Evaluating Structural Economy Claims in Relative Clause Attachment

Aniello De Santo
Department of Linguistics
University of Utah
aniello.desanto@utah.edu

So Young Lee
Department of English
Miami University
lees44@miamioh.edu

Abstract

Grillo and Costa (2014) argue for a pseudo-relative (PR) first account of relative clause attachment preferences (RC) such that, when faced with a sentence ambiguous between a PR and a RC interpretation, the parser prefers committing to a PR structure first, thus giving rise to what looks like a high-attachment preference. One possible explanation for this parsing choice is in terms of simplicity of the PR structure, and overall economy principles. Here, we evaluate this hypothesis by testing the predictions of a parser for Minimalist grammars for PR and RC structures in Italian. We discuss the relevance of our results for PR-first explanations of the cross-linguistic variability of RC attachment biases, and highlight the role that computational models can play in evaluating the cognitive plausibility of economy considerations tied to fine-grained structural analyses.

1 Introduction

The idea that economy and simplicity principles affect syntactic derivations has been central to inquiries in Generative grammar. In earlier iterations, economy conditions were basically conceived as evaluation metrics for selecting grammars from the format permitted for rule systems. In a lexicalized, non-rule based framework such as the Minimalist Program, economy has come to play a different but still central guiding role in the theoretical architecture of the grammar — motivating, for instance, the preference for applying some grammatical operations over others (Chomsky, 1995; Collins, 2001).

These appeals to economy considerations, ubiquitous even in the most recent syntactic literature, occasionally reference general *parsing* and computational motivations (Kayne, 1994; Motut, 2010; Fukuda, 2011; Razaghi et al., 2015; Bošković and Messick, 2017, a.o.). In this sense, a mutual exchange of questions and insides across the syntactic and psycholinguistic literature has been fruitful, inspiring a vast array of research questions. For instance, there are numerous detailed formalizations of the role of *locality* considerations in our understanding of grammatical and processing principles (Frazier, 1987, 1978; De Vincenzi, 1991; Gibson, 2000). However, within the theoretical literature there is sometime the tendency to rely on economy explanations without overtly specifying what kind of assumptions are made about fine-grained syntactic details and their relation to broader principles of *cost*. For example, while it is possible to find many claims of *structural simplicity* made to motivate syntactic and/or psycholinguistic predictions, it is often unclear in these contexts how *simplicity* is actually quantified, how these computational demands would be implemented in a precise parsing architecture, and how these costs are linked to cognitive resources. Ideally, it would be desirable to formally spell-out the kind of complexity assumptions underlying different aspects of syntactic representations, so to explore the plausibility of the predictions made by economy claims with respect to behavioral responses in psycholinguistic studies (Bresnan, 1978, 1982; Rambow and Joshi, 1994; Kobele et al., 2013; Demberg and Keller, 2009).

Following these intuitions, in this paper we focus on economy principles as referenced in the context of the cross-linguistic variation of relative clause attachment ambiguity preference. We suggest that a transparently specified computational model which takes syntactic assumptions seriously can help shed light on these issues. In particular, we propose the use of a parser for Minimalist grammars (MGs; Stabler, 2013), coupled with complexity metrics measuring memory usage (Kobele et al., 2013; Gerth, 2015; Graf et al., 2017, a.o.), in order to investigate the predictions of the so-called *pseudo-relative first* hypothesis (Grillo and Costa, 2014) in a framework that actually formalizes

Proceedings of the Society for Computation in Linguistics (SCiL) 2022, pages 65-75.
Held on-line February 7-9, 2022

economy considerations. As a starting point in this enterprise, we evaluate the predictions of the model for the processing preferences reported in recent attachment ambiguity studies for Italian (De Vincenzi and Job, 1993; Grillo and Costa, 2014).

2 Parsing Principles and RC Attachment Preferences

One of the most researched topics in the sentence processing literature is the cross-linguistic variation in *attachment ambiguity* preferences. Notoriously, when a complex Determiner Phrase (e.g., DP_1 *of* DP_2) is followed by a RC, languages are known to show varying biases for the RC modifying either DP_2 (Low Attachment, **LA**) or DP_1 (High Attachment, **HA**). Consider the following sentence:

(1) Pearl saw the Commander of the Gem that run

a. Pearl saw the Commander of [the Gem that run] **LA**
b. Pearl saw [[the Commander of the Gem] that run] **HA**

This sentence is ambiguous between two interpretations. In the LA interpretation, the relative clause [*that run*] modifies the second DP: e.g. in (1a), it is *the Gem* that is doing the running. However, a HA interpretation is also available (1b), according to which it is *the commander* that was running, with the RC modifying the whole complex DP [*the Commander of the Gem*]. While it is well established that English speakers will generally prefer the LA interpretation, it has been shown that languages vary significantly in this respect (Cuetos and Mitchell, 1988). For instance, Spanish, Greek, and Italian speakers show a general preference for a HA interpretation (Cuetos and Mitchell, 1988; Carreiras and Clifton Jr, 1993; De Vincenzi and Job, 1993; Papadopoulou and Clahsen, 2003, a.o.), while in Basque and Chinese speakers pattern similarly to English (Gutierrez-Ziardegi et al., 2004; Shen, 2006, a.o.). Additionally, variation in attachment preference within the same language has also been reported (Fernández, 2003), as well as variation across online and offline tasks (De Vincenzi and Job, 1993).

