Testing a Minimalist Grammar Parser on Italian Relative Clause Asymmetries

Aniello De Santo
Department of Linguistics
Stony Brook University
aniello.desanto@stonybrook.edu

Abstract

Stabler's (2013) top-down parser for Minimalist grammars has been used to account for off-line processing preferences across a variety of seemingly unrelated phenomena cross-linguistically, via complexity metrics measuring "memory burden". This paper extends the empirical coverage of the model by looking at the processing asymmetries of Italian relative clauses, as I discuss the relevance of these constructions in evaluating plausible structure-driven models of processing difficulty.

1 Introduction

Recent studies have shown that a top-down parser for Minimalist grammars (MGs; Stabler, 1996, 2013)) can be combined with complexity metrics to relate parsing behavior to memory usage, and successfully used to model sentence processing preferences across a variety of phenomena cross-linguistically (Kobele et al., 2013; Gerth, 2015; Graf et al., 2017). This kind of work follows a line of research on syntactic processing that sees computational models provide a transparent, interpretable linking theory between syntactic assumptions and processing behavior (Joshi, 1990; Rambow and Joshi, 1994; Hale, 2011). Importantly, at the core of the particular approach adopted here is a theory of grammatical structure driving off-line processing cost, thus connecting longstanding ideas about memory resources with explicit syntactic analyses in rigorous ways. Extending the range of phenomena correctly modeled by the parser is then going to be crucial to confirm the empirical feasibility of the approach.

Here, I adopt Kobele *et al.*'s (2013) implementation of Stabler's (2013) top-down traversal algorithm, coupled with the set of complexity metrics defined by Graf et al. (2017). We test the MG parser's performance on the processing asymmetries reported for Italian relative clauses, which have been object of extensive study in the psycholinguistic literature. Apart from conforming to a well-attested cross-linguistic preference for subject over object relatives, Italian speakers also show increased processing difficulties when encountering relative clauses with subjects in postverbal position. This difficulty gradient has often been accounted for in the literature in terms of the cost of local ambiguity resolution. Since in the particular formulation of Kobele et al. (2013) the MG parser acts as an oracle and deliberately ignores structural ambiguity, these constructions thus make for a challenging testing ground for a model attempting to account for processing contrasts *just* in terms of *structural complexity*.

The paper is structured as follows. Section 2 presents an informal introduction to MGs and Stabler's (2013) top-down parser, and an overview of previous work on combining the MG parser with complexity metrics measuring memory burden. Section 3 discusses Italian relative clause asymmetries and our modeling assumptions. Section 4 looks at the modeling results, and shows how the MG parser succeeds in predicting the correct processing preferences. Section 5 concludes with a brief discussion of possible limits of the model, and promising future work.

2 Preliminaries

2.1 Minimalist Grammars

MGs (Stabler, 1996, 2011) are a highly lexicalized, mildly context-sensitive formalism incorporating the structurally rich analyses of Minimalist syntax — the most recent version of Chomsky's transformational grammar framework. Therefore, they have proven to be a fruitful grammar formalism in investigating how ideas from theoretical syntax weight on sentence processing. While

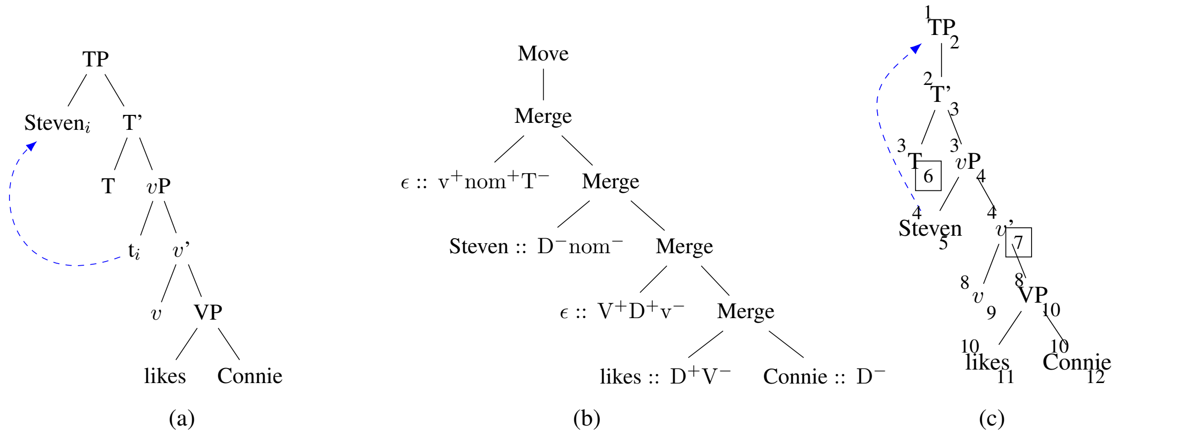


Figure 1: Phrase structure tree (a), MG derivation tree (b), and annotated derivation tree (c) for *Steven likes Connie*. Boxed nodes in (c) are those with tenure value greater than 2, following (Graf and Marcinek, 2014).

much work has been done on the formal properties of MGs, the fine-grained details of the formalism are unnecessary given the focus of this paper. Thus, I introduce MGs in a mostly informal way, as my main goal is to provide the reader with an intuitive understanding of the core data structure the parser is going to operate on: derivation trees.

