down*, verifies that the words in the structure match the input string, and outputs a tree encoding of the sentence structure. However, due to the fact that in a derivation tree the order of lexical items does not fully match the linear surface order, simple left-to-right scanning of the leaf nodes yields the wrong word order. Thus, the MG variant must also keep track of the derivational operations affecting the linear word order.

Without too many technical details, the parsing procedure can be outlined slightly more clearly as follows: I) hypothesize the top of structure and add nodes downward (toward words) and left-to-right; II) if *move* is predicted, it triggers the search for mover ⇒ build the shortest path towards predicted mover; III) once the mover has been found, continue from the point where it was predicted (Kobele *et al.* 2013). The step in (II) makes this a *string driven* recursive descent strategy. Memory mechanisms are essential to this procedure: if a node is hypothesized at step i, but cannot be worked on until step j, it must be stored for $j-i$ steps in a priority queue.

To make the traversal strategy easy to follow, I adopt Kobele *et al.* (2013)'s tree annotation approach. The annotation indicates for each node in the tree when it is first conjectured by the parser (*index*, superscript) and placed in the memory queue, and at what point it is considered completed and flushed from memory (*outdex*, subscript). Since the details of the feature calculus are mostly irrelevant to the memory metrics adopted later on in the paper, here I also rely on a simplified version of derivation trees, discarding the features of each LI, and labelling internal nodes as standard in minimalist syntax (Figure 1c). While not technically part of the representation, dashed arrows are included to make movement relations explicit. Note, however, that intermediate movement steps are not marked by arrows, since intermediate landing sites do not affect the traversal strategy.

Finally, Stabler's original parser is equipped with a search beam discarding the most unlikely predictions. Consistently with previous work, I follow Kobele *et al.* (2013) in ignoring the beam and assuming that the parser is equipped with a perfect oracle, which always makes the right choices when constructing a tree. This idealization is clearly implausible from a psycholinguistic point of view. However, it is made with a precise purpose in mind: to ignore the cost of choosing among several possible predictions and, by assuming a deterministic parse, to focus on exclusively evaluating contribution of syntactic complexity to processing difficulty without concern for (local) ambiguities (lexical or structural). As a corollary of this modeling commitment, the present paper focuses on evaluating *off-line* processing asymmetries: differences in complexity profiles as registered over a whole sentence, instead of word-by-word predictions.

2.3 Complexity Metrics

In order to allow for psycholinguistic predictions, the behavior of the parser must be related to processing difficulty via a *linking theory*. Adopting Stabler's MG parser allows us to be explicit about the nature of the structures being built, and about the time-course of the structure building operations connecting linear input to hierarchical representations. What remains to be specified then, is a psychologically reasonable theory of how cognitive resources are *linked* to parsing operations to derive measures of cognitive load.

There are, of course, several ways in which the relation between grammatical theories and processing mechanisms can be specified (the reader is referred to

Berwick & Weinberg 1982, 1983; Stabler 1984; Berwick & Weinberg 1985: for a classic example of such discussions). For instance, a common assumption in the early days of transformational grammar was that grammatical principles should guide processing strategies directly, with parsing mechanisms somehow mirroring the rules of the grammar. This is what Berwick & Weinberg (1983) refer to as the *Type Transparency Hypothesis* which, in its strongest interpretation, demanded a direct relation between "*the theoretical objects of grammar and those of parsing*". Without committing to an ontological distinction between grammatical and parsing objects, and even considering the drastic changes undergone by generative theories of syntactic representations during the years, Berwick & Weinberg (1983)'s fundamental questions about the *cost of mental computations* remain relevant (Phillips 2003).

In this sense, the approach I build on in this paper follows the ideas of Hale (2001) in adopting a framework – based on a weak version of the Derivational Theory of Complexity (Miller & Chomsky 1963) and more in line with a lexicalized syntax driven by Merge and Move operations – in which a computational cost is not associated with single grammar rules directly, but with parsing operations building the surface structure. The cost of one grammatical "step" can thus be spread during the processing phase upon different parsing operations.

Specifically, the linking theory adopted in this paper takes the form of complexity metrics that predict *off-line* processing difficulty based on how the geometry of the trees affects memory usage during a parse (Rambow & Joshi 1994; Gibson 2000; Kobele *et al.* 2013; Graf & Marcinek 2014; Gerth 2015). This link is *transparent*, in that a metric's value for a specific derivation tree can always be fully reconstructed given the geometry of that tree, the linear representation of a sentence, and knowledge about the parsing strategy. For a deeper discussion of different ways of approaching the relation between grammatical representations and processing behavior, and the particular way the model in this paper fits in this long-standing debate, the reader is referred to Vasishth & Lewis (2006) and De Santo (2020b: Chpt. 2), respectively.

The *MG model* distinguishes several cognitive notions of memory usage (Graf *et al.* 2017). Two of those are particularly relevant for the sake of this paper: I) how long a node is kept in memory (*tenure*); II) how much memory a node consumes (*size*). Tenure for each node *n* in the tree can be easily computed via the node annotation schema of Kobele *et al.*: a node's tenure is equal to the difference between its index and its outdex. Introducing size in an informal way is slightly trickier, as its formal definition is based on how information about movers is stored by Stabler's top-down parser (for a technical discussion, see Graf *et al.* 2015). In practice, size encodes how many nodes in a derivation consume more memory because a certain phrase *n* moves across them, and it is thus sensitive to the *hierarchical* distance between the filler and the gap. Procedurally, the size of the parse item corresponding to a moving node *n* can be computed, exploiting our simplified derivation trees, as the index of the mover n minus the index of its target site m. For example, referring to the annotated tree in Figure 1c, the size of *Pearl* is 3.

These broad notions of memory can then be used to define a vast set of complexity metrics measuring processing difficulty over a full derivation tree. Kobele *et al.* (2013) show that tenure can be associated to quantitative values by defining metrics like MaxT $:= max(\{tenure\text{-}of(n)\})$ and SumT $:= \sum_n tenure\text{-}of(n)$. MaxT measures the maximum amount of time any node stays in memory during processing, while SumT measures the overall amount of memory usage for all nodes whose tenure is not trivial (i.e., >2). It thus captures total memory usage over the course of a parse. Building on these findings, Graf & Marcinek (2014) show that MaxT (restricted to pronounced nodes) makes the right difficulty predictions for several phenomena, such as right embedding vs. center embedding, nested dependencies vs. crossing dependencies, as well as a set of contrasts involving relative clauses.

Extending Graf & Marcinek (2014)'s analysis of relative clause constructions, Graf *et al.* (2015) argue for the insufficiency of MaxT and introduce several new metrics. For example, they define an equivalent of SumT for size, measuring the overall cost of maintaining long-distance filler-gap dependencies over a derivation. Let M be the set of all nodes of derivation tree t that are the root of a subtree undergoing movement. For each $m \in M$, $i(m)$ is the index of m and $f(m)$ is the index of the highest Move node that m's subtree is moved to. SumS is defined as $\sum_{m \in M} i(m) - f(m)$. Graf *et al.* (2015) also introduce the idea of ranked metrics of the type $\langle M_1,...,M_n \rangle$, similar to constraint ranking in Optimality Theory (Prince & Smolensky 2008): a lower ranked metric matters only if all higher ranked metric have failed to pick out a unique winner (e.g., if two constructions result in a *tie* over MaxT).

Due the large number of metrics generated by this ranking approach, the evaluation space might seem too vast to give us any fruitful insights. However, as for most computational models, the issue of potential empirical indeterminacy is addressed by rigorously searching for a restricted set of metrics that account for a variety of diverse phenomena across languages. In this sense, previous work has ruled out the vast majority of the existing metrics, by showing their insufficiency in accounting for some crucial constructions across a variety of possible grammatical analyses (De Santo 2020b). Surveying the variety of previously modeled phenomena, a ranked combination of MaxT and SumS is supported by recent work on several different constructions cross-linguistically (Graf *et al.* 2017; Liu 2018; Lee 2018; De Santo 2019, 2020a). Building on previous work then, in the rest of this paper I will focus on the individual performance of MaxT and SumS on the processing phenomena of interest.

3. MODELING ITALIAN POSTVERBAL CONSTRUCTIONS

With our understanding of the computational model solidly in place, this section reviews the psycholinguistics literature on the processing of Italian postverbal subject, and presents the test cases modeled in the rest of the paper. I also discuss several modeling choices in respect to the syntactic analyses of the constructions under study. A detailed analysis of the modeling results is then the focus of Section 4.