This cross-linguistic variation in Relative Clause (RC) attachment preferences has been the object of extensive investigation in theoretical linguistics and psycholinguistics. A long-standing hypothesis in the sentence processing literature has been that *processing economy* principles are a core feature of the human parser. In this sense, RC attachment ambiguity is of interest as a perfect case study for the exploration of such general economy considerations. As mentioned, such ambiguity is due to the possibility of attaching the RC to either the first DP or the second DP. An intuitive interpretation of locality of structure building would favor the latter, under the assumption that local attachment reduced the processing load of the parser (Frazier, 1990; Gibson et al., 1996; Gibson, 1998). While LA languages perfectly conform to the predictions made by such an hypothesis, HA languages present a problem. Importantly, numerous studies have unveiled a variety of factors that can modulate RC attachment — such as prosodic, semantic, and pragmatic variables (MacDonald et al., 1994; Gilboy et al., 1995; Acuna-Farina et al., 2009; Fernández, 2005; Fraga et al., 2005; Hemforth et al., 2015). These additional variables seem to behave somewhat consistently across languages, and the many existing proposals in the literature (Cuetos and Mitchell, 1988; Clifton Jr and Frazier, 1996; Gibson et al., 1996; Hemforth et al., 2000, a.o.) still leave the full pattern of cross-linguistic variation somewhat unexplained. If the goal is to provide explanatory insights into parsing mechanisms, even accounts that try to reduce variation in cross-linguistic preferences to statistical/exposure distributions would need to address whether HA/LA is less frequent in a specific language because of the inherent complexity of one construction over the other, or because of some external reason. All else being equal then, if the variance in the interpretative biases for RC attachment is not to be located in language specific grammatical distinctions, it might pose an issue for theories of language processing that see universal parsing mechanisms underlying human language processing behavior (see Grillo and Costa, 2014; Grillo et al., 2015; Aguilar et al., 2021, for a discussion).

While acknowledging the complicated array of variables affecting RC interpretation, Grillo and Costa (2014) point out a possible confounding factor in previous experiments reporting HA in languages like Italian and Spanish: namely, the availability of a *pseudo-relative* interpretation. Their claim is that in HA languages there is an additional structural representation available for sentences like (1): a *pseudo-relative* clause (PR) construction. Although linearly identical to RCs, PRs have different structural and semantic properties — essentially, they behave as NP/DP modifiers denoting events.

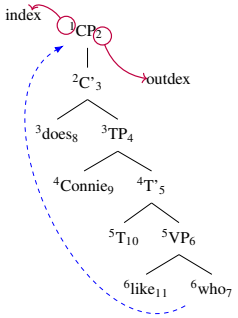
Importantly, the main structural difference is that in a PR parse, the matrix verb takes the whole PR as its complement (akin to what happens for English Small Clauses), and what looks like the "modified" DP is the subject of that clause. As PRs are complement/adjuncts of VPs, the most local DP is not grammatically available to the PR, and thus they are only compatible with what looks like a HA interpretation. Interestingly, it is possible to control for PR availability by modulating the syntactic and semantic environment of a sentence. With these considerations in mind, Grillo and Costa (2014) report that when participants are tested with sentences for which the RC interpretation is the only possible one (i.e., PR is made unavailable based on the properties of the main clause verb), a LA parse is preferred over the HA one (see also De Vincenzi and Job, 1993; Branco-Moreno, 2014; Aguilar et al., 2021).

These facts are accounted for by formulating a **pseudo-relative first** hypothesis. This hypothesis states that, when faced with a sentence ambiguous between a PR and a RC interpretation, the parser prefers committing to a PR structure first, thus giving rise to what looks like a HA preference. Then, if a PR analysis is made unavailable, the parser will prefer the LA parse over the HA due to universal locality principles. Grillo and Costa (2014) argue that the parsing preference for PR constructions might be due to the richer functional domain usually associated to RCs, making the latter dispreferred.

