Processing French RCs with Postverbal Subjects in a Minimalist Parser

Daniel Del Valle & Aniello De Santo
Dept. of Linguistics, University of Utah
{name.surname}@utah.edu

Overview   Computational models with explicit assumptions about the connection between syntactic representations and processing difficulty can help strengthen bridges between theoretical linguistics and psycholinguistics. In this sense, a model based on Stabler's parser for Minimalist grammars (MGs; Stabler 2013) has been shown to predict a variety of off-line processing preferences, by exploiting complexity metrics tracking how syntactic structure affects memory load (Graf et al. 2017:a.o.). This model provides an interpretable linking theory between fine-grained syntactic structure in the generative tradition and precise sentence processing results, through transparently specified notions of cognitive cost. Here, we build on recent work on the processing of non-canonical word order constructions in Italian (De Santo 2021), and explore the model's performance on off-line results for cases of *subject inversion* within restrictive relative clauses (RCs) in French. This provides an opportunity to probe the sensitivity of the approach to small but critical syntactic differences across similar constructions cross-linguistically.

MG Parser   MGs (Stabler 2013) are a formalization of the Minimalist program, the syntactic framework adopted in this investigation. We follow previous literature in employing a top-down incremental parser for this formalism, evaluating the precise sequence of merge and move operations needed to derive a sentence via *derivational* representations (Figure 1). Then, complexity metrics are used to link structure building to aspects of memory usage, and provide a way to estimate and contrast the cognitive load of building one structure over anothe (e.g. subject vs. object relatives Kobele et al. 2013; Graf et al. 2017). Consider, for instance, Figure 1. Each node in the tree is annotated with a superscript (index), indicative of when the node enters the parser's memory stack, and a subscript (outdex), indicative of the point where the node is confirmed and thus is taken out of the memory stack. Then, TENURE, a measure of how long a node is kept in memory, is estimated by subtracting the index from the outdex of each node. Because we are interested in modeling off-line processing preferences, we then adopt MAXT (MAXIMUM TENURE), a metric which identifies the maximum tenure over a full syntactic derivation. For instance, in Figure 1 MAXT is 9 $(14-5)$, measured on *q*ui. Interestingly, previous work has identified MAXT as an especially helpful metric in evaluating the effectiveness of the model across a variety of processing phenomena (Graf et al. 2017; De Santo 2021).

Figure 1: Annotated MG derivation tree for the SRC in (1), following a Wh-movement analysis. Movement landing sites are marked by unary branches. Boxed nodes are those for which tenure is greater than 2.

French Postverbal Subjects   French is a SVO language with postnominal relative clauses (RCs), and conforms to a general preference for subject RCs (SRC, 1) over object RCs (ORC, 2) (Pozniak et al. 2016). While French declarative clauses are canonically thought as SVO, subject inversion — i.e. with the subject linearized postverbally — is possible under certain contexts. While in some other Romance languages (e.g. Italian or Spanish) subject inversion is allowed in matrix clauses with little restriction, French is more strict in the thematic contexts which license it (Kayne and Pollock 1978; Lahousse 2006). RCs are one environment which allow subject inversion in all contexts (see 1,2,3), with reported additional processing difficulty of ORCp compared to preverbal RCs such that: SRC < ORC < ORCp (Pozniak and Hemforth 2015).
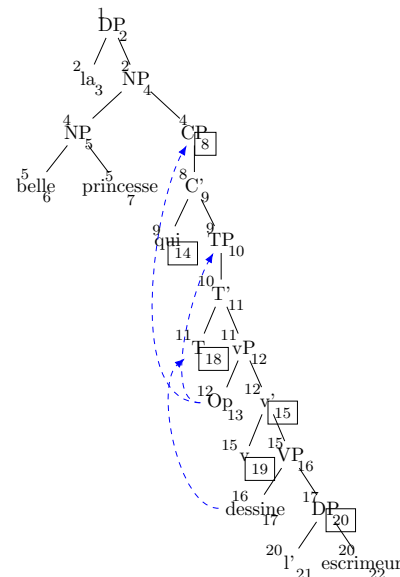
(1) la belle   princesse qui dessine l'escrimeur **SRC**
    the beautiful princess$_{sbj}$ that draws   the fencer$_{obj}$
    'the beautiful princess that draws the fencer'