3.1 Processing Asymmetries

While it is generally accepted that standard Italian's basic word order is Subject-Verb-Object (SVO), the variety of non-canonical word order constructions available in the language has been object of extensive study through the years. In particular, here I am interested in the off-line processing profiles associated with *postverbal* subjects in declarative clauses, usually classified as "free" inversion (in contrast with the "obligatory" type of subject inversion observed, for example, in wh-questions) and tied to the information structure of a clause (Longobardi 2000; Belletti 2004; Belletti & Leonini 2004; Cardinaletti 2004; Samek-Lodovici 2015; Leonetti 2018; Bianchi *et al.* 2017; Cardinaletti 2018). Consider Italian declarative sentences like in (1).

(1) *Ha chiamato Gio*
Has-SG called Gio
a. "He/she/it called Gio" **SVO**
b. "Gio called" **VS**

Without contextual/discourse cues, sentences like (1) are structurally ambiguous between a null-subject interpretation (1a) and a postverbal subject one (1b), with a marked processing preference for (1a) as compared to (1b) (De Vincenzi 1991: a.o.). This type of preference is also known to be modulated by properties of the verb as, for instance, its argument structure. Consider the sentences below, with (2) carrying an *unaccusative* verb and (3) an *unergative* one.

(2) *È arrivato Gio*
Is-SG arrived Gio
"Gio arrived" VS UNACCUSATIVE

(3) *Ha corso Gio*
Has-SG ran Gio
"Gio ran" VS UNERGATIVE

While on the surface these sentences look very similar, they differ in the underlying organization of the argument structure of the verb (Belletti 1988). Importantly, De Vincenzi (1991) reports faster reading times and higher comprehension accuracy for (2) over (3) (see also Greco *et al.* 2020: a.o.).

Additionally, postverbal subject constructions interact in interesting ways with the processing of *restrictive relative clauses* (RCs), which in Italian have been the focus of extensive experimental studies from the perspective of comprehension (Volpato & Adani 2009), production (Belletti & Contemori 2009), and (L1 and L2) acquisition (Friedmann *et al.* 2009; Volpato 2010: a.o.). Italian speakers conform to the general cross-linguistic preference for subject over object RCs (Frauenfelder *et al.* 1980; King & Kutas 1995; Schriefers *et al.* 1995; Lau & Tanaka 2021: a.o.), so that (4) is easier to process than (5):

(4) *Il cavallo che ha inseguito i leoni*
The horse-SG.M that has-SG chased the lions-PL.M
"The horse that chased the lions" SRC

(5) *Il cavallo che i leoni hanno inseguito*
The horse-SG.M that the lions-PL.M have-PL chased
"The horse that the lions chased" ORC

However, Italian also allows for sentences like (6), ambiguous between a SRC interpretation (6a) and an ORC interpretation (6b) with the embedded subject expressed postverbally (ORCp):

(6) *Il cavallo che ha inseguito il leone*
The horse-SG.M that has-SG.M chased the lion-SG.M
a. "The horse that chased the lion" SRC
b. "The horse that the lion chased" ORCp

In these ambiguous cases, native speakers show a marked preference for the SRC interpretation over the ORCp one. As Italian is morphologically rich, sentences like (6) can also be disambiguated by grammatical cues like subject-verb agreement. For instance, in (7) the DP *i leoni* is plural, while the DP *il cavallo* is singular.

(7) *Il cavallo che hanno inseguito i leoni*
The horse-SG.M that have-PL.M chased the lions-PL.M
"The horse that the lions chased" ORCp

Since the verb agrees in number with its subject, and in this case the embedded verb is marked for plurality, (7) can only be interpreted as an ORCp construction. Crucially, even in unambiguous cases studies report increased efforts with ORCps (Utzeri 2007: a.o.), leading to the following difficulty gradient: SRC $<$ ORC $<$ ORCp (where $x < y$ is henceforth used to convey that x is preferred over y).

The cross-linguistic contrast between SRCs and ORCs has been well studied in the past (see Lau & Tanaka 2021: for a recent critical review), and it is compatible with a variety of theories of processing difficulty, such as surprisal (Levy 2013), cue-based memory retrieval (Lewis & Vasishth 2005), the active filler strategy (Frazier 1987), the Dependency Locality Theory (Gibson 1998, 2000), the Competition Model (Bates & MacWhinney 1987), the Minimal Chain Principle (De Vincenzi 1991), featural Relativized Minimality (Rizzi 1990; Friedmann *et al.* 2009; Contemori & Belletti 2014), among many. The increased complexity reported for ORCs with postverbal subjects is more of a challenge to some of these models (e.g., for the Competition model and Dependency Locality Theory; Arosio *et al.* 2009), as (on the surface) the gap between the moved object and its base position is identical in both configurations. However, the processing profile of postverbal matrix constructions and ORCps can be accounted for by a variety of processing models (De Vincenzi 1991; Friedmann *et al.* 2009; Adani 2011; Arosio *et al.* 2017). For instance, the increased difficulty of ORCps could be explained in terms of *economy* of gap prediction and cost of structural re-analysis, due to the possible ambiguity at the embedded subject site – where the parser has the choice of either postulating a null pronominal subject or establishing a filler-gap dependency. Crucially though, such an account has to come with extra assumptions about *why*, when building these dependencies, it is preferable to choose one strategy over the other.

Importantly, the aim of this paper is not to argue for the correctness (or lack thereof) of any of these accounts per se. The goal here is to instead evaluate these results through the lens of the MG parsing model. Adopting an explicit computational model forces us to intentionally commit to detailed syntactic representations. Thus, this approach allows us to explore the repercussion of fine-grained syntactic choices towards our psycholinguistic predictions/explanations.

Moreover, the MG parser has already been successful in accounting for RC asymmetries cross-linguistically (Graf *et al.* 2017; Zhang 2017). Thus, the interaction of post-verbal constructions and RCs in Italian is the perfect next step in expanding the empirical coverage of the model, allowing us to build on the insights provided by previous work while incrementally exploring new structural configurations. Section 5 will then discuss whether/how adopting this kind of model also highlights the contributions/limitations of existing theoretical accounts of processing difficulty.

3.2 Syntactic Assumptions

The discussion in Section 2 should have made it clear that, in order to explore how different aspects of sentence structure drive processing cost, the MG model fully commits to structural details. The choice of a syntactic analysis is then particularly important. In what follows, I highlight the main choices made for the patterns under

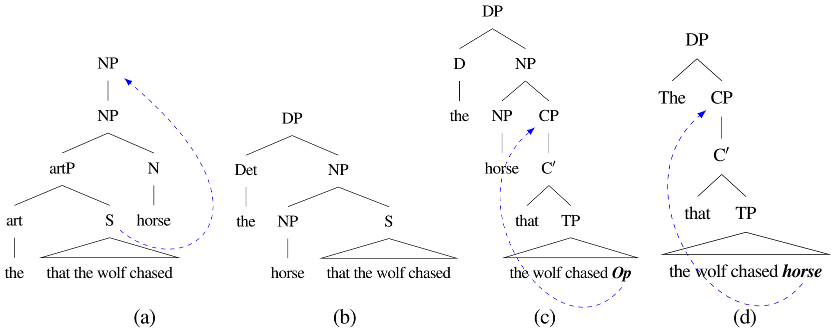


FIGURE 2: SKETCHES OF THE *(a)* EXTRAPOSITION, *(b)* DP, *(c)* WH-MOVEMENT, AND *(d)* PROMOTION ANALYSES FOR THE OBJECT RELATIVE CLAUSE *The horse$_i$ [that the wolf chased* t$_i$*].*

study: the structure of relative clauses, and the analysis of Italian postverbal subjects.

Structure of the Relative Clause

Building on Zhang (2017) and De Santo (2019), here I consider four distinct analyses of relative clauses: the Extraposition analysis (Chomsky 1965), the DP analysis (Abney 1987), the Wh-movement analysis (Chomsky 1977), and finally the Promotion analysis (Kayne 1994).

While some of couple of these approaches are fairly outdates, and there is an abundance of theoretical discussions on the advantages of some analyses over others (Bianchi 2002a,b), this set is representative of a broad range of theoretical accounts differing from each other both in terms of major structural configurations (e.g., the relation between RC and modified DP; whether the RC is an adjunct or an argument), and in terms of more subtle details (e.g., whether the RC head is base generated internally or externally to the RC; the specific landing sites for movement). These aspects are crucial, as they will lead to movement dependencies varying in direction (left vs. right) and distance (short vs. long) – thus potentially impacting memory load in fundamentally different ways.