An MG grammar is a set of of lexical items (LIs) consisting of a phonetic form and a finite, non-empty string of features. They distinguish two types on features, each with either *positive* or *negative* polarity: *Merge* features (written here in upper caps, with the exception of little v), and *Move* features (in lower caps). LIs are assembled via two feature checking operations: *Merge* and *Move*. Informally, Merge combines two LIs if their respective first unchecked features are Merge features of opposite polarity. Move removes a phrase from an already assembled tree and displaces it to a different position (Stabler, 2011). Intuitively, Merge encodes subcategorization, while Move long-distance movement dependencies.

MGs succinctly encode the sequence of Merge and Move operations required to build the phrase structure tree for a specific sentence into *derivation trees* (Michaelis, 1998; Harkema, 2001; Kobele et al., 2007). For instance, Fig. 1a and Fig. 1b compare these two kind of trees for a simplified analysis of the sentence *Steven likes Connie*. In the derivation tree (Fig. 1b), 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 fully 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 — can be regarded as a simple variant of context-free grammars (CFG), which have been studied extensively in the computational parsing literature. This is the crucial insight behind Stabler's top-down parser.

2.2 MG Parsing

Stabler's (2013) parser for MGs is a variant of a standard depth-first, top-down parser for CFGs: it takes as input a sentence represented as string of words, hypothesizes the structure top-down, verifies that the words in the structure match the input string, and outputs a tree encoding of the sentence structure. Basically, the parser scans the nodes from top to bottom and from left to right; but since the surface order of lexical items in the derivation tree is not the phrase structure tree's surface order, simple left-to-right scanning of the leaf nodes yields the wrong word order. Thus, while scanning the nodes, the MG parser must also keep tracking the derivational operations which affect the linear word order.

Without delving too much in 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).

Essential to this procedure is the role of memory: 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.*'s (2013) notation, in which each node in the tree is annotated with an *index* (superscript) and an *outdex* (subscript). Intuitively, the annotation indicates for each node in the tree when it is first conjectured by the parser (index) and placed in the memory queue, and at what point it is considered completed and flushed from memory (outdex). In the rest of the paper I adopt an annotated, simplified version of derivation trees, with internal nodes explicitly labelled and dashed arrows indicating movement relations (as shown in Fig. 1c).[1]

Finally, note that in Stabler's original formulation the parser is equipped with a search beam discarding the most unlikely predictions. In this paper though, 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 the specific contribution of syntactic complexity to processing difficulty. In Sec. 3 I will discuss how assuming an idealized parser is exactly what makes Italian RCs an interesting test case.

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, which here takes the form of complexity metrics. Specifically, I employ complexity metrics that predict processing difficulty based on how the geometry of the trees built by the parser affects memory usage.

Extending previous work on MG parsing (Kobele et al., 2013; Graf and Marcinek, 2014; Gerth, 2015), Graf et al. (2017) distinguish three cognitive notions of memory usage: I) how long a node is kept in memory (*tenure*); II) how many nodes must be kept in memory (*payload*); or III) how many bits a node consumes in memory (*size*). Tenure and payload for each node n in the tree can be easily computed via the node annotation scheme of Kobele *et al.*: a node's tenure is equal to the difference between its index and its outdex; the payload of a derivation tree is computed as the number of nodes with a tenure strictly greater than 2. Defining size in an informal way is slightly trickier, as its original conception was based on how information about movers is stored by Stabler's top-down parser (for a technical discussion, see Graf et al., 2015). Procedurally, the size of the parse item corresponding to each node n can be simply computed by exploiting our simplified representation of derivation trees: it corresponds to the number of nodes below n that have a movement arrow pointing to somewhere above n.[2] For example, referring to the annotated tree in Fig. 1c, the size of vP is 1, while the size of VP is 0. In practice, size encodes how many nodes in a derivation consume more memory because a certain phrase m moves across them.

With the exception of payload, these concepts are not exactly metrics we can use to directly compare derivations. What we are missing is a way for them to be applied to a given derivation as measures of overall processing difficulty. In order to do so, these notions of memory have been used to define a vast set of complexity metrics measuring processing difficulty over a full derivation tree. In this paper, we look at Italian relative clause asymmetries using the full set of 1600 metrics as defined in Graf et al. (2017). However, in what follows we only give a general intuition of how such metrics can be defined, and we refer the reader to Graf *et al.* for the detailed formal definitions. Importantly, just a few of these metrics are enough to account for the contrasts we are interested in.

Kobele et al. (2013) show that tenure can be associated to quantitative values by defining metrics like MAXT $:= max(\{\textit{tenure-of}(n)\})$ and SUMT $:= \sum_n \textit{tenure-of}(n)$. 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 and Marcinek (2014) show that MAXT (restricted to

[1] Note that, due to the fact that intermediate landing sites for moved phrases do not affect the traversal strategy, they not explicitly marked by movement arrows.

[2] Thus, as a reviewer correctly notes, size is sensitive to the hierarchical distance between the filler and the gap.

