

What processing and computational modeling can tell us about syntax:

The case of Persian relative clauses

Nazila Shafiei & Aniello De Santo (Stony Brook University)

nazila.shafiei@stonybrook.edu

Overview In an ideal world, sentence processing effects would inform our syntactic choices and, for instance, help discriminate between competing analyses of complex phenomena. However, this is made especially complicated by the fact that such analyses often differ in subtle, fine-grained details. We suggest that this problem can be overcome by exploiting a computationally specified parsing model as a bridge between processing results and theoretical syntax [6]. We adopt a top-down parser for Minimalist grammars (MGs, [8]), combined with a set of metrics measuring memory usage — a model which has been shown to successfully explain processing difficulty across a variety of phenomena [3, a.o.]. As a case study, we look at attachment ambiguities in Persian relative clauses (RCs), comparing Kayne’s raising analysis of RCs [5] to Karimi’s base-generation approach [4].

Persian RCs and Processing Asymmetries While resembling the post-nominal structure of English RCs [1], Persian RCs show processing preferences closer to those of pre-nominal languages like Korean: in ambiguous cases, speakers show a marked preference for high attachment (HA) interpretations over low attachment (LA) ones [7]. Crucially, the internal structure of these clauses has been the topic of extensive debate. In particular, the position and scope properties of the Differential Object Marker *râ* have been used by Karimi [4] to argue in favor of a base-generation approach of the relative clause head, against Kayne’s better known raising analysis [5]. While these two analyses differ significantly from a theoretical perspective, it is unclear whether those subtle structural changes make different predictions in terms of processing. To probe this connection, we need a transparent linking theory, fine-grained enough to rigorously connect subtle structural details to processing performance.

Structure, Memory, and a Minimalist Grammar Parser Stabler’s [8] MG parser can be associated to three notions of memory: I) **tenure**: how long a node is kept in memory; II) **payload**: how many nodes must be kept in memory; or III) **size**: how many bits a node consumes in memory. These cognitive notions are then precisely quantified in a set of offline metrics — for instance, **maximum**, **average**, and total (**sum**) tenure — which crucially tie processing difficulty to differences in grammatical structure [3]. This is the model we exploit to compare the analyses for Persian RCs.

Computational Modeling Results, and Discussion We tested sentences like in (1), comparing the parser performance on trees corresponding to HA and LA interpretations, built according either to Kayne’s raising analysis [5], or Karimi’s base-generation proposal [4]. The parser correctly predicts a preference for HA over LA *only* when following a base-generation analysis, while producing the opposite prediction over trees built according to a raising approach. Two metrics in particular drive the parser performance towards the correct predictions — maximum tenure and sum size — consistent with previous MG parser results cross-linguistically [2, 3]. When adopting Karimi’s account, these metrics pick up on the effects of the length of the movement dependencies and overall memory commitment caused by the interaction of DP structure and base generation in LA cases. However, when adopting a Kaynian analysis, these subtle differences are lost due to an increase in memory usage common to both constructions, caused by the raising of the RC head. Overall then, these results are in support of Karimi’s argument for a base-generation approach to the structure of RCs in Persian, and highlight the plausibility of the MG parser as a transparent way to have sentence processing data inform syntactic theory. The next logical step will be to investigate different kinds of processing phenomena for RCs in Persian, such as the preferences for subject over object RCs. The success of this approach also opens the way to a careful comparison of attachment preferences across languages with pre- and post-nominal RCs, in order to better understand how fine differences in the syntax of these constructions drive processing cost.

- (1) Ali doxtar-e bâzigar-i ro ke ruye bâlkon bud did.
 Ali daughter-EZ actor-REL DOM that on balcony be.PST.3SG see.PST.3SG
- (i) 'Ali saw [[the daughter of the actor] that was on the balcony].' **HA**
- (ii) 'Ali saw [the daughter [of the actor that was on the balcony]].' **LA**

References

- [1] A. Arabmofrad and H. Marefat. Relative clause attachment ambiguity resolution in Persian. *Iranian Journal of Applied Linguistics*, 11(1):29–49, 2008.
- [2] S. Gerth. *Memory Limitations in Sentence Comprehension: A Structural-based Complexity Metric of Processing Difficulty*, volume 6. Universitätsverlag Potsdam, 2015.
- [3] 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>.
- [4] S. Karimi. Persian complex dps: How mysterious are they? *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 46(1-2):63–96, 2001.
- [5] R. S. Kayne. *The antisymmetry of syntax*. Number 25. MIT Press, 1994.
- [6] G. M. Kobele, S. Gerth, and J. Hale. Memory resource allocation in top-down minimalist parsing. In *Formal Grammar*, pages 32–51. Springer, 2013.
- [7] H. Marefat and M. Meraji. Parsing preferences in structurally ambiguous relative clauses: L1 vs. L2. *The International Journal of Humanities*, 12(1):111–127, 2005.
- [8] E. P. Stabler. Two models of minimalist, incremental syntactic analysis. *Topics in cognitive science*, 5(3):611–633, 2013.