The preference of the parser for a PR structure is thus accounted for in this literature in terms of *simplicity* of the PR structure, and overall *economy* principles. However, while the locality ideas that would lead to preferring a LA over a HA have been extensively discussed in the past, the specific parsing principle grounding the alleged PR vs. RC complexity asymmetry is left generally unspecified. While the idea that structure building operations correspond to some type of cognitive cost is certainly not new, it is unclear that simply postulating additional functional structure *per se* implies increased parsing cost (Miller and Chomsky, 1963; Bresnan, 1978). In fact, it is possible to conceive of pletora of ways in which a specific structure could be defined as being simpler than another, and none of these are guaranteed to have concrete effects on a specific parsing strategy (Bresnan, 1982; Berwick and Weinberg, 1983). If such hypotheses are to be thoroughly explored, it seems crucial to ground our theoretical stipulations in a transparent theory of exactly why certain operations are more costly for the parser than others. In the rest of the paper, we propose the use of a computational model grounded in a rich grammar formalism, as a way to evaluate the economy claims made by the PR-first hypothesis. The following section illustrates the core ideas behind the model, and clarifies why such an approach can offer insights when testing theories of structural and processing complexity. For recent, detailed overviews of the technical details of the approach the reader is referred to (Gerth, 2015; Graf et al., 2017; De Santo, 2020b).

Steps	Parse Action
1	CP is conjectured
2	CP expands to C'
3	C's expands to does and TP
4	TP expands to Connie and T'
5	T' expands to T and VP
6	VP expands to like and who
7	**who** is found
8	**does** is found
9	**Connie** is found
10	**T** is found
11	**like** is found

(a)



(b)

Figure 1: Example of a string-driven top-down tree traversal for an MG derivation tree.

3 MG Parsing

MGs (Stabler, 1996, 2011) are a lexicalized formalism rigorously implementing an early version of Minimalist syntax. These grammars consist of a sets of lexical items (LIs), each with 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. In this paper, we will ignore the feature component of the LIs, and focus on the fact that 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).

Merge and Move operations are represented in these trees as binary and unary branching nodes, respectively. The main difference between a more traditional phrase structure tree and a derivation tree is that in the latter, the final word order of a sentence is not directly reflected in the order of the leaf nodes in a derivation tree. This is because moving phrases remain in their base position, and their landing site can be deterministically reconstructed via the

feature calculus.

Given that MGs are able to represent the structurally rich analyses now common in Minimalist syntax, they have been focus of a line of work aimed at connecting syntactic assumptions to offline processing behavior. Specifically, this work has shown that a top-down parser for MGs (Stabler, 2013) can successfully predict a variety of processing difficulty contrasts, via metrics that relate offline parsing difficulty to memory usage (Kobele et al., 2013; Graf et al., 2017; De Santo, 2020b, a.o.).

Stabler (2013)'s parser is a variant of a standard recursive-descent parser for CFG, modified to take care of the fact that the order of lexical items in a derivation tree does not fully match the linear surface order. 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. In order to keep track of the derivational operations affecting the linear word order, the MG variant follows the standard approach of predicting nodes downward (toward words) and left-to-right until a Move node is predicted. At that point, the parser discards the top-down strategy and builds the shortest path towards the predicted mover. After the mover has been found, the parser continue from the point where the search for the mover started (Figure 1a). The memory stack associated to the parser therefore plays a fundamental role: 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. For instance, considering the derivation tree in Figure 1b, the node for *does* is predicted at step 3. However, since a movement dependency for Spec,CP has been postulated, the parser is not following a pure top-down strategy and will not match that prediction against the linear input until a node for *who* has been predicted and confirmed (at step 6 and 7).

To make the traversal strategy easy to follow, we 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). Index and Outdex allow the *MG model* to rigorously link parser behavior, syntactic structure, and processing difficulty by connecting the stack states of the top-down parser to memory usage. In order to allow for psycholinguistic predictions, it is then possible to use these annotations to predict processing difficulty based on how the structure of a derivation tree affects memory usage during a parse (Rambow and Joshi, 1994; Gibson, 2000; Kobele et al., 2013; Graf and Marcinek, 2014; Gerth, 2015).

The MG model distinguishes several cognitive notions of memory usage (Graf et al., 2017). Here, we focus on a measure of how long a node is kept in memory through a derivation (TENURE). Tenure for each node is computed considering the moment a node was first postulated into the structure (and thus placed in the memory stack of the parser) and the moment such prediction was confirmed and the node could be taken out of memory. Essentially then, a node's tenure is equal to the difference between its index and its outdex. For instance, considering the annotated MG tree in Figure 1b, tenure for *Connie* is $Outdex(Connie) - Index(Connie) = 9-4 = 5$.