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's (2014) analysis of relative clause constructions, Graf et al. (2015) argue for the insufficiency of MAXT as a single, reliable metric. They then introduce several new metrics, inspired by those defined for tenure. For example, they define an the equivalent of SUMT for size, which measures 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. Then 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 $\langle M_1, M_2, \ldots, M_n \rangle$, similar to constraint ranking in Optimality Theory (Prince and 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). This suggestion is fully explored in Graf et al. (2017), which show that when complexity metrics are allowed to be ranked in such a way the total number of possible metrics quickly reaches an astronomical size. However, surveying the variety of previously modeled phenomena, the authors also suggest that the number of metrics truly needed to account for human processing contrasts can be reduced to a small number of core metrics (particularly, they point toward a combination of MAXT and SUMS), an hypothesis that seems supported by recent work on several different constructions (Liu, 2018; Lee, 2018).

3 Modeling Italian RCs

3.1 Processing Asymmetries

Restrictive relative clauses (RCs) in Italian have been the focus of extensive experimental studies from the perspective of comprehension (Volpato and Adani, 2009), production (Belletti and Contemori, 2009), and acquisition (Volpato, 2010; Friedmann et al., 2009). Italian speakers conform to the general cross-linguistic preference for subject over object RCs (Frauenfelder et al., 1980; King and Kutas, 1995; Schriefers et al., 1995, a.o.), so that (1) is easier to process than (2):

(1) Il cavallo che ha inseguito i leoni
The horse that has chased the lions
"The horse that chased the lions" **SRC**

(2) Il cavallo che i leoni hanno inseguito
The horse that the lions have chased
"The horse that the lions chased" **ORC**

Interestingly, Italian also allows for sentences like (3), ambiguous between a SRC interpretation (3a) and an ORC interpretation (3a) with the embedded subject expressed postverbally:

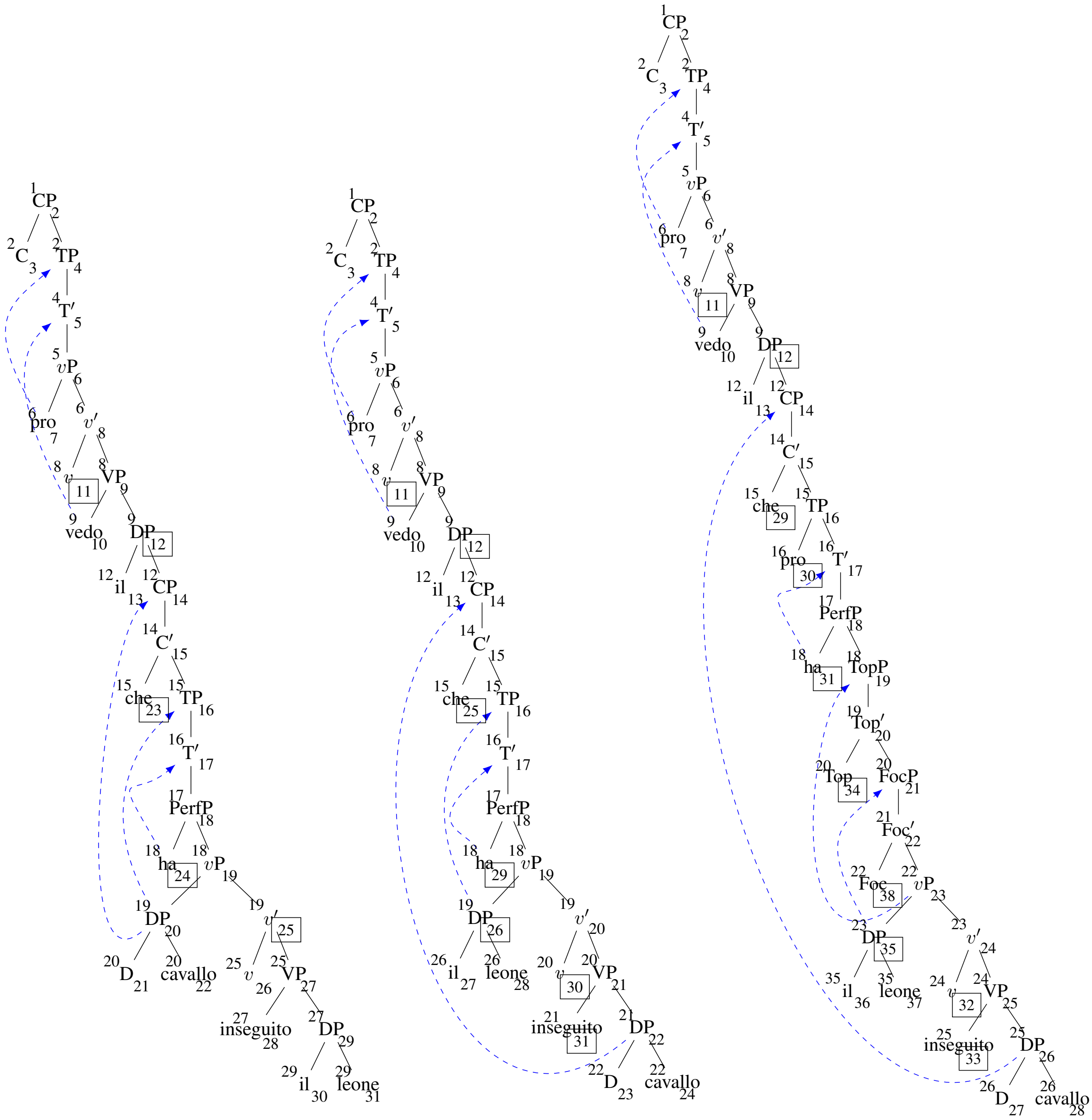
(3) Il cavallo che ha inseguito il leone
The horse that has chased the lion
a. "The horse that chased the lion" **SRC**
b. "The horse that the lion chased" **ORCp**