Extraposition Analysis - In Chomsky (1965)'s Extraposition approach, the RC is a complement of the article head, which projects an article phrase then selected by the N head. This derivation would not yield the correct word order for languages like Italian, which see the RC clause appear post-nominally in the surface order. Thus, the RC needs to undergo extraposition to the right of the relativized noun (Figure 2a).

DP Analysis - I use the DP analysis as an instance of a broad group of analyses that treat the RC as adjoining to a certain projection, with no internal movement inside the RC itself. Following Abney (1987), here the determiner heads its own projection and takes as its complement an NP with a right-adjoined RC (Figure 2b).

Wh-movement Analysis - Chomsky (1977)'s Wh-movement analysis treats the construction of an RC as an instance of wh-movement. This analysis was initially proposed to construct wh-relatives, but it can be easily applied to that-clauses in English. The complementizer position is overtly filled by *that*, while a silent wh-operator *Op* moves from the base position to Spec,CP. Then the whole CP merges with the relativized NP as its adjunct (Figure 2c). The silent *Op* is co-indexed with the NP to which the RC is adjoining. This is the approach adopted here for Italian.

Promotion Analysis - Lastly, I consider a Promotion analysis (Kayne 1994). In this case, the head noun starts out as an argument of the embedded verb and undergoes movement into the specifier of the RC. The RC itself is then selected by the determiner that would normally select the head noun in head-external accounts, like the wh-movement cases above (Figure 2d).

Postverbal Subjects

Consider now the following declarative clause with a postverbal subject:

(8) *Inseguono*	*il cavallo*	*i leoni*	
Chase	the horse	the lions	

"The lions chase the horse"

The specific structure of free inversion constructions in Romance languages has been (and is) topic of extensive debate – especially due to the complex interaction between the syntactic properties of postverbal subjects and the information structure of the clause they appear in (Antinucci & Cinque 1977; Longobardi 2000; Samek-Lodovici 2015; Leonetti 2018; Cardinaletti 2018: a.o.). An in-depth discussion of the literature on this topic is beyond the scope of this paper. Recall however that our goal is to explore the consequences of syntactic decisions on the predictions of the MG model. Thus, as in the RC case, I focus on two popular approaches that substantially differ in their structural assumptions.

Smuggling Approach - First, I consider an analysis of postverbal constructions due to Belletti & Leonini (2004: a.o.) – which in the past has been often referenced in the psycholinguistic literature (Arosio *et al.* 2017: a.o.). According to Belletti & Leonini, in postverbal constructions the subject DP (*[i leoni]*) is merged in preverbal subject position Spec,*v*P, and then raised to a Spec,FocP position in the clause-internal *v*P periphery. The whole verbal cluster is raised to a clause-internal Spec,TopP position; and an expletive *pro* is base generated in Spec,TP and co-indexed with the postverbal subject (Figure 3a).[2]

Scrambling Approach - A different approach to the sentence in (8) is to assume that the subject does not move, but instead remains *in situ*. A variety of proposals in this direction assume that the postverbal subject is in its base position (e.g., in Spec,*v*P) as the argument of the verb. The VOS order then is derived by head movement of the verb to an aspectual projection (Cinque 1999), and by leftward *scrambling* of the object to a position above the subject but below the landing position of the verb (Ordóñez 1998; Cardinaletti 1998; Brunetti 2003; Cardinaletti 2004; Bocci 2013; Samek-Lodovici 2015; Cardinaletti 2018). The specific label for the projection the object lands into varies depending on the analysis (cf. Ordóñez 1998; Cardinaletti 1998), and it is irrelevant for our purposes (Figure 3b).

4. MODELING RESULTS

With all preliminaries in place, we can now move to modeling the Italian processing asymmetries with the MG parser, following the approach detailed in Section 2.[3] In particular, derivations for each test sentence are fed to the parser, together with the processing contrasts reported by the psycholinguistic literature – reframed in terms of pairwise comparisons. As outlined in Section 3, I consider the following processing contrasts (where $x < y$ stands for x being preferred over y):

Postverbal Subjects in Matrix Clauses

- SVO < VS (1a) < (1b)
- unaccusative < unergative (2) < (3)

Postverbal Subjects in RCs

- SRC < ORC (4) < (5)

---

[2] Technically, Belletti & Leonini (2004) assume that VP, not *v*P, raises to Spec,TopP. This follows from the authors adopting Collins (2005)'s smuggling analysis of passives directly. However, if we follow the traditional view of active verbs moving out of their base position to adjoin to little *v*, this analysis cannot hold as it would derive the wrong word order. Thus, I raise the whole *v*P cluster to TopP. This also seems to be in the spirit of what suggested by Belletti & Contemori (2009). But note that the modeling results in the following section would remain mostly unchanged even if we were to leave the *v*P shell in its base position, while VP raises above.

[3] All simulations in this paper were run on the open source code made available by Graf *et al.* (2017) at https://github.com/CompLab-StonyBrook/mgproc.

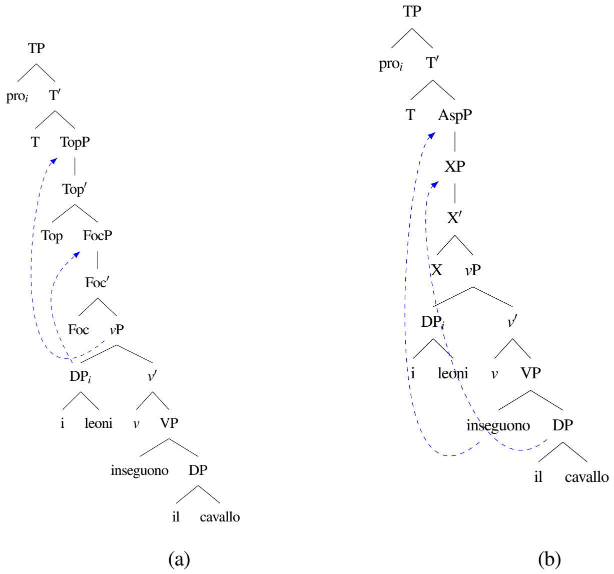


FIGURE 3: A SMUGGLING
(A) AND SCRAMBLING (B) ANALYSIS FOR THE POSTVERBAL CONSTRUCTION IN (8).

- SRC $<$ ORCp (4) $<$ (7)
- ORC $<$ ORCp (5) $<$ (7)

Derivations are built for each example sentence in Section 3, modulated across the syntactic choices discussed before.[4]

In order to derive processing predictions, the parser is then equipped with the all complexity metrics defined by Graf *et al.* (2017). However, since the relationship between complexity metrics and the structure of a specific derivation tree is subtle, in what follows I focus exclusively on two metrics that have been noted in previous studies as consistent predictors of processing difficulty: MAXT and SUMS. Given their previous success in accounting for off-line preferences cross-linguistically, focusing on MAXT and SUMS also furthers the goal of reducing empirical indeterminacy potential to every computational model, by evaluating a subset of metrics

---

[4] It is always possible to add extra dimensions of syntactic variation, of course. For instance, under a smuggling approach the Top and Foc projections expand on whichever "base" *v*P clause structure one assumes. We could then follow approaches that first require an Aspectual projection immediately above *v*P (Kempchinsky 2000; Borer 1994). However, since this would be done for all three sentences under consideration (not just postverbal ones) and since the modeling results crucially rest on a cross sentence (not cross analysis) comparison, the general conclusion of the model would not be affected (though we would obviously see a small change in the individual numerical scores). Similar considerations extend to any type of additional levels of projection that could be postulated for the structures adopted here.

that is limited and cognitively plausible.[5] As mentioned before, the metrics are not sensitive to lexical information. Thus, the model leads to strict categorical results, and one test sentence per construction is sufficient to conduct the evaluation.