Based on how this cognitive notion of memory usage interacts with the geometry of the underlying syntactic structure, the MG parser then assigns a cost to each sentence. 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. A metric like MAXT can then be used to derive categorical processing contrasts, by comparing the tenure values assigned by the MG model to derivation trees corresponding to sentences with stark asymmetries in reported offline processing preferences. For instance, building on these intuitions, Graf and Marcinek (2014) show that MAXT 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 cross-linguistic contrasts involving relative clauses. Importantly, while the space of possible metrics defined by this model is potentially vast, in what follows we will focus our discussion on MAXT exclusively, given the attention that this specific metric has received in recent work (Gerth, 2015; Graf et al., 2017; Liu, 2018; Lee, 2018; De Santo, 2019, 2020a).

Finally, note that Stabler's original parser is

equipped with a search beam discarding the most unlikely predictions. Consistent with previous work, we 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, and might seem controversial when modeling structurally ambiguous sentences. However, it is made with a precise purpose in mind: by assuming a deterministic parse, we aim to evaluate structural economy claims by focusing on the specific contribution of syntactic complexity to memory load.

4 PRs vs RCs in a Computational Model

Consider now the following Italian sentence:

(2) (Io) Ho visto la nonna della ragazza
(I) have seen the grandma of the girl
che gridava
that screaming-3SG

"I saw the grandma of the girl that was screaming"

This sentence is ambiguous between a HA interpretation (*the grandma was screaming*) and a LA interpretation (*the girl was screaming*). Additionally, the HA interpretation is ambiguous between two structural analyses: a PR analysis and a true HA, RC analysis.

As mentioned, the *pseudo-relative first* hypothesis as stated in previous literature predicts that a PR parse should be preferred over RC parses (both LA and HA), due to the overall simplicity of PRs over RCs. Additionally, the hypothesis then predicts that, in absence of an available PR parse, a LA parse should be preferred over an HA one, possibly due to locality principles. The relevant pairwise contrasts are summarized in Table 1. Note that, while the PR $<$ LA[1] contrast might seem counter intuitive, it is crucial for the PR-first hypothesis to pan out: when faced with a choice, the parser follows a PR strategy first as it is (in some ways) simpler. A conceivable weaker version of the hypothesis, which makes a prediction only for the HA vs. PR contrast with nothing to say about the relative simplicity of the PR structure when compared to relative clauses with LA constructions would be insufficient, as it would not explain why the parser does not follow a LA strategy to begin with. With this in mind, we test this hypothesis over sentences in Italian as reported in (De Vincenzi and Job, 1993; Grillo and Costa, 2014).

[1]Henceforth in the paper, processing contrasts are summarized as x $<$ y, to be interpreted as x is preferred over y.

4.1 Syntactic Choices

Adopting MGs as the core grammar formalism makes the model sensitive to fine-grained syntactic choices. Exploring how different syntactic analyses impact the main results is thus important to the explanatory aims of the approach (De Santo, 2021). In this sense, the psycholinguistic literature tends to be fairly non-committal with respect to the details of the structural hypotheses underlying relative clauses. Therefore, here we evaluate two different analyses of RC constructions currently popular in minimalist syntax (Bianchi, 2002a,b): the promotion analysis (Kayne, 1994) and the *wh*-movement analysis (Chomsky, 1977).

Promotion Analysis Under a promotion analysis (Kayne, 1994), the head of the RC is a noun starting out as an argument of the embedded verb and undergoing movement into the specifier of the RC. The RC itself is selected by the determiner that would normally select the head noun in head-external accounts, like the *wh*-movement case below (Figure 2a).

***Wh*-movement Analysis** Chomsky (1977)'s *wh*-movement analysis treats the construction of an RC as an instance of *wh*-movement. The complementizer position is overtly filled by *that*, while a silent wh-operator *Op* moves from the base position to Spec,CP. The whole CP merges with the relativized NP as its adjunct (Figure 2b). The silent *Op* is co-indexed with the NP to which the RC is adjoining.

Pseudo-relatives Similarly to RCs, there are various potential analyses to pseudo-relative clause constructions. Here, we follow Grillo and Costa (2014) and adopt an approach to PR structures as small clauses (Cinque, 1992). Essentially, as mentioned before, in PR parses the matrix verb takes the whole PR as its complement, and the modified DP is the head of that clause (Figure 2c). Thus, there is no movement extracting the head DP from within the PR. The modified DP is linked to its interpreted position by co-indexing it with a null *pro*, resembling what is done with RC in the *wh*-movement analysis.