Although postverbal subject constructions are very common in Italian, in such cases native speakers show a marked preference for the SRC interpretation over the ORCp one. Sentences like (3) can be disambiguated by grammatical cues like subject-verb agreement:

(4) Il cavallo che hanno inseguito i leoni **ORCp**
The horse that have chased the lions
"The horse that the lions chased"

However, even in unambiguous cases like (4), studies report increased efforts with ORCp, leading to the following difficulty gradient: SRC < ORC < ORCp (Utzeri, 2007, a.o.).

The contrast between SRCs and ORCs has been well studied in the past, and it is compatible with a variety of models, such as surprisal (Levy, 2013), cue-based memory retrieval (Lewis and Vasishth, 2005), the active filler strategy (Frazier, 1987), the Dependency Locality Theory (Gibson, 1998, 2000), the Competition Model (Bates and MacWhinney, 1987), the Minimal Chain Principle (De Vincenzi, 1991), among many. The increased complexity reported for ORCs with postverbal subjects comes as a challenge to some of these models (e.g., for the Competition model and Dependency Locality Theory; Arosio et al., 2009). However, their processing profile can be explained in terms of economy of gap prediction and cost of structural re-analysis, due to the possible ambiguity in ORCps at the embedded subject site — where the parser has the choice of either postulating a null pronominal subject or establishing a filler-gap dependency. Importantly though, the



(a) (b) (c)

Figure 2: Annotated derivation trees for right-embedding (a) SRC, (b) ORC, and (c) ORCp.

aim of this paper is not to argue for the correctness (or lack thereof) of these accounts. Our purpose is to extend previous evaluations of memory metrics for a top-down MG parser as a reliable model of processing difficulty.

As discussed above, the MG parser has already been successful in accounting for RC asymmetries cross-linguistically (Graf et al., 2017; Zhang, 2017). Thus, Italian RCs are the perfect next step in understanding the plausibility of the model, allowing us to build on the insights provided by previous work while incrementally exploring new structural configurations. In particular, the fact that by assumption the MG parser ignores structural ambiguity (thus potential costs associated to reanalysis) and deterministically builds only the correct parse, makes ORCs with postverbal subjects an intriguing test case.

3.2 Syntactic Assumptions

The central tenant of the MG model is to take syntactic commitments seriously, so to explore how different aspects of sentence structure drive processing cost. The choice of a syntactic analysis is then particularly important. In line with most of the psycholinguistic literature on Italian RCs, this paper's analysis of postverbal subjects follows Belletti and Leonini (2004, a.o.). Specifically, I assume that in ORCp constructions the subject DP *[i leoni]* is merged in preverbal subject position Spec,vP, and then raised to a Spec,Focus position in the clause-internal vP periphery. The whole verbal cluster is raised to a clause-internal Spec,Topic position; and an expletive *pro* is base generated in Spec,TP and co-indexed with the postverbal subject (Fig. 2c).[3] Furthermore, again consistently with the Italian psycholinguistic literature (Arosio et al., 2017, a.o.), we adopt a promotion analysis of relative clauses (Kayne, 1994). That is to say, 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 treated as an NP, and selected by the determiner that would normally select the head noun in more traditional, head-external accounts (Chomsky, 1977).

4 Modeling Results

4.1 Core Results

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 RC head raising to the matrix object position, and the embedded relative clause either an SRC (1), an ORC (2), or an ORCp (4). The corresponding derivation trees, annotated by the MG parser with index and outdex values at each node, are shown in Fig. 2a, Fig. 2b, and Fig. 2c respectively. Recall 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 agreement relationships). Contrasting (1) and (4) is then equivalent to contrasting (3a) and (3b). Thus, to reiterate the central tenants of the approach, these comparisons aim to model both the preference for SRC in ambiguous cases, and the overall increased processing difficulty of ORCps, just in terms of structural differences.

Modeling results show that the parser correctly predicts the gradient of difficulty observed for Italian RCs (SRC $<$ ORC $<$ ORCp), across a variety of complexity metrics.[4] In fact, the increased difficulty of ORCps over both SRCs and ORCs is predicted by *every* base (i.e., non ranked) metrics defined in (Graf et al., 2017). However, since the relationship between complexity metrics and the structure of a specific derivation tree is subtle, a detailed discussion of why each metric fares the way it does is not feasible within the scope of this paper. In what follows, I focus on two metrics that have been noted in previous studies as consistent predictors of processing difficulty: MAXT and SUMS.