4.1 Postverbal Subjects in Matrix Clauses

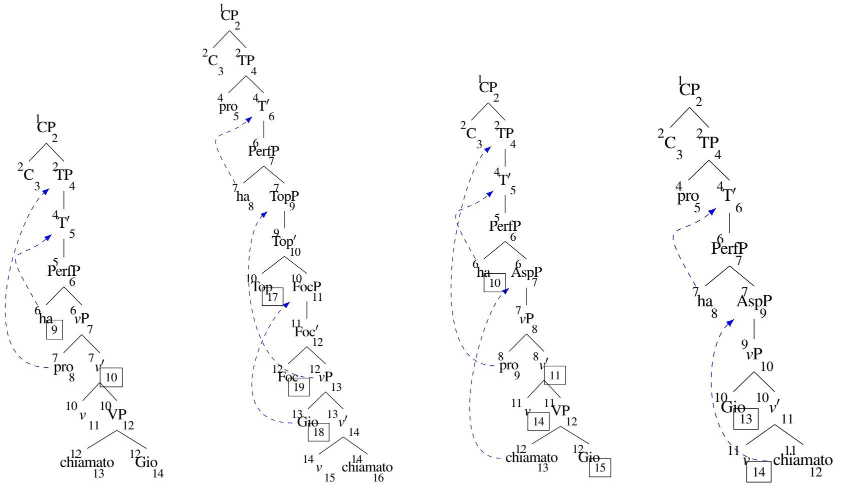
In order to understand the complexity of the grammatical assumptions made for the postverbal subjects, we can first look at processing asymmetries of postverbal constructions outside of RC environments (Table 1).

Recall that, when considering a matrix clause ambiguous between an SVO and VS interpretation (as in (1)), we expect a marked processing preference for SVO < VS (De Vincenzi 1991). As summarized in Table 1, both MAXT and SUMS predict the correct preferences under Belletti & Leonini (2004)'s smuggling analysis, as the Top and Foc heads have to wait for the whole *v*P to be found, before they can be discharged from memory themselves (Figure 4). However, neither metric is able to account for the correct preference under a scrambling analysis – in fact, there is *no* metric among the ones defined by Graf *et al.* (2017) that is able to capture this contrast given the scrambling assumption. Specifically, the parser makes a prediction that is the opposite of what expected, with the VS structure inducing lower memory load than the SVO one. To understand why that is, consider MAXT (Table 2). In both SVO and VS, the verb needs to raise to AspP, leading to tenure on *v* increasing to 3 $(14-11)$. Importantly though, in the SVO case the null subject (*pro*) needs to raise to Spec,TP, thus increasing tenure on *ha*. In the VS case the subject can stay *in situ*, and since there is no object movement, nothing else affects tenure significantly.

Let us now turn our attention to declarative sentences containing intransitive verbs of two classes: unaccusatives (2) and unergatives (3). The desired contrast is unaccusatives < unergatives (De Vincenzi 1991). While on the surface these sentences look very similar, they differ in the base position of the subject (Figure 5): postverbal for unaccusatives, preverbal for unergatives (Belletti 1988). Once again, both MAXT and SUMS correctly capture the processing preference under the smuggling analysis (Table 1). Due to the fact that unaccusative subjects are base-generated postverbally, MAXT for these constructions is the lowest it can be (2, the tenure of any right sibling which is predicted and immediately discharged; see Table 2). However, the scrambling approach has a similar issue as before, with MAXT registered on *v* $(13-10)$ due to the verb raising to AspP. This time both metrics predict a tie between the two structures, since in neither case there is subject movement to Spec,TP.

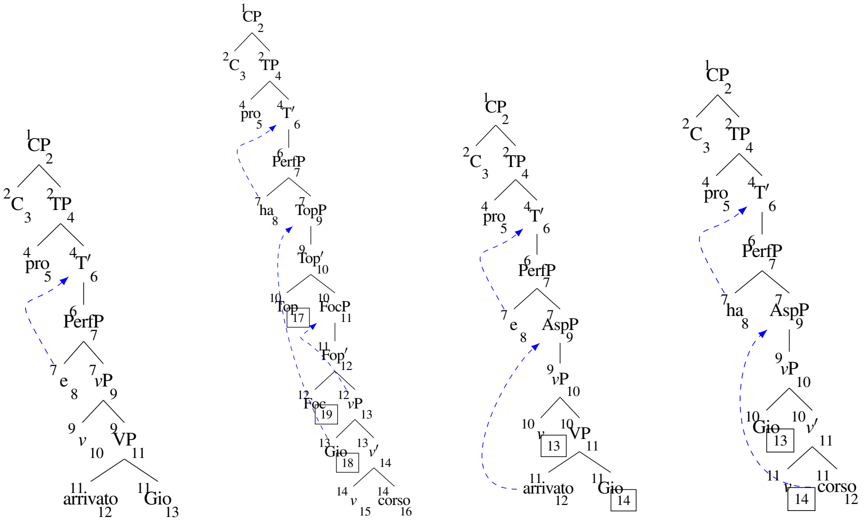
---

[5] Due to space limitations, it is also not possible to show full annotated trees for every syntactic combination. However, all annotated trees, together with LaTex files to replicate the results are available at: https://osf.io/8j2kx/?view_only=251032b6d3ae484fb751e8abc57d8b10.



(a) (b) (c) (d)

FIGURE 4: ANNOTATED DERIVATION TREES FOR *(a,c)* THE SVO SENTENCE IN (1A), AND *(b,d)* THE VS SENTENCE IN (1B), FOLLOWING A SMUGGLING *(a,b)* OR A SCRAMBLING *(c,d)* ANALYSIS.



(a) (b) (c) (d)

FIGURE 5:
ANNOTATED DERIVATION TREES FOR AN UNACCUSATIVE *(a,c)* SENTENCE AND AN UNERGATIVE *(b,d)* SENTENCE, FOLLOWING A SMUGGLING *(a,b)* OR A SCRAMBLING *(c,d)* ANALYSIS.

	SVO < VS		unacc < unerg	
analysis	MaxT	SumS	MaxT	SumS
Smuggling	✓	✓	✓	✓
Scrambling	×	×	tie	tie

Table 1: Predictions of the MG parser for the matrix sentences by contrast.

clause type	MaxT	SumS
matrix SVO	3/*ha*/*v*’	7
matrix VS	7/*Top*/*Foc*	11
VS unacc	2/*v*P	3
VS unerg	7/*Top*/*Foc*	11

(a) Smuggling

clause type	MaxT	SumS
matrix SVO	4/*ha*/*v*	14
matrix VS	3/*v*/*Gio*	6
VS unacc	3/*v*/*Gio*	7
VS unerg	3/*v*/*Gio*	7

(b) Scrambling

Table 2: Summary of MaxT (*value*/*node*) and SumS by construction and postverbal analysis, for the trees in Figure 4 and Figure 5. The expected difficulty gradient is SVO < VS, and unacc < unerg.

4.2 Relative Clauses

For consistency with psycholinguistic stimuli, and with previous MG parsing work, RCs are not modeled by themselves, but are embedded in a template sentence.[6] Thus, I tested the parser performance on *right-branching* restrictive RCs of the form *(pro) vedo il cavallo [$_{RC}$ che ...]* (*I see the horse [$_{RC}$ that ...]*) – the relativized NP occupying the matrix *object* position, and the relative clause either an SRC (4), an ORC (5), or an ORCp (7). As discussed, each construction is modulated across two syntactic dimensions: one of the four RC analyses, and one of the two postverbal analyses. Recall once again that by assumption the parser is equipped with a perfect oracle, and that the complexity metrics are *only* sensitive to structural differences (i.e., the MG model is blind to lexical differences and agreement relationships). Contrasting (4) and (7) is then equivalent to contrasting (6a) and (6b). Thus, to reiterate the central tenants of the approach, these comparisons aim to model both the preference for SRC in structurally ambiguous cases (as done for the matrix SVO/VS sentence before), and the overall increased processing difficulty of ORCps, just in terms of *structural differences*.

Table 3 shows how each of the two metrics fares on the contrasts under consideration, based on the different syntactic choices made for each construction. The numerical values of each metric for the different sentence types are then decomposed

---

[6] Note that this choice does not actually change the results in this section, as the matrix clause structure is consistent across comparisons.

