Eye-tracking Evidence for Frequency and Integration Cost Effects in Corpus Data

Vera Demberg (University of Edinburgh) & Frank Keller (University of Edinburgh) & Roger Levy (University of California, San Diego)
v.demberg@sms.ed.ac.uk

Relative clauses have been investigated extensively in the literature on syntactic processing difficulty. Experimental results (Gibson, 1998; King & Just, 1991) show that English subject relative clauses (SRCs) as in (1a) are more difficult to process than object relative clauses (ORCs) as in (1b).

(1) a. The reporter who [R1 attacked] the senator admitted the error.

b. The reporter who [R2 the] senator [R3 attacked] admitted the error.

These result were obtained for isolated sentences in a fully controlled experimental setting. This raises the question whether the ORC/SRC asymmetry generalizes to naturally occurring, contextualized text. We tested this hypothesis on the Dundee Corpus (Kennedy, 2003), which contains the eyetracking record of 10 subjects reading 50,000 words of newspaper text.

Syntactic Prediction Locality Theory (SPLT; Gibson, 1998) explains the ORC/SRC asymmetry based on the cost of integrating the verb's arguments when the verb is reached. ORCs incur a higher integration cost, as more discourse referents intervene between the relative pronoun and the verb than in SRCs. This predicts that the reading time for R3 in (1b) should be higher than that for R1 in (1a). Furthermore, SPLT predicts a higher integration cost on R1 than on R2 because the relative pronoun has to be integrated with the verb in R1 but not with the NP in R2.

An alternative explanation is offered by probabilistic accounts that predict shorter reading times for more probable linguistic sequences. For example, McDonald and Shillcock (2003) show that P(w_n|w_n-1), the transitional probability of word w_n given the previous word w_n-1, is predictive of reading times. Thus, reading time differences between R1 and R2 and between R1 and R3 maybe the explainable in terms of transitional probability.

To test these predictions, we extracted all subject and object relative clauses headed by "who", "which", and "that" from the Dundee Corpus, and computed the reading times for regions R1 and R3 for each item and each subject (approx. 3000 data points). We fitted a generalized linear model with reading time as the dependent variable and the predictors subject, item, word frequency, word length, RC type and transitional probability. The results show a significant main effect of RC type ($p < 0.001$) for R1 and R3: SRC verbs were read more quickly than ORC verbs, as predicted by SPLT. We fitted a second generalized linear model for the reading times of R1 (SRC verb) and R2 (ORC determiner, noun, or adjective). Again, we found a significant main effect of RC type ($p < 0.001$). Region R1 was read more slowly than region R2, which is again in line with SPLT predictions.

In both of these regions, transitional probability emerged as a significant factor in the regressions ($p < 0.001$) for total reading times. This effect was additional to the RC type effect. On first fixation times, we only found an RC type effect but no transitional probability effect for R1 vs. R2.

Our findings provide a new type of evidence in favor of locality-based theories of processing complexity such as Gibson's SPLT. To our knowledge, this is the first time that SPLT effects have been found for reading time data extracted from naturally occurring, contextualized texts. Furthermore, we confirmed that transitional probability is a predictor of reading time, but that SPLT effects occur in addition to effects of transitional probability.

References

Gibson, E. 1998. Linguistic complexity: locality of syntactic dependencies, Cognition 68, 1-76.

Kennedy, A. 2003. The Dundee Corpus. Poster at the 12th European Conference on Eye Movements, Dundee.

King, J. and Just, M. A. 1991. Individual differences in syntactic processing: the role of working memory. Journal of Memory and Language 30, 580-602.

McDonald, S.A., and Shillcock, R.C. (2003). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. Vision Research, 43, 1735-1751