

Evaluating a Minimalist grammar parser on Italian relative clause asymmetries

Aniello De Santo (Stony Brook University)

aniello.desanto@stonybrook.edu

Overview Recent studies have shown that a top-down parser for Minimalist grammars (MGs; [9]) 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 seemingly unrelated phenomena cross-linguistically [8, 5, 7]. At the core of this approach is a theory of grammatical structure driving 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, we test the parser performance on the processing asymmetries reported for Italian relative clauses (RCs), which have been object of extensive study in the psycholinguistic literature and, especially due to the *pro drop* nature of the language, make for an interesting testing ground for the MG parser.

Italian RCs Italian speakers conform to a reported cross-linguistic preference for subject (SRC, 1) over object (ORC, 2) relative clauses [12, a.o.]. Interestingly, Italian also allows for object relatives with the embedded subject expressed postverbally, and a null *pro* postulated in the embedded preverbal position (ORCp, 3). Although postverbal subject constructions are very common, studies have shown increased efforts with ORCp, leading to the following difficulty gradient: $SRC < ORC < ORCp$ [10]. The fact that SRCs are easier than ORCs is consistent with the predictions of a variety of theories [6, 2, 4]. However, the increased difficulty with ORCs including a postverbal subject has been subject of much interest in previous literature, and it is usually discussed in terms of costly local-ambiguity resolution [cf. 4, 12]. Thus, these asymmetries offer a fascinating challenge for the MG model since, with its *perfect oracle* strategy, the parser ignores ambiguity resolution in favor of comparing memory resources allocated to successful parses.

MG Parsing and Memory Metrics The MG parser systematically links syntactic structure to processing difficulty by connecting the stack states of a top-down parser for MGs to memory burden. In particular, the model refers to 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**). Based on these concepts, Graf et al. [7] define a set of off-line metrics of processing difficulty — measuring, for instance, **maximum**, **average** or total (**sum**) tenure over a whole derivation — sensitive to fine-grained structural changes.

Modeling Results and Discussion Because of the parser's sensitivity to grammatical structure, the choice of syntactic analysis is crucial. In line with most of the psycholinguistic literature on this topic, we adopt a promotion analysis of relative clauses and follow Belletti and Contemori [3] in our analysis of postverbal subjects. Then, to control for the top-down sensitivity to the depth of left- vs. right-embedding constructions, we test the parser performance on sentences of the form *I saw the horse* [_{RC} *that ...*] and *The horse* [_{RC} *that ...*] *jumped the fence* — the embedded relative clause either an SRC (1), an ORC (2), or an ORCp (3). Our simulations¹ show that the MG parser correctly predicts the gradient of difficulty observed for Italian RCs ($1 < 2 < 3$) across a variety of memory metrics. In particular, **max. tenure** and **sum size** play a crucial role in discriminating between alternatives, reflecting how the additional movement dependencies postulated for postverbal subject constructions drive the increased processing efforts for ORCps. These two metrics have already been noted as a good measure of processing difficulty in previous studies [7], which we take as additional support to the plausibility of the MG parser as a reliable model of sentence processing. These results also suggest how to reinterpret popular theories of Italian RC processing in a formal model of parsing. In particular, the MG parser offers a way to unify popular accounts of filler-gap dependencies and theories of intervention effects based on *featural Relativized Minimality* [1, 11], in a framework that takes both economy claims and syntactic assumptions seriously.

¹Code available at <https://github.com/CompLab-StonyBrook/mgproc>

• **Example sentences:**

- | | | |
|-----|--|-------------|
| (1) | Il cavallo che insegue i leoni
“The horse that chases the lions” | SRC |
| (2) | Il cavallo che i leoni inseguono
“The horse that the lions chase” | ORC |
| (3) | Il cavallo che <i>pro_i</i> inseguono i leoni;
The horse that <i>pro_i</i> chase the lions;
“The horse that the lions chase” | ORCp |

References

- [1] F. Arosio, F. Panzeri, B. Molteni, S. Magazù, and M. T. Guasti. The comprehension of Italian relative clauses in poor readers and in children with specific language impairment. *Glossa: a journal of general linguistics*, 2(1), 2017.
- [2] E. Bates and B. MacWhinney. Competition, variation, and language learning. *Mechanisms of Language Acquisition*, 20:157, 1987.
- [3] A. Belletti and C. Contemori. 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, 2009.
- [4] M. De Vincenzi. *Syntactic parsing strategies in Italian: The minimal chain principle*, volume 12. Springer Science & Business Media, 1991.
- [5] S. Gerth. *Memory Limitations in Sentence Comprehension: A Structural-based Complexity Metric of Processing Difficulty*, volume 6. Universitätsverlag Potsdam, 2015.
- [6] E. Gibson. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, Language, Brain*, pages 95–126, 2000.
- [7] T. Graf, J. Monette, and C. Zhang. Relative clauses as a benchmark for Minimalist parsing. *Journal of Language Modelling*, 5:57–106, 2017. doi: 10.15398/jlm.v5i1.157. URL <http://dx.doi.org/10.15398/jlm.v5i1.157>.
- [8] G. M. Kibele, S. Gerth, and J. Hale. Memory resource allocation in top-down minimalist parsing. In *Formal Grammar*, pages 32–51. Springer, 2013.
- [9] E. P. Stabler. Two models of minimalist, incremental syntactic analysis. *Topics in cognitive science*, 5(3):611–633, 2013.
- [10] I. Utzeri. The production and acquisition of subject and object relative clauses in Italian: a comparative experimental study. *Nanzan Linguistics*, 2, 2007.
- [11] S. Villata, L. Rizzi, and J. Franck. Intervention effects and relativized minimality: New experimental evidence from graded judgments. *Lingua*, 179:76–96, 2016.
- [12] F. Volpato and F. Adani. The subject/object relative clause asymmetry in Italian hearing-impaired children: evidence from a comprehension task. *Studies in Linguistics*, 3:269–281, 2009.