Relative clause processing in a flexible word order language: Evidence from Hungarian

Eszter Ronai & Ming Xiang (The University of Chicago)
ronai@uchicago.edu

Hungarian subject-extracted (SRC) and object-extracted (ORC) relative clauses both have flexible word order. In a self-paced reading experiment, we find an advantage for the more frequent SRC and ORC types, despite there being a longer distance between the filler and gap. Our findings support an expectation-based, rather than a memory-based account of relative clause processing.

Background. An important case study in the processing of syntactic complexity is the asymmetry we observe between the English SRC (1a) and ORC (1b): ORC is harder to process than SRC.

(1) a. The engineer [$_{RC}$ who _ annoyed the analyst] wrote a report about the project. (SRC)
b. The engineer [$_{RC}$ who the analyst annoyed _] wrote a report about the project. (ORC)

There are two competing classes of accounts. Memory-based accounts generally predict a locality preference – shorter filler-gap dependencies are preferred (Gibson, 1998; Lewis & Vasishth, 2005). Expectation-based accounts attribute greater processing cost to less expected structures (e.g. surprisal, Hale, 2001; Levy, 2008). In English, the predictions of both accounts converge, since English SRCs instantiate a shorter filler-gap dependency than ORCs, and they are also more frequent.

Experiment and Predictions. Hungarian is particularly suitable for teasing apart the predictions of these two accounts, because extraction site and word order (i.e. the local proximity of the relative clause verb to the relative pronoun) can be varied independently. SRCs (2) can occur either in a VO (local) or OV (non-local) configuration, while ORCs (3) can occur in VS (local) or SV (non-local).

(2) A mérnök, aki {idegesítette az elemzőt} / {az elemzőt idegesítette},...
the engineer.NOM who.NOM {annoyed the analyst.ACC }/ {the analyst.ACC annoyed}...
Both: 'The engineer who annoyed the analyst... (wrote a report about the project).'

(3) A mérnök, akit {idegesített az elemző}/ {az elemző idegesített},...
the engineer.NOM who.ACC {annoyed the analyst.NOM}/ {the analyst.NOM annoyed}...
Both: 'The engineer who the analyst annoyed... (wrote a report about the project).'

Under the Memory account, word orders supporting local/shorter filler-verb dependencies would be less costly (VO, VS) than non-local ones (OV, SV), irrespective of SRC vs. ORC status. To check the predictions of the Expectation account, we carried out corpus searches in the Hungarian National Corpus (Oravecz et al., 2014). For both SRC and ORC, non-local structures are more common than local ones (see Table 1). Thus Expectation makes the opposite predictions to Memory: non-local structures should be easier to process. Expectation accounts also predict a general advantage for SRCs, since they are overall more frequent than ORCs. A self-paced reading experiment was conducted with 81 native speakers of Hungarian (item N=32). A comprehension question followed each sentence. In order to avoid a confound between event plausibility and extraction type, nouns (e.g. *engineer*, *analyst*) were counterbalanced to occur both in head NP and RC NP positions.

Results. Analyses on log-transformed reading times (RTs) (Figure 1) revealed that for both SRC and ORC, the RC Verb (*annoyed* in (2, 3)) had significantly shorter RTs in the non-local than the local configurations ($p<.001$). This supports the predictions of Expectation, but not Memory-based accounts. Surprisingly, there was no effect on the rel. pronoun, where case marking disambiguates SRC vs. ORC – even though SRCs are overall more frequent. However, the predicted SRC advantage did manifest in comprehension question accuracy (Figure 2): questions following SRCs (especially local) were easiest to answer (RC type ($p<.001$); interaction of locality-RC type ($p<.05$)).

Conclusion. Levy et al. (2013), using a similar manipulation in Russian, found a clear reading time advantage for local sentences at the RC Verb. The Hungarian results presented here are the opposite: non-local structures are easier. Crucially, in Russian, local structures are also the more frequent ones, whereas in Hungarian they are not. Thus Hungarian better teases apart the predictions of the two competing accounts, and provides support for the Expectation-based account.

