

Coreference information guides human expectations during natural reading

Evan Jaffe, Cory Shain, and William Schuler (The Ohio State University)
jaffe.59@osu.edu

Models of human sentence processing effort tend to focus on costs associated with retrieving structures and discourse referents from memory (memory-based) and/or on costs associated with anticipating upcoming words and structures based on contextual cues (expectation-based) [4]. Although expectation and memory may play separable roles in language comprehension [5], theories of coreference processing have largely focused on memory: how comprehenders identify likely referents of linguistic expressions. In this study, we hypothesize that coreference tracking also informs human expectations about upcoming words, and test by evaluating the degree to which incremental surprisal measures generated by a novel coreference-aware semantic parser explain human response times in a naturalistic self-paced reading experiment. Results indicate (1) that coreference information indeed guides human expectations and (2) that coreference effects on memory retrieval exist independently of coreference effects on expectations. Together, these findings suggest that the language processing system exploits coreference information both to retrieve referents from memory and to anticipate upcoming material.

To generate incremental coreference-aware surprisal, we augment a left-corner parser [3, 9] with referential contexts indicating predicate and argument information provided by a categorial grammar reannotation [7]. The parser is currently implemented as a series of maximum entropy submodels that make parsing subdecisions using discrete, symbolic features. An example referential context is `CLOUD_1`, indicating the first argument of a ‘being a cloud’ predicate. At each word, a coreference submodel chooses an earlier antecedent index and inherits its referential contexts (or else the submodel chooses a special null antecedent which does not provide additional contexts). These contexts are additional features that the other parsing submodels condition on, whose joint decisions form a parse hypothesis.

(1) Aqua_i thought he_i was the luckiest guy to see a cloud_j pouring its_j rain onto him_i .

For example, at the word *its* in Ex. (1), the system would consider coreference antecedent offsets including those corresponding to *he* and *cloud*. With the correct antecedent, the referential context of *its* would inherit the `CLOUD_1` context, making the following word *rain* more likely than if it had inherited the incorrect antecedent context `HE_1`. Left-corner parsing decisions and word emissions are conditioned on the coreference decision, resulting in an incremental generative semantic parsing model that is able to generate word-by-word surprisal estimates while explicitly tracking coreference, which is a novel contribution, to the authors’ knowledge.

The parser is trained on the coreference-annotated subset of OntoNotes [10] and used to generate surprisal estimates for the Natural Stories corpus [1], which includes 768,584 self-paced reading time observations for 181 subjects, divided into two equally sized partitions for exploratory analysis and hypothesis testing. We use ablative likelihood ratio testing of linear mixed effects models to test whether incremental surprisal estimates from the coreference-aware parser improve over those of an otherwise identical semantic parser that lacks coreference information. Models also include control variables for word length, percent narrative completed, and 5-gram surprisal. All predictors are z-scored prior to fitting and single-best spillover [8] position is optimized on held-out data, resulting in the surprisal estimates and MentionCount being spilled over by one position. MentionCount [2] is an index of topicality as measured by entity repetition, where repetition is understood to correlate with activation strength in memory. All models have full by-subject random effects for all predictors.

Results show a significantly improved fit ($p=5.6e-5$) to reading times when adding a fixed effect for coreference-aware surprisal, supporting the claim that humans use coreference information to guide word expectation (Tbl. 1). We further show that a previously proposed estimator of coreference resolution difficulty [2] explains additional variance over that explained by the coreference-aware parser, suggesting that the memory phenomenon is at least partially independent of expectation (Tbl. 2). Lastly, a comparison of by-word mean residuals show that the coreference-aware surprisal model has significantly improved reading time prediction for words immediately following pronouns, which we interpret as evidence that specifically coreference information from pronoun antecedent resolution improves overall model fit (Tbl. 3).

Together, these results indicate that comprehenders incrementally exploit coreference information both to identify referents of linguistic expressions (a memory effect) and to inform predictions about upcoming words (an expectation effect). Our study thus joins previously reported evidence of the need for both memory-related and expectation-related mechanisms in accounting for the full range of human sentence processing phenomena [6].

Effect	$\beta(z)$	$\beta(ms)$
Word Length	7.487	3.22
Story Position	-15.47	-41.26
NgramSurp	9.958	5.40
NoCorefSurpS1	0.676	0.118
CorefSurpS1	3.198	0.554

Table 1: Experiment 1 fixed effect estimates for full model on held-out data partition. Positive effects correspond to increases in reading time duration while negative effects correspond to a decrease in reading time duration.

Effect	$\beta(z)$	$\beta(ms)$
Story Position	-11.26	-30.03
NgramSurp	13.11	7.11
CorefSurpS1	3.32	0.58
MentionCountS1	-3.92	-0.20

Table 2: Fixed effects from Experiment 2 full model. MentionCount is significant when added to a strong baseline including lexical, syntactic **and** coreference controls. Word Length is omitted for convergence reasons as the weakest predictor. There are 54,026 observations from 180 subjects. With a range of 1 to 90 in this corpus, MentionCountS1 accounts for approximately 20ms of reading time facilitation at its max value.

Model	Mean Residual (ms)	
	All Words	Post Pronoun
Full	0.003	-7.67
Ablated	0.004	-9.13

Table 3: CorefSurpS1 full model has significantly lower mean residual magnitudes for words immediately after pronouns on dev data. There is no significant difference between residuals across models when comparing all words (9287), but there is when looking at words immediately after a pronoun (444), with $p = 3.79e-48$.

References

- [1] Futrell, R., Gibson, E., Tily, H. J., Blank, I., Vishnevetsky, A., Piantadosi, S. T., and Fedorenko, E. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, 2018.
- [2] Jaffe, E., Shain, C., and Schuler, W. In *Cognitive Modeling and Computational Linguistics (CMCL)*, 2018.
- [3] Johnson-Laird, P. N. *Mental models: towards a cognitive science of language, inference, and consciousness*, 1983.
- [4] Levy, R. *Cognition*, 2008.
- [5] Levy, R., Fedorenko, E., and Gibson, E. *Journal of Memory and Language*, 2013.
- [6] Levy, R. and Gibson, E. *Frontiers in Psychology*, 2013.
- [7] Nguyen, L., van Schijndel, M., and Schuler, W. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING '12)*, 2012.
- [8] Rayner, K. *Psychological Bulletin*, 1998.
- [9] van Schijndel, M., Exley, A., and Schuler, W. *Topics in Cognitive Science*, 2013.
- [10] Weischedel, R., Pradhan, S. S., Ramshaw, L., Kaufman, J., Franchini, M., El-Bachouti, M., Xue, N., Palmer, M., Hwang, J., Bonial, C., Choi, J., Mansouri, A., Foster, M., Hawwary, A.-a., Marcus, M., Taylor, A., Greeberg, C., Hovy, E., Blevin, R., and Houston, A. *OntoNotes*. Technical report, 2012.