The fact that MAXT (SRC: 8/*che*; ORC: 11/*ha*; ORCp:16/*Foc*) succeeds in predicting the reported processing preferences is encouraging, given the past success of this metric on many different constructions.[5] In particular, observe how the string-driven traversal strategy of the MG parser makes tenure sensitive to minor structural differences. In the SRC, *che* is introduced at step 15. Since, based on information in the input string, the parser is looking for the the subject DP *il cavallo*, *che* has to be kept in memory until the latter is found. Thus, it is 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 CFG top-down strategy, ignore the subject DP, and keep expanding nodes until *il cavallo* is found. Thus, *che* cannot be flushed from memory until step 25.

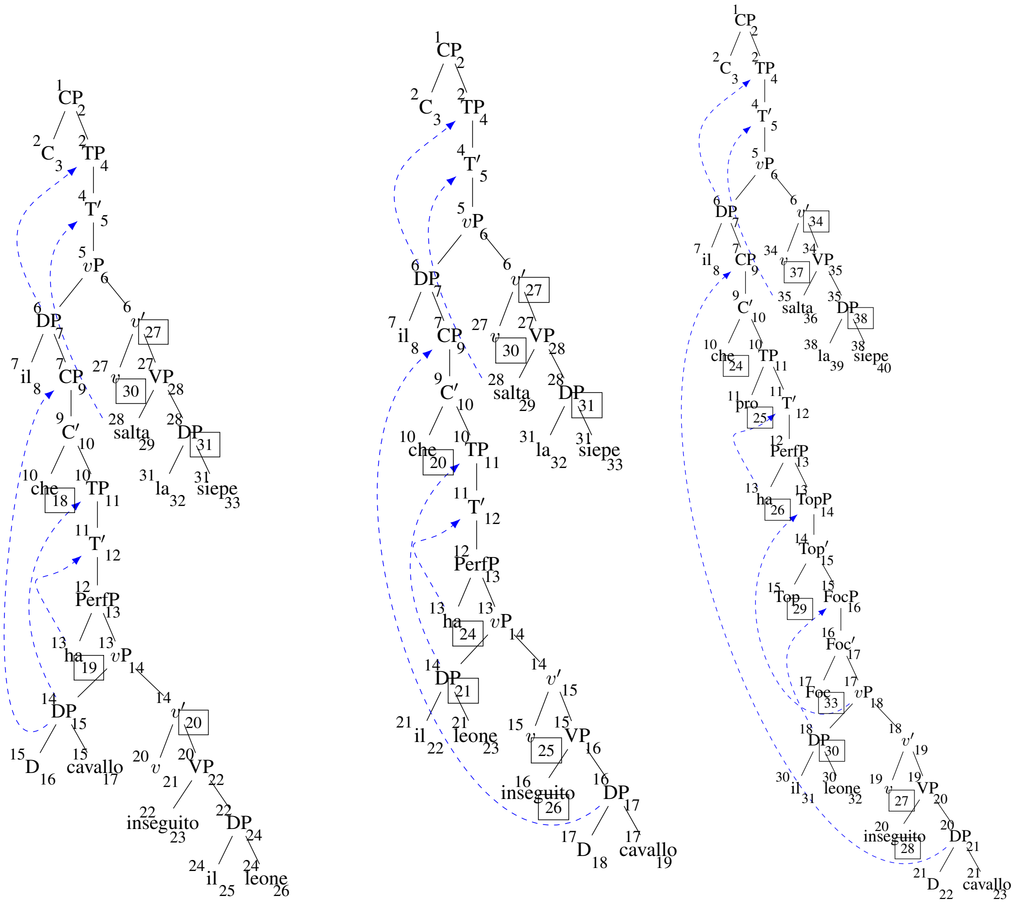
The difference between SRC and ORC also highlights how tenure interacts with movement. Once *che* has been found in the SRC tree, the next node in the stack is *ha*, which can be discharged from memory immediately after. In the ORC however, the parser still has to find the subject DP. Thus, *ha* has 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 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

[3] Technically, Belletti & Leonini (2004) assume that VP, not vP, raises to Spec,Topic. 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 vP cluster to Topic. This also seems to be in the spirit of what suggested by Belletti and Contemori (2009). But note that the modeling results in the following section would remain mostly unchanged even if we were to leave the vP shell in its base position, while both verb and object raise above.

[4] https://github.com/CompLab-StonyBrook/mgproc.

[5] These predictions hold even if we ignore tenure on unpronounced nodes — as suggested by Graf et al. (2017) — since we would obtain (SRC: 8/*that*; ORC: 11/*has*; ORCp:14/*that*.



(a) (b) (c)

Figure 3: Annotated derivation trees for left-embedding (a) SRC (b) ORC and (c) ORCp.

queue.

4.2 Additional Simulations

From one side, the successful predictions made by MAXT are a welcome result, as they confirm the sensitivity of tenure-based metrics to fine-grained structural details. From the other though, one might wonder exactly how much these differences depend on the specific case study we are modeling. In this section, I partially address this issue by looking at variations in the construction of the RCs, and at two more processing asymmetries involving Italian post-verbal subjects.[6] I return to the general issue of the sensitivity of the MG results to syntactic choices in Sec. 5.