		SRC < ORC		SRC < ORCp		ORC < ORCp	
postverbal	RC Type	MaxT	SumS	MaxT	SumS	MaxT	SumS
Smuggling	Promotion	✓	✓	✓	✓	✓	✓
	Wh-movement	✓	✓	✓	✓	✓	✓
	Extraposition	✓	✓	✓	✓	✓	✓
	DP analysis	✓	✓	✓	✓	✓	✓
Scrambling	Promotion	✓	✓	✓	✓	✓	✓
	Wh-movement	✓	✓	✓	✓	✓	✓
	Extraposition	✓	✓	✓	✓	tie	tie
	DP analysis	✓	✓	✓	✓	tie	tie

Table 3: Predictions of the MG parser for the RC contrast by analysis.

clause type	MaxT	SumS
SRC	8/*che*	18
ORC	11/*ha*	24
ORCp	16/*Foc*	31

(a) Promotion

clause type	MaxT	SumS
SRC	6/*che*	18
ORC	9/*ha*	24
ORCp	14/*Foc*	31

(b) Wh-movement

clause type	MaxT	SumS
SRC	3/*v*/*CP*	17
ORC	5/*ha*/*v*'	21
ORCp	9/*Foc*	22

(c) Extraposition

clause type	MaxT	SumS
SRC	3/*v*/*DP*	12
ORC	5/*ha*/*v*'	16
ORCp	10/*Foc*/*v*	16

(d) DP Analysis

Table 4: Summary of MaxT (*value*/*node*) and SumS predictions
for the right-embedding RC contrasts, varied by RC analysis and assuming a smuggling analysis of postverbal subjects. The expected difficulty gradient is SRC < ORC < ORCp.

by RC analysis in Table 4 and Table 5, based on the chosen approach to postverbal subjects. Importantly, this is where the interpretability of the model comes into play, especially since we are modeling contrasts categorically without attempting to fit the magnitude of the effects. While some of the numerical differences might appear small at a glance, inspecting the trees allows us to determine whether a particular contrast is won due to core structural differences or not.

First, consider the performance of the MG parser when adopting a *smuggling* analysis of postverbal subjects and varying the RC structure. Modeling results show that both MaxT and SumS correctly predict the gradient of difficulty observed for Italian RCs (SRC < ORC < ORCp), independently of RC analysis (Table 3).

Clause type	MaxT	SumS
SRC	8/*che*	18
ORC	11/*ha*	24
ORCp	16/*che*/*v*	31

(a) Promotion

Clause type	MaxT	SumS
SRC	6/*che*	18
ORC	9/*ha*	24
ORCp	11/*che*/*v*	34

(b) Wh-movement

Clause type	MaxT	SumS
SRC	3/*v*/*CP*	17
ORC	5/*ha*/*v*'	21
ORCp	5/*v*	21

(c) Extraposition

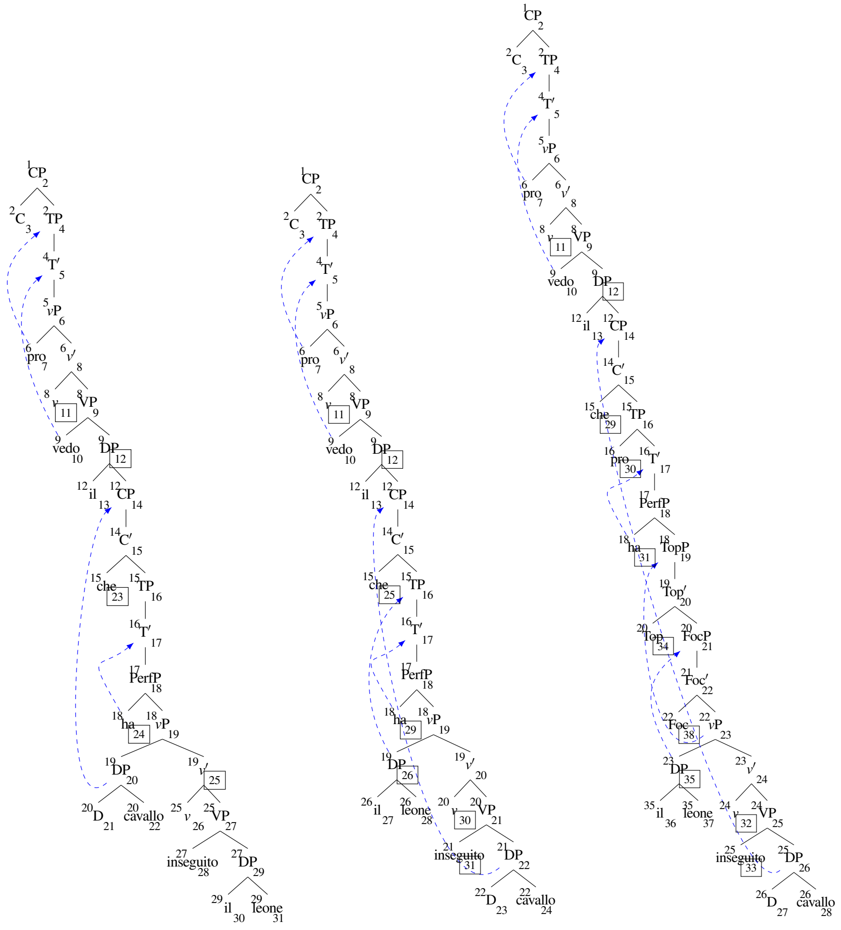
Clause Type	MaxT	SumS
SRC	3/*v*/*DP*	12
ORC	5/*ha*/*v*'	16
ORCp	5/*DP*/*v*	16

(d) DP Analysis

Table 5: Summary of MaxT (*value/node*) and SumS predictions for the right-embedding RC contrasts, varied by RC analysis and assuming a scrambling analysis of postverbal subjects. The expected difficulty gradient is SRC < ORC < ORCp.

However, the Promotion and Wh-movement analyses significantly differ from the DP/Extraposition analyses in *how* the memory predictions are derived, highlighting how the string-driven traversal strategy of the MG parser makes the memory metrics sensitive to *details of the structural representations*. Consider the derivation for the Promotion analysis (Figure 6). In the SRC, *che* is introduced at step 15. Based on information in the input string, the parser is looking for the subject DP *il cavallo*. Thus, *che* has to be kept in memory until the latter is found, and it is only flushed from memory at step 23. In the ORC, *che* is also put in memory at step 15. However, since the head of the relative clause is the embedded object, the parser will discard the standard recursive-descent strategy, ignore the subject DP, and keep expanding nodes until *il cavallo* is found. Thus, *che* cannot be flushed from memory until step 25.

The tenure difference between SRC and ORC (8 vs. 11) also highlights how tenure interacts with movement. Once *che* has been found in the SRC tree, the next node in the stack is *ha*, which can be discharged from memory immediately after. In the ORC however, the parser still has to find the subject DP. Thus, *ha* has to be kept in memory for the three additional steps that are required to conjecture and scan *il leone*. Similarly, the maximum tenure recorded on the Foc head in ORCp (16) highlights the cost of the additional movement steps postulated for this construction. The Foc node needs to wait until both the RC object *and* subject are built and scanned, before being itself discharged from the memory queue. The results for the Wh-movement analysis are derived in a similar fashion, with small differences in tenure due to the fact that in this case it is the operator *Op* that raises from within the RC to Spec,CP (Table 4).



(a) (b) (c)

Figure 6: Annotated derivation trees for an *(a)* SRC, *(b)* ORC, and *(c)* ORCp assuming a promotion analysis of RCs and a smuggling analysis of postverbal subjects.

Consider now the Extraposition analysis (Figure 7). In the SRC case, MaxT (3) is driven by *il* and *cavallo* being in two separate constituents. The parser needs to expand artP and scan *il* before moving to the right daughter of the NP and scan *cavallo*. Importantly though, since the head of the RC does *not* need to raise from within the RC itself, what increases tenure in the ORC case (5) is simply the need to raise the subject DP to Spec,TP. Finally, the additional movement dependencies assumed for ORCps drive the ORC < ORCp ($5 < 9$) result. Crucially, the rightward displacement of the CP containing the RC does not affect the results in any way, since the displacement is identical in all three constructions and it does not interact with the linear string in a way that affects the tree traversal strategy. Because of this fact, these same considerations mostly hold also for the DP analysis, which is lacking any type of movement exclusive to the RC structure.

