Surprising parser actions and reading difficulty

Marisa Ferrara Boston[1], John T. Hale[1], Reinhold Kliegl[2], and Shravan Vasishth[2]

[1]Michigan State University [2]Universität Potsdam, Germany

Introduction

This study examines whether surprisal (Hale 2001), a parser-based complexity measure, can predict German readers' eye-fixation durations, an empirical measure of sentence processing difficulty.

We use the Potsdam Sentence Corpus (PSC), a 144 sentence eye movement corpus read by 272 native German speakers.

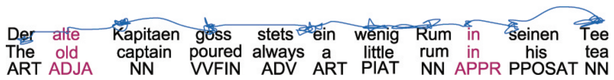

Figure 1: Fixations for a PSC sentence.

Baseline predictors like log frequency (lf), log bigram frequency (bi), word length (len), and human predictability given the left context (pr) model the total reading time (TRT) of eye-fixations per word (Ehrlich & Rayner 1981). We hypothesize that the addition of syntactic effects, as measured by surprisal, will better model this data than these word-level factors alone. We also test the role of the beam-size k in surprisal calculation.

$$\log(TRT) = 5.4 - 0.02lf - 0.01bi - 0.59len^{-1} - 0.02pr$$

A dependency parser calculates surprisal...

Surprisal is the logarithm of the prefix-probability α eliminated in the transition from one word to the next.

$$\text{surprisal}(n) = \log_2\left(\frac{\alpha_{n-1}}{\alpha_n}\right)$$

$$\alpha_n = \sum_{d \in \mathcal{D}(G,wv)} \text{Prob}(d)$$

When the transition from previous word to current word is low probability (as at *in*, compared with *alte* in 1), the surprisal is high. The psycholinguistic claim is that behavioral measures should register increased cognitive difficulty: rare parser actions are cognitively costly.

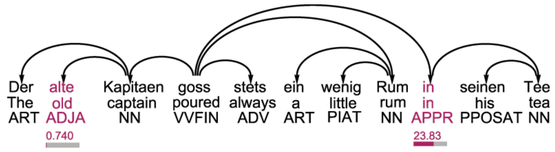

Figure 2: Surprisal for a PSC sentence.

We calculate surprisal using an incremental dependency parser built to the specifications of Nivre 2004 with an added k-best search, as in (3) and (4). We calculated surprisal for 10 total models, k=1...9,100.

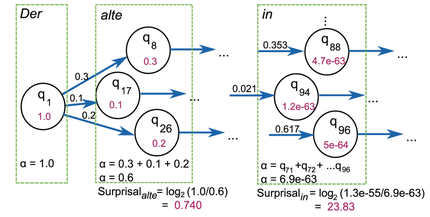

Figure 3: State-based surprisal calculation.

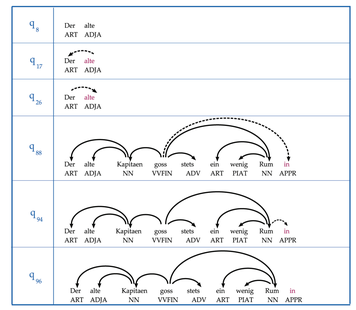

Figure 4: Dependency claims in parser states q.

...which predicts fixation durations independent of baseline predictors

Using TRT as a dependent measure, we compared the quality of fit of the baseline model above to the 10 models which had all the predictors of the baseline plus one of the 10 surprisals. We evaluated the change in relative quality of fit due to surprisal with the Deviance Information Criterion (DIC) (Spiegelhalter et al. 2002). Low DIC = better fit.

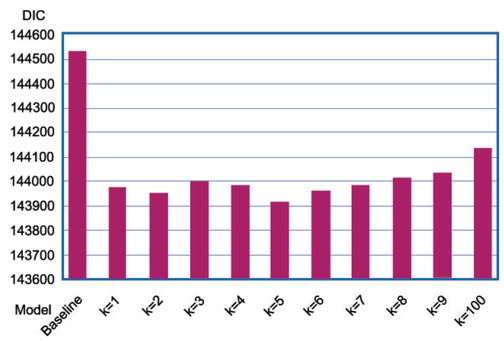

Figure 5: DICs from the multiple regressions using different versions of surprisal.

Surprisal predicts fixation durations and improves model fit for TRT measures when used in addition to the baseline measures. Small k (1,2,3) is more valuable than high k (100), challenging the assumption that cognitive functions are global optima.

Conclusion

This study demonstrates that surprisal calculated with a dependency parser is a significant predictor of reading times, an empirical measure of cognitive difficulty. Predictions derived even from very narrow-beamed parsers improve a baseline eye-fixation duration model, and support a boundedly rational view of the human parser.

S. F. Ehrlich and K. Rayner. 1981. Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20:641655.

J. Hale. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of 2nd NAACL*, 18.

J. Nivre. 2004. Incrementality in deterministic dependency parsing. In *Incremental Parsing: Bringing Engineering and Cognition Together, Barcelona, Spain.* Association for Computational Linguistics.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society*, 64(B):583639.

{mferrara, jthale}@msu.edu, kliegl@uni-potsdam.de, vasishth@acm.org

http://www.msu.edu/~mferrara/