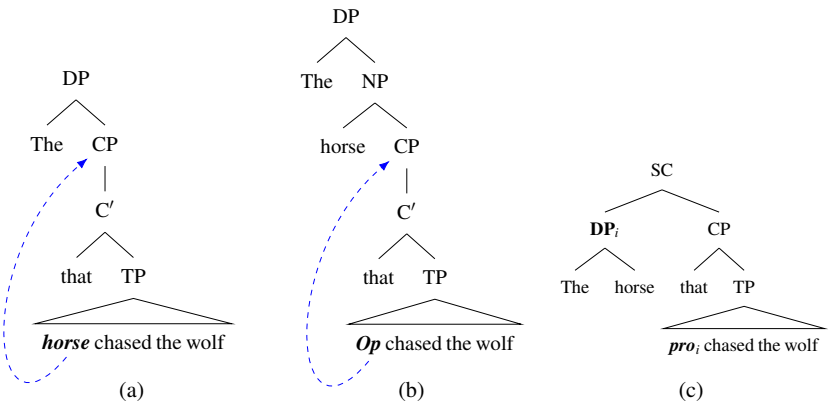


Figure 2: Sketches of the *(a)* RC with promotion, *(b)* RC with *wh*-movement, and *(c)* PR analyses for the sentence *The horse that the wolf chased*.

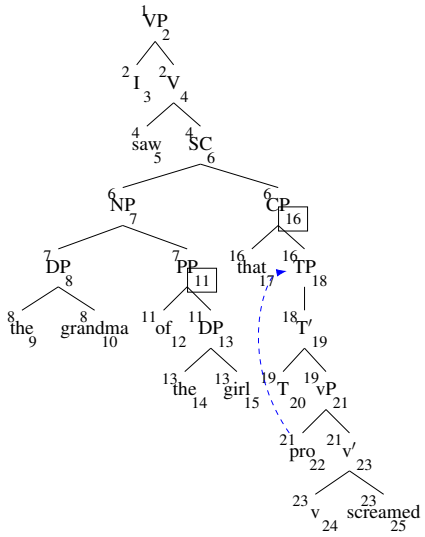


Figure 3: Annotated derivation trees for the Italian sentence *I saw the grandma of the girl that screamed*, according to a pseudo-relative clause analysis. The root of the tree is treated as a VP since additional structure in the matrix clause would be identical across comparisons. Boxed nodes are those with tenure value greater than 2, following (Graf and Marcinek, 2014).

4.2 Modeling Results[2]

As the model's results are categorical, only one test item per construction is needed. Consider once again the ambiguous sentence in (4). According to what discussed in the previous section, we can build five derivations for that single linear string: one derivation using a PR analysis, and then two derivations for RC/LA and RC/HA, each modulated by syntactic analysis (*wh*-movement or promotion). Annotated derivation trees for these configurations can be seen in Figure 3 and Figure 4. With all of this in place, we can finally look at the modeling results. Table 1 and Table 2 report overall performance of the model, and MAXT values for the three constructions considered here.

First, MAXT successfully captures the $LA < HA$ preference, independently of syntactic analysis. This is because in the HA cases the parser has to search for the whole NP [*the grandma of the girl*] before being able to work on the rest of the RC (Figures 4a and 4b vs. Figures 4c and 4d). This is encouraging, in the sense that it shows how the model captures the well-established intuition about locality of attachment for these two constructions.

We can then move on to the pseudo-relative contrasts. Under a promotion analysis, the parser correctly captures $PR < HA$, due to the additional movement dependencies hypothesized for the RC/HA structure (Figure 4a). However, MAXT predicts no difference between the two when the RC is built according to a *wh*-movement approach (Figure 4b). Looking at the annotated derivation trees for the HA case, it is possible to infer that in the promotion case MAXT (measured on the complementizer *that*) is driven by the fact that the whole head NP raises to Spec,CP. Thus, the parser needs to expand it in its base position (Spec,*v*P) before being able to work on the rest of the CP. This contrasts starkly with what is done when building the PR structure: since there is no movement dependency to resolve, having to build the big NP first does weight on the CP node somewhat, but it does not affect how long CP internal nodes have to be maintained in memory

[2] 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.