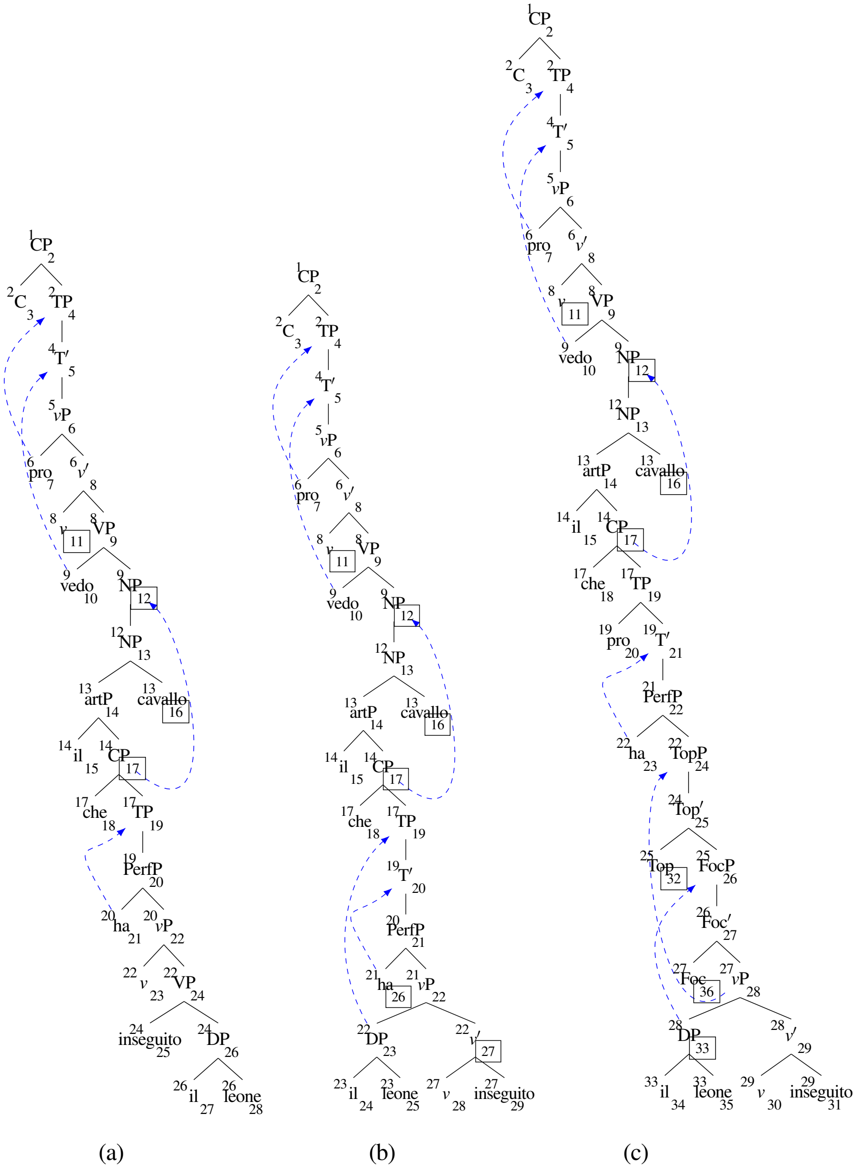


FIGURE 7: ANNOTATED DERIVATION TREES FOR AN *(a)* SRC, *(b)* ORC, AND *(c)* ORCP ASSUMING AN EXTRAPOSITION ANALYSIS OF RCS AND A SMUGGLING ANALYSIS OF POSTVERBAL SUBJECTS.

Differences between RC analyses are made more evident when adopting a *scrambling* approach to postverbal subjects.Note that the discussion above about the SRC and ORC contrast is unchanged, as those structures are unaffected by the choice of postverbal subject analysis. Let us summarize what we learned for those cases. For the Promotion/Wh-movement analyses, MAXT in the SRC/ORC cases is mostly driven by the movement of the RC head to Spec,CP, starting from subject or object position. In the ORC case, this movement also interacts with the need to raise the embedded subject from Spec,*v*P to Spec,TP – thus increasing tenure beyond what would be due to a simple increase in embedding depth. In the case of the Extraposition/DP analyses, there is no real movement within the RC, and MAXT is mostly driven by the movement of the verb to T. In this case, the difference between SRCs and ORCs is not really due to the RCs themselves, but can be reduced to the effect of the absence of subject movement to Spec,TP in SRCs.

These facts explain why the parser is still able to derive the the ORC < ORCp contrast when considering the scrambling approach combined with the Promotion/Wh-movement analyses, but fails to do so when considering the Extraposition/DP analyses (Table 3). Under a Promotion/Wh-movement account, MAXT is measured on the complementiser *che* due to movement of the RC head to Spec,CP, which interacts in non-trivial ways with the fact that the scrambling approach assumes movement of the verb to AspP. Specifically, this additional structure postulated for ORCp but not ORC forces the parser to ignore the subject DP and first expand the verb, leading to MAXT on *v* (16 and 11). In contrast, the tie predicted between ORC and ORCp with the the Extraposition/DP analyses is explained by the fact that under these analyses tenure is simply driven by subject movement to Spec,TP in ORCs. Since in ORCps there is a null *pro* directly in Spec,TP, moving the verb to Asp does not really affect how the parser approaches the search for the subject DP. The fact that MAXT for the two structures is on significantly different nodes highlights how the metrics' tie is mostly a coincidence, due to the additional projection introduced by the scrambling approach compensating for the lack of movement – and highlights the advantage of having a transparent computational model which provides results interpretable upon investigation (Table 5).

5. DISCUSSION

This paper explored how a parser for Minimalist grammars, equipped with measures of memory usage, can model the processing asymmetries reported for Italian postverbal subject constructions. As pointed out early in the paper, the model's approach to memory load is explicitly influenced by how the syntactic derivation assumed for a sentence interacts with its expected linear order. While this explicit link allows us to connect psycholinguistic results to work in theoretical syntax, it also makes it important to understand how much the predictions of the model depend on the specific syntactic analysis of choice. Thus, this paper modulated a variety of processing contrasts across different approaches to the structure of RCs and postverbal constructions. A summary of these results can be found in Table 1 and Table 3.

By focusing on the performance of two specific metrics (SUMS and MAXT), Section 4 showed us how the parsing model easily picked up on the additional projections and movement dependencies postulated by a smuggling analysis of subject inversion – predicting the correct asymmetries for a variety of matrix clause processing profiles. This contrasted with the results obtained with a scrambling analysis. This latter approach still postulates extra projections and additional movement dependencies in the structure of postverbal subjects. However, the subject of the clause is assumed to remain *in situ* in its base argument position. This crucial change makes it so that the tree traversal strategy is less affected by the extra structure of the inverted constructions, and made the model unable to predict the correct contrasts.

When considering RC structures, the MG model was fully successful in predicting the expected SRC $<$ ORC $<$ ORCp gradient under a smuggling analysis, independently of the approach to RCs. However, important differences between analyses still arised when looking at what drives memory load. Specifically, under a Promotion/Wh-movement analysis MAXT differences between SRCs and ORCs are due to fundamental properties of the RC (i.e., the base position of the RC head). Instead, tenure differences in the Extraposition/DP analysis case are due to assumptions about subject movement that are independent of RCs themselves. In turn, this lead to an inability of the model to produce the correct results when the Extraposition/DP analyses are combined with a scrambling approach to subject inversion.

Overall, the success of the top-down parser in accounting for the Italian processing contrasts adds support to the MG model as a valuable theory of how processing cost is tied to structure (Graf *et al.* 2017; De Santo 2020b). Moreover, by modulating the phenomena under study across different syntactic choices explicitly, the MG model also put us in the position to fully explore how structural assumptions affect claims about memory load. This was highlighted in Section 4, when discussing the effects of different structural assumptions on the performance of Tenure and Size metrics. The contributions of this line of inquiry are thus twofold.

From one side, the results in this paper improve our understanding of the MG model itself, by clarifying which aspects of sentence structure drive the parser's performance, and how they weight on the recruitment of memory resources as measured by different metrics. Importantly, they also highlight how important it is to evaluate multiple interacting constructions, in order to truly understand the effects syntactic choices have on the parser's predictions.

From the other side, the MG model is obviously not the first associating some kind of memory cost to structure building operations. However, the precise specification of the parsing model and its transparent linking theory between structure and memory load allow us to reinterpret previous theories in a quantifiable framework that directly connects parsing processes to cognitive resources.

Consider De Vincenzi (1991)'s Minimal Chain Principle (MCP), which has been used in the past as a way to ground Italian postverbal asymmetries in parsing effects. The MCP postulates that shorter dependencies are computationally less demanding than longer dependencies: thus SRCs are easier than ORCs because the filler gap distance in the former is shorter than in the latter (cf. Gibson 2000). *Economy* principles also predict the increased difficulty found for ORCs with a postverbal subject. However, the MCP leaves unspecified how these computational demands would be implemented in a precise parsing architecture, and how these costs are linked to cognitive resources like working memory, known to affect processing effects (Utzeri 2007). The MG parser explicitly connects processing differences to the additional memory resources involved in keeping track of long movement dependencies. Thus, it offers a way to reinterpret De Vincenzi (1991)'s theory, and economy claims more generally, in a framework that takes structural assumptions seriously.