4.2.1 Left-Embedding RCs

Due to the string-driven nature of its traversal strategy, the MG parser is peculiarly sensitive to the depth of left- vs. right-embedding constructions. To control for this, I tested the parser predictions on sentences of the form *Il cavallo [$_{RC}$ che ...] salta la siepe* (*The horse [$_{RC}$ that ...] jumps the fence*, Fig. 3), with the head noun raising to the *subject* position in the matrix clause. Here, MAXT predicts that SRC and ORC should have the same processing complexity (they *tie*), since their memory differences are flattened by the increased tenure on the matrix *v'* (the Merge node expanding the matrix vP). The tenure of this node depends on the size of the matrix subject — thus, on the size of the relative clause. Since the size of the SRC and of the ORC is the same (the only thing changing being the site of extraction), MAXT for the whole sentence will never vary between the two constructions. This issue is solved by SUMS, which correctly predicts SRC $<$ ORC, as well as the SRC/ORC $<$ ORCp contrast.

Interestingly, MAXT also correctly predicts the increased difficulty of ORCps in these left-embedding cases. As seen above, MAXT flattens the differences in clauses with subject-modifying SRC/ORCs because the size of the RCs in subject

[6] Trees for these simulations can be found in Appendix A.

position is identical. This is not the case for ORCps, due the sequence of projections and movement steps involved in deriving postverbal subjects from the base SVO order. Thus, while MAXT in these sentences in still measured on the matrix *v'* (28), this value is also picking up on the additional steps required to derive the internal structure of the ORCp construction.

4.2.2 Postverbal Subjects in Matrix Clauses

In order to understand the complexity of the grammatical assumptions made for the postverbal subjects, we can look at processing asymmetries of postverbal constructions outside of RC environments. Consider Italian declarative sentences with a lexically empty subject position, like in (5).

(5) Ha chiamato Gio
Has called Giovanni
a. "He/she/it called Gio" **SVO**
b. "Gio called" **VS**

Without contextual/discourse cues, sentences like (5) are structurally ambiguous between a null-subject interpretation (5a) and a postverbal subject one (5b), with a marked processing preference for (5a) as compared to (5b) (De Vincenzi, 1991).

As summarized in Tbl. 1, both MAXT and SUMS predict the correct preferences under Belletti and Leonini (2004)'s analysis, as the Top and Foc heads have to wait for the whole vP to be found, before they can be discharged from memory themselves (cf. Fig. 4 and Fig. 5).

4.2.3 Unaccusatives vs. Unergatives

Finally, it is interesting to look at declarative sentences containing intransitive verbs of two classes: unaccusatives (6) and unergatives (7).

(6) È arrivato Gio
Is arrived Gio
"Gio arrived" **Unaccusative**

(7) Ha corso Gio
Has ran Gio
"Gio ran" **Unergative**

While on the surface these sentences look very similar, they differ in that the subject originates in postverbal position for unaccusatives but in preverbal position for unergatives (Belletti, 1988). Importantly, De Vincenzi (1991) reports faster reading times and higher comprehension accuracy for (6) over (7), a preference that is again correctly captured both by MAXT and SUMS (cf. Fig. 6 and Fig. 7). In particular, due to unaccusative subjects being 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).

Clause Type	MaxT	SumS
obj. SRC	8/*che*	18
obj. ORC	11/*ha*	24
obj. ORCp	16/*Foc*	31
subj. SRC	21/v'	37
subj. ORC	21/v'	44
subj. ORCp	28/v'	56
matrix SVO	3/*ha*/v'	7
matrix VOS	7/*Top*/*Foc*	11
VS unacc	2/vP	3
VS unerg	7/*Top*/*Foc*	11

Table 1: Summary of MAXT (*value*/*node*) and SUMS by construction. Obj. and subj. indicate the landing site of the RC head in the matrix clause.

5 Discussion

The success of a top-down parser in modeling the processing difficulties of Italian RCs adds support to the MG model as a valuable theory of how processing cost is tied to structure.

As some reviewers point out though, one potential concern with the plausibility of the approach is in the degrees of freedom that are left to the model. In particular, the processing predictions depend on the interaction of three factors: the parsing strategy, the syntactic analysis, and the complexity metrics. Here, I put aside the choice of parsing strategy (but see Hunter, 2018; Stanojević and Stabler, 2018), and briefly address concerns about the remaining two factors.

Due the large number of existing metrics, it is conceivable that some combination of syntactic analysis and metric could have explained any other processing ranking among sentences. Similarly, it is possible that any syntactic analysis would make the right (i.e., empirically supported) predictions with some metric. Both these possibilities would undermine the relevance of this kind of modeling. Luckily, this does not seem to be the case. In fact, 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 (Graf et al., 2017). Thus, it seems that underspecification is not an issue in practice.