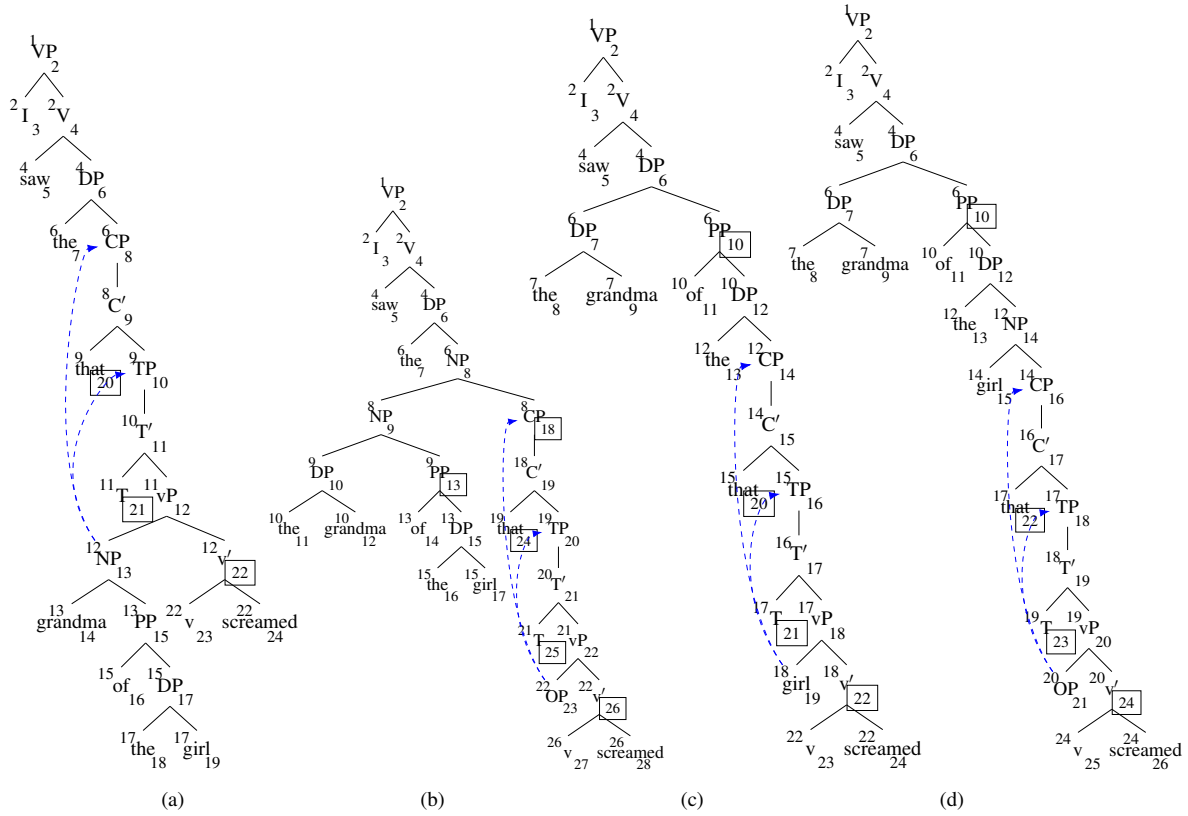


Figure 4: Annotated derivation trees for Italian Relative Clauses: *(a)* HA with a promotion analysis, *(b)* HA attachment with a *wh*-movement analysis, *(c)* LA with a promotion analysis, and *(d)* LA with a *wh*-movement analysis for the sentence *I saw the grandma of the girl that screamed.* Trees are treated as VPs since additional structure in the matrix clause would be identical across comparisons. Boxed nodes are those with tenure value greater than 2, following (Graf and Marcinek, 2014).

(Figure 3). Crucially though, this is very similar to what has to be done for RCs according to the *wh*-movement analysis. According to this approach, there is no movement of the whole NP from within the RC, but just of an operator to Spec,CP. Thus, while there are some subtle structural differences between RCs and PRs under the *wh*-movement analysis too, they do not end up affecting overall memory load in any significant way (beyond the specific node on which MAXT is measured).

Finally, we look at the last contrast relevant to the PR-first hypothesis. Under neither of the RC analyses considered the model is able to capture the fact that a PR construction should be more efficient to parse than a LA attachment RC one. This is because for both PR and HA structures, the parser has to explore the full complex NP before being able to expand on the rest of the structure (thus increasing memory load on the hypothesized embedded CP), while in the LA case only the lower DP needs to be fully built and discarded from memory.

	MG Parser	
Hypothesis	**Promotion**	**Wh-mov**
PR < HA	✓	Tie
PR < LA	×	×
LA < HA	✓	✓

Table 1: Summary of the predictions made by a *pseudo-relative first* account, and corresponding parser's predictions based on MAXT, as pairwise comparisons (x < y: x is preferred over y).

5 Conclusion

In this paper, we exploited a transparent computational model connecting grammatical representations to memory cost via parsing, in order to explicitly test an economy-based hypothesis about why pseudo-relative clauses are preferred over relative clauses in psycholinguistic experiments. This PR-first hypothesis has been put forward in the literature as a way to account for the reported cross-linguistic variation between high-attachment

	MaxT	
	Promotion	**Wh-mov**
PR	10/CP	
HA	11/that	10/CP
LA	5/that	7/that

Table 2: MaxT values (*value*/*node*) by construction, with RCs modulated across a promotion and *wh*-movement analysis.

(HA) and low-attachment (LA) parsing preferences — which would then arise as an artifact of a syntactic difference between languages with pseudo-relative constructions and languages without it. In order to evaluate the broader implications of the PR-first idea then, what seems crucial is the ability to explore the soundness of the complexity predictions made by this hypothesis when interacting with the broad range of fine-grained syntactic assumptions for a minimalist derivation.