(2) la belle   princesse que l'escrimeur dessine **ORC**
    the beautiful princess$_{obj}$ that the fencer$_{sbj}$ draws
    'the beautiful princess that the fencer draws'

(3) la belle   princesse que dessine l'escrimeur **ORCp**
    the beautiful princess$_{obj}$ that draws   the fencer$_{sbj}$
    'the beautiful princess that the fencer draws'

Increased difficulty of ORCp contructions is not trivial to derive for some approaches to processing difficulty (Arosio et al. 2017), but De Santo (2021) shows how it follows for Italian from the structural predictions made by the MG Model. Because of the particular environments that differentiate French subject inversion from the Italian cases, the French contrast poses a valuable case study for a parsing model sensitive to fine-grained structural differences.

Results and Discussion   Beyond exploring whether the parser is able to reproduce the processing gradient reported above, we examine the effects that alternative syntactic choices have on processing. Following De Santo (2021), we explore two common relative clause theories: the promotion analysis (Kayne 1994) and the wh-movement analysis (Chomsky 1977), and their interactions with a French postverbal subject analysis (Drijkoningen and Kampers-Manhe 2008). Simulations revealed the MG-Parser was indeed capable of reflecting the expected processing gradient, SRC < ORC < ORCp, for both relative clause theories (see Table 1). Such results mirror what De Santo (2021) reported for Italian, despite the significant differences in structural analysis.

A more in-depth qualitative inspection of the results allows us to begin exploring differences in exactly what drives memory load in each analysis (cf. Table 1), capitalizing on the model's sensitivity to subtle structural details. Notably, RC preferences cross-linguistically can be modulated by a variety of lexical, semantic, and pragmatic factors, which also seem to be relevant for the relative acceptability of subject inversion (Pozniak et al. 2021). Crucially, this model is not a commitment

	Prom		Wh	
	MaxT	**Node**	**MaxT**	**Node**
SRC	11	T	7	T
ORC	14	T	10	T
ORCp	19	Foc	15	Foc

Table 1: Performance of MaxT for each of the RC derivations, according to a promotion (Prom) or WH-movement (Wh) analysis.

to a reductionist, syntax-only approach to the RC asymmetries, and these factors suggest ways in which the MG Model itself could be extended. Nonetheless, the ability of the parser to capture these processing contrasts in controlled syntactic environments helps probing how the effects of particular structural assumptions interact with commitments to notions of cognitive cost.

Arosio, F. et al. (2017). The comprehension of Italian relative clauses in poor readers and in children with specific language impairment. Glossa. • **Chomsky, N.** (1977). On wh-movement. • **De Santo, A.** (2021). Italian postverbal subjects from a minimalist parsing perspective. Lingue e linguaggio. • **Drijkoningen, F. & Kampers-Manhe, B.** (2008). On inversions and the interpretation of subjects in French. • **Graf, T., et al.** (2017). Relative clauses as a benchmark for Minimalist parsing. JoLM. **Kayne, R. and Pollock, J.-Y.** (1978). Stylistic inversion, successive cyclicity, and move NP in French. LI. • **Kayne, R.** (1994). The antisymmetry of syntax. MIT press. • **Kobele, G. M., et al.** (2013). Memory resource allocation in top-down minimalist parsing. In FG. • **Lahousse, K.** (2006). NP subject inversion in french: two types, two configurations. Lingua. • **Pozniak, C., et al.** (2021). Subject inversion in French object relatives: What's your preference?. • **Pozniak, C. and Hemforth, B.** (2015). Processing subject and object relative clauses in French and Mandarin Chinese. Poster at CUNY 28th. • **Pozniak, C., et al.** (2016). French object relatives: evidence against DLT but not entirely explained by frequency. • **Stabler, E. P.** (2013). Two models of minimalist, incremental syntactic analysis. T. in CogSci.