A different line of research associates efforts in ORC processing to *interference effects* caused by the relative head moving across an embedded subject endowed with a similar feature set (Friedmann *et al.* 2009; Belletti & Contemori 2009; Arosio *et al.* 2009; Villata *et al.* 2016: featural Relativized Minimality). This approach is unique in that structure building operations only matter for processing based on how they modulate the relation of lexical items based on their underlying features. Crucially though, these predictions are also going to be affected by specific structural assumptions. Additionally, interference effects do not fully predict differences in performance on ORC over ORCp, and thus a Relativized Minimality account needs to refer to additional mechanisms. For instance, past work attributes this contrast to different subject-verb agreement operations found in the two structures (Volpato 2010; Volpato & Adani 2009). That is, some postulate that in ORCs agreement is more *robust*, since it is double checked under AGREE and under a local Specifier-Head checking operation. In contrast, in ORCps agreement is supposed to be more "fragile" since it is only checked once under a non-local AGREE (Arosio *et al.* 2017).

Importantly, the evidence for effects associated to lexical features modulating processing difficulty profiles is compelling. While the current metrics do not take features into account when computing memory load, MGs offer direct ways to explore interference effects, since they are fundamentally a feature-driven formalism. Incorporating the effect of features into this model would thus require no change to the parser's implementation, but simply the introduction of memory metrics sensitive to the feature component of lexical items (cf. Chesi & Canal 2019; De Santo 2021: a.o.).

Relatedly, the existing model idealizes several aspects of the parsing process. For instance, we abstracted non-determinism away, and consciously ignored the role that (local and global) ambiguity resolution plays in processing. The MG model adopted here might thus seem to support a syntactic “reductionist” view, in its attempt to explain processing asymmetries purely in terms of structure building operations. As already mentioned though, this is not to be interpreted as claiming structural cost to be a comprehensive explanation of processing difficulty. It is without doubt that a cognitively realistic theory would see multiple factors interact with each other to derive the correct contrasts (Demberg & Keller 2008; Brennan *et al.* 2016: a.o.). In fact, the current idealization guides the choice of phenomena that it is reasonable to model with this approach – so that, for instance, we investigate RC asymmetries but not garden-path effects (Townsend & Bever 2001). Crucially, the current model seems to be compatible with a variety of approaches to ambiguity resolution both in terms of broader perspectives – e.g., parallelism vs. serial backtracking – and specific computational metrics (e.g., surprisal). Extensions to the current metrics would then allow us to clearly quantify the contribution of a variety of different components (length of dependencies, feature overlap, ambiguity, context, etc.) to sentence processing effects. Additionally, new metrics would also allow us to explore processing asymmetries that cannot simply be reduced to syntactic differences (Pasternak & Graf 2020).

Similar issues motivate this paper’s focus on off-line predictions exclusively. It is crucial to note that the memory types discussed here *can* derive measures of on-line complexity (e.g. tenure on every node). However, it seems reasonable that on-line complexity profiles should be more deeply affected by those elements of sentence processing we have been ignoring (e.g. local ambiguity) than off-line profiles. Exploring how this model fares for on-line predictions is obviously a essential area of future work. Some results in this direction can be found already in (Gerth 2015), which incorporates structural metrics similar to tenure with information theoretic measures, and even sketched in (Kobele *et al.* 2013). A fruitful way of exploring these connections might be to adopt multiple complexity metrics modeling different aspects of sentence processing, and comparing how they fare in predicting the magnitude of processing effects throughout a sentence (Brennan *et al.* 2016).

Finally, comparing the performance of the parser on the same construction across different analyses sheds light on the effect that competing structural assumptions can have on theories of syntactic processing. In this sense, this paper’s results stress how crucial it is to explore how hypotheses about the underlying structure of different parts of a clause can interact in unexpected ways. Far from being an issue exclusive to the model presented here, this observation extends to every theory of processing difficulty that appeals to structural dependencies – and calls for explicit theoretical commitments to syntactic details not just in computational modeling, but in the broader design of psycholinguistic experiments testing sentence processing phenomena. This seems to be a desirable perspective, if one believes that theories of syntactic representations have any kind of cognitive reality (Bresnan 1978; Joshi

1990; Kush & Dillon 2021). Ideally, it should be possible to use particular analyses supported by theoretical and comparative linguistic research in order to derive (or gain insights into) processing contrasts (Kobele *et al.* 2013). Additionally, while the different analyses adopted here are not particularly controversial from a theoretical standpoint, it is encouraging that the results of the model support the independently motivated consensus in the syntactic literature. It should then be possible to use this framework to approach linguistic constructions that are more theoretically opaque (Rambow & Joshi 1994), and have experimental observations not substitute but complement other types of linguistic data.

REFERENCES

Abney, S.P. (1987). *The English noun phrase in its sentential aspect*. Ph.D. thesis, Massachusetts Institute of Technology.

Adani, F. (2011). Rethinking the acquisition of relative clauses in Italian: towards a grammatically based account. *Journal of Child Language*, 38(1). 141–165.

Antinucci, F. & G. Cinque (1977). Sull'ordine delle parole in italiano: l'emarginazione. *Studi di grammatica italiana VI, pp. 121-146*.

Arosio, F., F. Adani & M.T. Guasti (2009). Grammatical features in the comprehension of Italian Relative Clauses by children. *Merging Features: Computation, Interpretation, and Acquisition*. 138–158.

Arosio, F., F. Panzeri, B. Molteni, S. Magazù & M.T. Guasti (2017). The comprehension of Italian relative clauses in poor readers and in children with specific language impairment. *Glossa: a journal of general linguistics*, 2(1).

Bates, E. & B. MacWhinney (1987). Competition, variation, and language leaning. *Mechanisms of language acquisition*.

Belletti, A. (1988). The case of unaccusatives. *Linguistic inquiry*, 19(1). 1–34.

Belletti, A. (2004). Aspects of the low ip area. *The structure of CP and IP. The cartography of syntactic structures*, 2. 16–51.

Belletti, A. & C. Contemori (2009). Intervention and attraction. on the production of subject and object relatives by Italian (young) children and adults. In *Language acquisition and development, 3. Proceedings of Gala*. 39–52.

Belletti, A. & C. Leonini (2004). Subject inversion in L2 Italian. *EUROSLA yearbook*, 4. 95–118.

Berwick, R.C. & A.S. Weinberg (1982). Parsing efficiency, computational complexity, and the evaluation of grammatical theories. *Linguistic Inquiry*, 13. 165–291.

Berwick, R.C. & A.S. Weinberg (1983). The role of grammar in models of language use. *Cognition*, 13. 1–61.

Berwick, R.C. & A.S. Weinberg (1985). The psychological relevance of transformational grammar: a reply to Stabler. *Cognition*, 19. 193?204.

Bianchi, V. (2002a). Headed relative clauses in generative syntax. part i. *Glot International*, 6(7). 197–204.

Bianchi, V. (2002b). Headed relative clauses in generative syntax. part ii. *Glot International*, 6(8). 1–12.

Bianchi, V., G. Bocci & S. Cruschina (2017). Two types of subject inversion in italian wh-questions. *Revue roumaine de linguistique*, 62(3). 233–252.

Bocci, G. (2013). *The Syntax Prosody Interface: A cartographic perspective with evidence from Italian*, volume 204. John Benjamins Publishing.

Borer, H. (1994). The projection of arguments. *University of Massachusetts occasional papers in linguistics*, 20(1). 3.

Brennan, J.R., E.P. Stabler, S.E. Van Wagenen, W.M. Luh & J.T. Hale (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, 157. 81–94.

Bresnan, J. (1978). A realistic transformational grammar. *Linguistic theory and psychological reality*. 1–59.

Brunetti, L. (2003). *A unification of focus*. Ph.D. thesis, Università degli Studi di Firenze.

Cardinaletti, A. (1998). A second thought on" emarginazione": Destressing vs." right dislocation". *Working Papers in Linguistics, 8.2, 1998, pp. 1-28*.

Cardinaletti, A. (2004). Toward a cartography of subject positions in the structure of ip and cp: The cartography of syntactic structures 2, ed. l. rizzi, 115-165.

Cardinaletti, A. (2018). On different types of postverbal subjects in italian. *Italian Journal of Linguistics*, 30(2). 79–106.

Chesi, C. & P. Canal (2019). Person features and lexical restrictions in italian clefts. *Frontiers in psychology*, 10. 2105. doi:10.3389/fpsyg.2019.02105.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*, volume 11. MIT Press.