Using complexity metrics calculated on Minimalist Grammar derivations for Italian sentences, we showed that a preference for PR over HA is predicted, as well as the more traditional preference for LA over HA, but the required preference for PR over LA is not. Overall then, our results do not support a memory-based, parsing economy explanation for a PR preference in Italian.

Importantly, these modeling results do not call into question the strong experimental evidence for PR availability modulating attachment preferences (De Vincenzi and Job, 1995; Branco-Moreno, 2014; Grillo and Costa, 2014; Grillo et al., 2015; Aguilar et al., 2021, a.o.), and thus do not weaken the analysis of LA/HA variation as an artifact of PR availability per se. In fact, the MG model's predictions are *consistent* with the idea that, when comparing genuine RC structures, a LA derivation should be easier than a HA derivation. What these results invite us to consider however, is the importance of deeper evaluations of "simplistic" explanations of processing facts based on un-specified parsing simplicity principles. While our model might not tell us *why* PRs are preferred over RCs, it suggests ways to narrow down the space of plausible accounts.

Obviously, there are a variety of ways in which simplicity claims can be incorporated into a parsing model (Boston, 2012). Moreover, here we only considered structural differences between PRs and RCs while, as Grillo and Costa (2014) themselves suggest , notions of complexity driven by semantic/pragmatic differences might be playing an important role (Crain, 1985; Altmann and Steedman, 1988).

Cross-linguistic validation is also fundamental in this type of inquiry. Note that under standard assumptions a corresponding Spanish sentence would only differ lexically for the Italian cases, and the numerical contrasts would be virtually identical across the two languages. Thus, while the discussion in this paper was focused on Italian, these results straightforwardly extend to Spanish too. However, in the future it will be important to extend this evaluation to HA languages with wider syntactic differences, for which a PR advantage has been also established experimentally (e.g., French; Koenig and Lambrecht, 1999; Pozniak et al., 2019). In this sense, the pairwise differences needed by the MG parser might also suggest ways to design fine-grained experimental contrasts for languages in which PR availability still lacks experimental support.

Finally, the difference between the performance under a promotion vs. *wh*-movement account highlights once again the model's sensitivity to syntactic details, and reveals how different syntactic choices might affect notions of *simplicity* grounded in parsing intuitions in unexpected ways. Crucially though, what our results reveal is that quantified implementations of simplicity and economy might differ significantly from more broadly specified, general intuitions. Transparent computational models, coupled with more extensive cross-linguistic experimental comparisons, can then play a crucial role in building theories of the interface between grammatical principles and sentence processing mechanisms that are explicit and explanatory.

Acknowledgments

The authors would like to thank SCiL 2022 anonymous reviewers for their helpful suggestions.

References

Carlos Acuna-Farina, Isabel Fraga, Javier García-Orza, and Ana Piñeiro. 2009. Animacy in the adjunction of spanish rcs to complex nps. *European Journal of Cognitive Psychology*, 21(8):1137–1165.

Miriam Aguilar, Pilar Ferré, José M Gavilán, José A Hinojosa, and Josep Demestre. 2021. The actress was on the balcony, after all: Eye-tracking locality and pr-availability effects in spanish. *Cognition*, 211:104624.

Gerry Altmann and Mark Steedman. 1988. Interaction

with context during human sentence processing. *Cognition*, 30(3):191–238.

Robert C. Berwick and Amy S. Weinberg. 1983. The role of grammar in models of language use. *Cognition*, 13:1–61.

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

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

Željko Bošković and Troy Messick. 2017. Derivational economy in syntax and semantics. In *Oxford research encyclopedia of linguistics*.

Marisa Ferrara Boston. 2012. *A computational model of cognitive constraints in syntactic locality*. Ph.D. thesis, Cornell University.

David Branco-Moreno. 2014. *The influence of pseudo-relatives on attachment preferences in Spanish*. City University of New York.

J. Bresnan. 1982. *The Mental representation of grammatical relations*. MIT Press series on cognitive theory and mental representation. MIT Press.

Joan Bresnan. 1978. A realistic transformational grammar. *Linguistic theory and psychological reality*, pages 1–59.

Manuel Carreiras and Charles Clifton Jr. 1993. Relative clause interpretation preferences in spanish and english. *Language and speech*, 36(4):353–372.

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

Noam Chomsky. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Guglielmo Cinque. 1992. The pseudo-relative and acc-ing constructions after verbs of perception. *Working Papers in Linguistics,¡ 2¿, 1992, pp. 1-31*.

Charles Clifton Jr and Lyn Frazier. 1996. Construal.

Chris Collins. 2001. *Economy conditions in syntax*. na.

Stephen Crain. 1985. On not being led up the garden path. *Natural language parsing*, pages 320–358.

Fernando Cuetos and Don C Mitchell. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in spanish. *Cognition*, 30(1):73–105.

Aniello De Santo. 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.