The results in this paper are indeed consistent with these observations, as they show SUMS as a reliable complexity metric. Importantly, as subject-modifying SRCs and ORCs only *tie* on MAXT, these findings are also consistent with Graf et al. (2017)'s hypothesis that SUMS should be used a secondary metric to adjudicate between constructions, after they tie on MAXT.[7]

A second, reasonable concern is how much the correct predictions depend on the specific syntactic analysis of choice. Due to the richness of existing analyses and to space constraints, in this paper I only considered an analysis of Italian RCs and postverbal constructions which had been extensively referred to in the psycholinguistic literature. To partially address this concern though, I showed how SUMS and MAXT not only make the right predictions for RC constructions under a few different syntactic configurations, but they also correctly account for postverbal subject asymmetries in different kind of sentences. Nonetheless, an important future enterprise will be to look at alternative approaches to postverbal subject configurations, such as *right dislocation* (Antinucci and Cinque, 1977; Cardinaletti, 1998), or *leftward scrambling* (Ordóñez, 1998). Note though that these analyses all assume additional movement dependencies in the structure of ORCps compared to clauses with preverbal subjects. Given what this paper taught us about SUMS and MAXT, it seems probable that such dependencies would also be picked up by these metrics.

Independently on the specific predictions of the parser for alternative analyses though, the contributions of this line of inquiry would be twofold. From one side, it will improve our understanding of the MG model, 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. Secondly, grounded in the discriminative power given to MAXT and SUMS by their success across empirical phenomena, comparing the predictions made by the parser for different analyses of the same construction might help adjudicate between competing theoretical assumptions, as was the original goal of Kobele et al. (2013).

Clearly, the fact that the parser relies on an idealized deterministic search strategy is one of the (potentially) most contentious assumption of the MG model, and could thus be used as yet another objection to the plausibility of the linking theory. As already mentioned, the goal is not to claim this as a comprehensive model of processing difficulty, as a cognitively realistic theory would see multiple factors interact with each other to derive the correct contrasts (Demberg and Keller, 2008; Brennan et al., 2016, a.o.). In principle though, the MG parser can be integrated with several of these additional factors (e.g., uncertainty; Hunter and Dyer, 2013; Yun et al., 2015). Crucially, the main advantage of the MG model is its transparent specification of the parser's behavior, which clarifies the effects of structural complexity on memory burden and would allow us to separate them from other effects contributing to processing load.

Moreover, while uncertainty is clearly a fundamental component of the human sentence processing system, the fact that an account deliberately abstracting away from all ambiguity can explain effects that would usually be attributed to it is an intriguing result. A fascinating open question is then whether we can characterize those phenomena where ambiguity really is the decisive factor, and cannot be "eliminated" from the model.

Finally, another advantage of having a computational model which provides a testable link between syntactic theory and behavioral data, is that it allows us to formally integrate structural hypotheses in existing psycholinguistic theories in a way that leads to precise quantitative predictions. However, as one reviewer observes, the complexity metrics exploited by the MG parser rely on very weak assumptions about the nature of human memory. In a sense, this could be considered a perk, as it leaves the model open to connections with a variety of sentence processing theories. In another sense though, this lack of cognitive plausibility weakens the impact of the approach, as it is often difficult to connect its results to more general concerns in the sentence processing literature. An important future research direction will thus be to re-evaluate the existing complexity metrics in light of psychological insights about human memory mechanisms (cf. Zhang, 2017).

Acknowledgments

I am extremely grateful to four anonymous reviewers for their insightful feedback.

[7] SUMS by itself does not seem to be enough, as it fails to predict the right preferences for contrasts like English right vs. center embedding (Graf et al., 2017).

References

Francesco Antinucci and Guglielmo Cinque. 1977. Sull'ordine delle parole in italiano: l'emarginazione. *Studi di grammatica italiana VI, pp. 121-146.*

Fabrizio Arosio, Flavia Adani, and Maria Teresa Guasti. 2009. Grammatical features in the comprehension of Italian Relative Clauses by children. *Merging Features: Computation, Interpretation, and Acquisition*, pages 138–158.

Fabrizio Arosio, Francesca Panzeri, Bruna Molteni, Santina Magazù, and Maria Teresa 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).

Elizabeth Bates and Brian MacWhinney. 1987. Competition, variation, and language leaning. *Mechanisms of language acquisition*.

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

Adriana Belletti and Carla 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*, pages 39–52.

Adriana Belletti and Chiara Leonini. 2004. Subject inversion in L2 Italian. *EUROSLA yearbook*, 4:95–118.

Jonathan R Brennan, Edward P Stabler, Sarah E Van Wagenen, Wen-Ming Luh, and John T Hale. 2016. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, 157:81–94.

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

Noam Chomsky. 1977. On wh-movement. *Formal syntax*, pages 71–132.

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

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

Vera Demberg and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193 – 210.

Ulrich Hans Frauenfelder, Juan Segui, and Jacques Mehler. 1980. Monitoring around the relative clause. *Journal of Verbal Learning and Verbal Behavior*, 19(3):328–337.

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

Naama Friedmann, Adriana Belletti, and Luigi Rizzi. 2009. Relativized relatives: Types of intervention in the acquisition of a-bar dependencies. *Lingua*, 119(1):67–88.

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

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

Edward Gibson. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, pages 95–126.

Thomas Graf, Brigitta Fodor, James Monette, Gianpaul Rachiele, Aunika Warren, and Chong Zhang. 2015. A refined notion of memory usage for minimalist parsing. In *Proceedings of the 14th Meeting on the Mathematics of Language (MoL 2015)*, pages 1–14, Chicago, USA. Association for Computational Linguistics.