Structure	Count	Searches
SRC, local	44	(Det) N.nom (,) Rel.Pronoun.nom V.3sg (Det) N.acc
SRC, non-local	466	(Det) N.nom (,) Rel.Pronoun.nom (Det) N.acc V.3sg
ORC, local	26	(Det) N.nom (,) Rel.Pronoun.acc V.3sg (Det) N.nom
ORC, non-local	50	(Det) N.nom (,) Rel.Pronoun.acc (Det) N.nom V.3sg

Table 1: Counts from the part-of-speech-tagged Hungarian National Corpus.

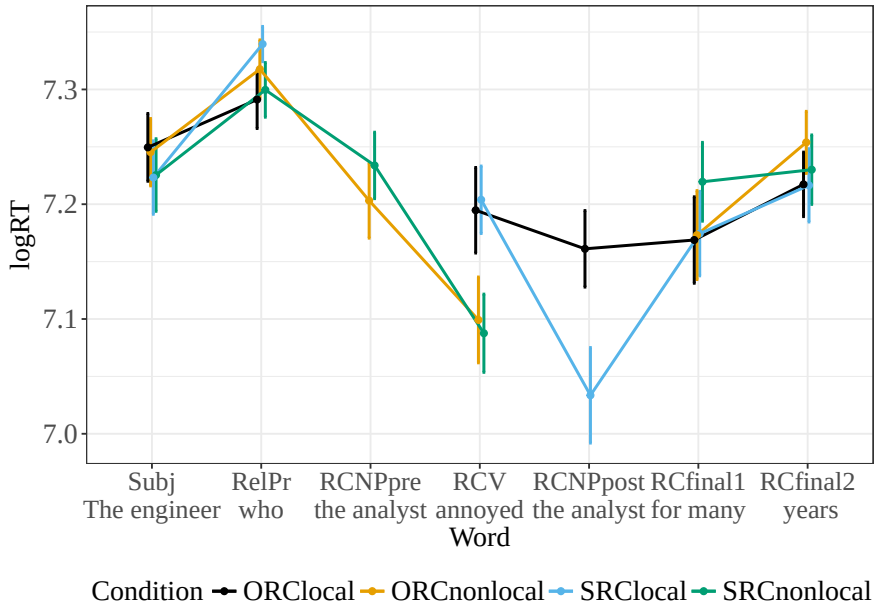

Figure 1: Log-transformed reading times and standard errors for the first six word. The position of the RC noun (RCNP) with respect to the RC verb (RCV) depends on locality. The RCfinal regions represent a prepositional phrase. Trials with an incorrectly answered question were excluded.

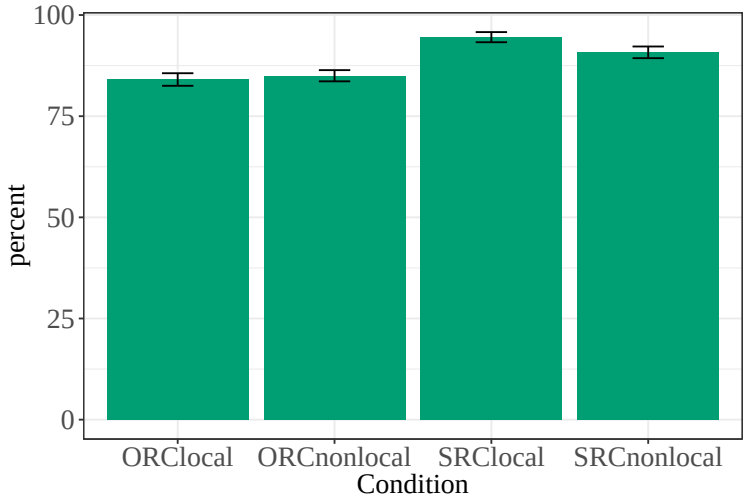

Figure 2: Percentage of correctly answered comprehension questions, following self-paced reading sentences in four conditions. Error bars indicate standard error.

References. Gibson. 1998. Linguistic complexity: locality of syntactic dependencies. Cognition. // Hale. 2001. A probabilistic earley parser as a psycholinguistic model. // Levy. 2008. Expectation-based syntactic comprehension. Cognition. // Levy, et al. 2013. The syntactic complexity of Russian relative clauses. JML. // Lewis & Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. Cognitive Science. // Oravecz, et al. 2014. The Hungarian Gigaword Corpus.