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

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

Aniello De Santo. 2021. Italian postverbal subjects from a minimalist parsing perspective. *Lingue e linguaggio*, 20(2):199–227.

Marcia De Vincenzi and Remo Job. 1995. An investigation of late closure: The role of syntax, thematic structure, and pragmatics in initial interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(5):1303.

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

Marica De Vincenzi and Remo Job. 1993. Some observations on the universality of the late-closure strategy. *Journal of Psycholinguistic Research*, 22(2):189–206.

Vera Demberg and Frank Keller. 2009. A computational model of prediction in human parsing: Unifying locality and surprisal effects. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 31.

E Fernández. 2005. The prosody produced by spanish-english bilinguals: A preliminary investigation and implications for sentence processing. *Revista da ABRALIN*, 4(1):109–141.

Eva Fernández. 2003. Bilingual sentence processing. *Amsterdam: John Benjamins. Ferreira, F.(2003). The misinterpretation of noncanonical sentences. Cognitive Psychology*, 47:164–203.

Isabel Fraga, Javier García-Orza, and Juan Carlos Acuña. 2005. La desambiguación de oraciones de relativo en gallego: Nueva evidencia de adjunción alta en lenguas romances. *Psicológica*, 26(2):243–260.

Lyn Frazier. 1978. On comprehending sentences: Syntactic parsing strategies. *Doctoral dissertation, University of Connecticut*.

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

Lyn Frazier. 1990. Parsing modifiers: Special purpose routines in the human sentence processing mechanism. *Comprehension processes in reading*, pages 303–330.

Minoru Fukuda. 2011. Locality in minimalist syntax by thomas s. stroik, linguistic inquiry monographs 51, mit press, cambridge, ma, 2009, x+ 149pp. *ENGLISH LINGUISTICS*, 28(2):321–333.

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. In *2000, Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, pages 95–126. MIT press.

Edward Gibson, Neal Pearlmutter, Enriqueta Canseco-Gonzalez, and Gregory Hickok. 1996. Recency preference in the human sentence processing mechanism. *Cognition*, 59(1):23–59.

Elizabeth Gilboy, Josep-MMaria Sopena, Charles Cliftrn Jr, and Lyn Frazier. 1995. Argument structure and association preferences in spanish and english complex nps. *Cognition*, 54(2):131–167.

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.

Nino Grillo and João Costa. 2014. A novel argument for the universality of parsing principles. *Cognition*, 133(1):156–187.

Nino Grillo, João Costa, Bruno Fernandes, and Andrea Santi. 2015. Highs and lows in english attachment. *Cognition*, 144:116–122.

E Gutierrez-Ziardegi, Manuel Carreiras, and Itziar Laka. 2004. Bilingual sentence processing: relative clause attachment in basque and spanish. In *ANNUAL CUNY CONFERENCE ON HUMAN SENTENCE PROCESSING*, volume 17.

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

Barbara Hemforth, Susana Fernandez, Charles Clifton Jr, Lyn Frazier, Lars Konieczny, and Michael Walter. 2015. Relative clause attachment in german, english, spanish and french: Effects of position and length. *Lingua*, 166:43–64.

Barbara Hemforth, Lars Konieczny, and Christoph Scheepers. 2000. Syntactic attachment and anaphor resolution: The two sides of relative clause attachment.

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

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

Gregory M. 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.

Jean-Pierre Koenig and Knud Lambrecht. 1999. French relative clauses as secondary predicates: A case study in construction theory. In *Unpublished ms. Communication préesentée au 2e Colloque de Syntaxe et Smantique de Paris*, volume 17.

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

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.

Maryellen C MacDonald, Neal J Pearlmutter, and Mark S Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological review*, 101(4):676.

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

George A. Miller and Noam Chomsky. 1963. Finitary models of language users. In R. Luce, R. Bush, and E. Galanter, editors, *Handbook of Mathematical Psychology*, volume 2. John Wiley, New York.

Alexandra Motut. 2010. Merge over move and the empirical force of economy in minimalism. *Toronto Working Papers in Linguistics*, 33.

Despina Papadopoulou and Harald Clahsen. 2003. Parsing strategies in l1 and l2 sentence processing: A study of relative clause attachment in greek. *Studies in Second Language Acquisition*, 25(4):501–528.

Céline Pozniak, Barbara Hemforth, Yair Haendler, Andrea Santi, and Nino Grillo. 2019. Seeing events vs. entities: The processing advantage of pseudo relatives over relative clauses. *Journal of Memory and Language*, 107:128–151.

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

Maryam Razaghi, Shahin Rahavard, and Firooz Sadighi. 2015. Economy, simplicity and uniformity in minimalist syntax. *International Journal on Studies in English Language and Literature*.

Xingjia Shen. 2006. Late assignment of syntax theory: evidence from chinese and english.

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.