Chomsky, N. (1977). On wh-movement. *Formal syntax*. 71–132.

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

Collins, C. (2005). A smuggling approach to the passive in english. *Syntax*, 8(2). 81–120.

Contemori, C. & A. Belletti (2014). Relatives and passive object relatives in Italian-speaking children and adults: Intervention in production and comprehension. *Applied Psycholinguistics*, 35(6). 1021–1053.

De Santo, A. (2019). Testing a Minimalist grammar parser on Italian relative clause asymmetries. In *Proceedings of the ACL Workshop on Cognitive Modeling and Computational Linguistics (CMCL) 2019*. June 6 2019, Minneapolis, Minnesota.

De Santo, A. (2020a). Mg parsing as a model of gradient acceptability in syntactic islands. *Proceedings of the Society for Computation in Linguistics*, 3(1). 53–63.

De Santo, A. (2020b). *Structure and memory: A computational model of storage, gradience, and priming*. Ph.D. thesis, State University of New York at Stony Brook.

De Santo, A. (2021). A minimalist approach to facilitatory effects in stacked relative clauses. *Proceedings of the Society for Computation in Linguistics*, 4(1). 1–17.

De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian: The minimal chain principle*, volume 12. Springer Science & Business Media.

Demberg, V. & F. Keller (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2). 193 – 210. doi:https://doi.org/10.1016/j.cognition.2008.07.008.

Frauenfelder, U.H., J. Segui & J. Mehler (1980). Monitoring around the relative clause. *Journal of Verbal Learning and Verbal Behavior*, 19(3). 328–337.

Frazier, L. (1987). Syntactic processing: evidence from Dutch. *Natural Language & Linguistic Theory*, 5(4). 519–559.

Friedmann, N., A. Belletti & L. Rizzi (2009). Relativized relatives: Types of intervention in the acquisition of a-bar dependencies. *Lingua*, 119(1). 67–88.

Gerth, S. (2015). *Memory Limitations in Sentence Comprehension: A Structural-based Complexity Metric of Processing Difficulty*, volume 6. Universitätsverlag Potsdam.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1). 1–76.

Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. In *2000, Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*. MIT press, 95–126.

Graf, T. & B. Marcinek (2014). Evaluating evaluation metrics for minimalist parsing. In *Proceedings of the 2014 ACL Workshop on Cognitive Modeling and Computational Linguistics*. 28–36.

Graf, T., B. Fodor, J. Monette, G. Rachiele, A. Warren & C. Zhang (2015). A refined notion of memory usage for minimalist parsing. In *Proceedings of the 14th Meeting on the Mathematics of Language (MoL 2015)*. Chicago, USA: Association for Computational Linguistics, 1–14.

Graf, T., J. Monette & C. Zhang (2017). Relative clauses as a benchmark for Minimalist parsing. *Journal of Language Modelling*, 5. 57–106. doi:10.15398/jlm.v5i1.157.

Greco, M., P. Lorusso, C. Chesi & A. Moro (2020). Asymmetries in nominal copular sentences: Psycholinguistic evidence in favor of the raising analysis. *Lingua*, 245. 102926.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

Harkema, H. (2001). A characterization of minimalist languages. In *International Conference on Logical Aspects of Computational Linguistics*. Springer, 193–211.

Joshi, A.K. (1990). Processing crossed and nested dependencies: An automation perspective on the psycholinguistic results. *Language and cognitive processes*, 5(1). 1–27.

Kayne, R.S. (1994). *The antisymmetry of syntax*. 25, MIT Press.

Kempchinsky, P. (2000). Aspect projections and predicate type. *Hispanic Linguistics at the Turn of the Millenium*. 171–187.

King, J.W. & M. Kutas (1995). Who did what and when? using word-and clause-level erps to monitor working memory usage in reading. *Journal of cognitive neuroscience*, 7(3). 376–395.

Kobele, G.M., C. Retoré & S. Salvati (2007). An automata-theoretic approach to minimalism. In *Model Theoretic Syntax at 10*. J. Rogers and S. Kepser, 73–82.

Kobele, G.M., S. Gerth & J. Hale (2013). Memory resource allocation in top-down minimalist parsing. In *Formal Grammar*. Springer, 32–51.

Kush, D. & B. Dillon (2021). Sentence processing and syntactic theory. *A Companion to Chomsky*. 305–324.

Lau, E. & N. Tanaka (2021). The subject advantage in relative clauses: A review. *Glossa: a journal of general linguistics*, 6(1).

Lee, S.Y. (2018). A minimalist parsing account of attachment ambiguity in English and Korean. *Journal of Cognitive Science*, 19(3). 291–329.

Leonetti, M. (2018). Two types of postverbal subjects. *Italian Journal of Linguistics*, 30(2). 11–36.

Levy, R. (2013). Memory and surprisal in human sentence comprehension. In *Sentence processing*, 90–126. Psychology Press.

Lewis, R.L. & S. Vasishth (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, 29(3). 375–419.

Liu, L. (2018). Minimalist Parsing of Heavy NP Shift. In *Proceedings of PACLIC 32 The 32nd Pacific Asia Conference on Language, Information and Computation*. The Hong Kong Polytechnic University, Hong Kong SAR.

Longobardi, G. (2000). “postverbal” subjects and the mapping hypothesis. *Linguistic Inquiry*, 31(4). 691–702.

Michaelis, J. (1998). Derivational minimalism is mildly context–sensitive. In *International Conference on Logical Aspects of Computational Linguistics*. Springer, 179–198.

Miller, G.A. & N. Chomsky (1963). Finitary models of language users. In R. Luce, R. Bush & E. Galanter (eds.) *Handbook of Mathematical Psychology*, volume 2. New York: John Wiley.

Ordóñez, F. (1998). Post-verbal asymmetries in Spanish. *Natural Language & Linguistic Theory*, 16(2). 313–345.

Pasternak, R. & T. Graf (2020). Cyclic scope and processing difficulty in a minimalist parser. *Manuscript. https://ling.auf.net/lingbuzz/005009*.

Phillips, C. (2003). Parsing: Psycholinguistic aspects. In *International Encyclopedia of Linguistics*. Oxford University Press, 2 edition.

Prince, A. & P. Smolensky (2008). *Optimality Theory: Constraint interaction in generative grammar*. John Wiley & Sons.

Rambow, O. & A.K. Joshi (1994). A processing model for free word order languages. *Perspectives on Sentence Processing*.

Rizzi, L. (1990). *Relativized minimality.* The MIT Press.

Samek-Lodovici, V. (2015). *The interaction of focus, givenness, and prosody: A study of Italian clause structure*. Oxford University Press.

Schriefers, H., A.D. Friederici & K. Kuhn (1995). The processing of locally ambiguous relative clauses in german. *Journal of Memory and Language*, 34(4). 499.

Stabler, E.P. (1984). Berwick and weinberg on linguistics and computational psychology. *Cognition*, 17(2). 155–179.

Stabler, E.P. (1996). Derivational minimalism. In *International Conference on Logical Aspects of Computational Linguistics*. Springer, 68–95.

Stabler, E.P. (2011). Computational perspectives on minimalism. In *The Oxford Handbook of Linguistic Minimalism.*

Stabler, E.P. (2013). Two models of minimalist, incremental syntactic analysis. *Topics in cognitive science*, 5(3). 611–633.

Townsend, D.J. & T.G. Bever (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Utzeri, I. (2007). The production and acquisition of subject and object relative clauses in Italian: a comparative experimental study. *Nanzan Linguistics*, 2.

Vasishth, S. & R.L. Lewis (2006). Symbolic models of human sentence processing.

Villata, S., L. Rizzi & J. Franck (2016). Intervention effects and relativized minimality: New experimental evidence from graded judgments. *Lingua*, 179. 76–96.

Volpato, F. (2010). *The acquisition of relative clauses and phi-features: evidence from hearing and hearing-impaired populations*. Ph.D. thesis, Università Ca'Foscari di Venezia.

Volpato, F. & F. Adani (2009). The subject/object relative clause asymmetry in Italian hearing-impaired children: evidence from a comprehension task. *Studies in Linguistics*, 3. 269–281.

Zhang, C. (2017). *Stacked Relatives: Their Structure, Processing and Computation*. Ph.D. thesis, State University of New York at Stony Brook.

Aniello De Santo
Department of Linguistics, University of Utah
Salt Lake City, UT
USA
e-mail: `aniello.desanto@utah.edu`