Thomas Graf and Bradley Marcinek. 2014. Evaluating evaluation metrics for minimalist parsing. In *Proceedings of the 2014 ACL Workshop on Cognitive Modeling and Computational Linguistics*, pages 28–36.

Thomas Graf, James Monette, and Chong Zhang. 2017. Relative clauses as a benchmark for Minimalist parsing. *Journal of Language Modelling*, 5:57–106.

John T Hale. 2011. What a rational parser would do. *Cognitive Science*, 35(3):399–443.

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

Tim Hunter. 2018. Formal methods in experimental syntax. *The Oxford Handbook of Experimental Syntax*.

Tim Hunter and Chris Dyer. 2013. Distributions on minimalist grammar derivations. In *Proceedings of the 13th Meeting on the Mathematics of Language (MoL 13)*, pages 1–11.

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

Richard S Kayne. 1994. *The antisymmetry of syntax*. 25. MIT Press.

Jonathan W King and Marta 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.

Gregory Kobele, Christian Retoré, and Sylvain Salvati. 2007. An automata-theoretic approach to minimalism. In *Model Theoretic Syntax at 10*, pages 73–82. J. Rogers and S. Kepser.

Gregory M Kobele, Sabrina Gerth, and John Hale. 2013. Memory resource allocation in top-down minimalist parsing. In *Formal Grammar*, pages 32–51. Springer.

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

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

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

Lei Liu. 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.

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

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

Alan Prince and Paul Smolensky. 2008. *Optimality Theory: Constraint interaction in generative grammar*. John Wiley & Sons.

Owen Rambow and Aravind K Joshi. 1994. A processing model for free word order languages. *Perspectives on Sentence Processing*.

Herbert Schriefers, Angela D Friederici, and Katja Kuhn. 1995. The processing of locally ambiguous relative clauses in german. *Journal of Memory and Language*, 34(4):499.

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

Edward P Stabler. 2011. Computational perspectives on minimalism. In *The Oxford Handbook of Linguistic Minimalism*.

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

Miloš Stanojević and Edward Stabler. 2018. A sound and complete left-corner parsing for minimalist grammars. In *Proceedings of the Eight Workshop on Cognitive Aspects of Computational Language Learning and Processing*, pages 65–74.

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

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

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

Jiwon Yun, Zhong Chen, Tim Hunter, John Whitman, and John Hale. 2015. Uncertainty in processing relative clauses across East Asian languages. *Journal of East Asian Linguistics*, 24(2):113–148.

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

A Appendix

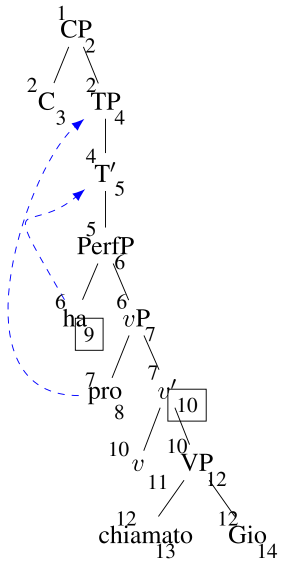


Figure 4: Annotated derivation tree for the SVO sentence in (5a)

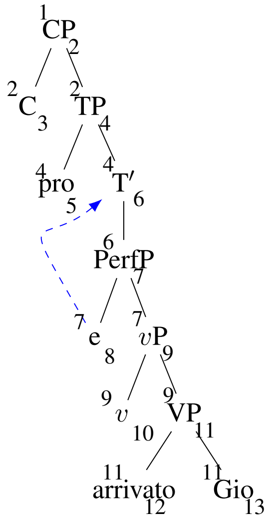


Figure 6: Annotated derivation trees for the unaccusative sentence in (6)

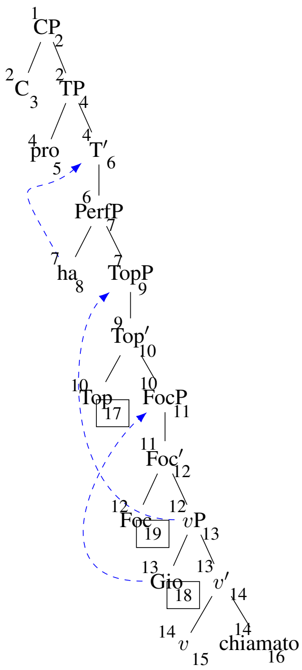


Figure 5: Annotated derivation trees for the VS sentences in (5b)

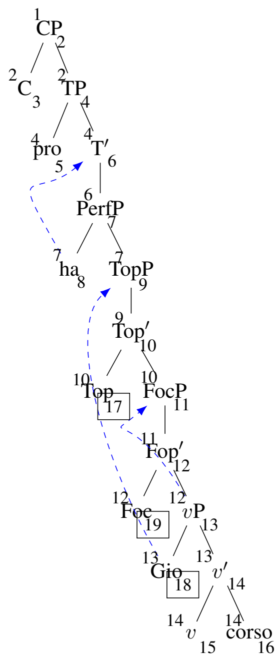


Figure 7: Annotated derivation trees for the unergative sentence in (7)