Evidence of semantic processing difficulty in naturalistic reading

Cory Shain[1], Richard Futrell[2], Marten van Schijndel[3], Edward Gibson[2], William Schuler[1], and Evelina Fedorenko[2]; [1]Ohio State, [2]MIT, [3]Johns Hopkins

Language is a powerful vehicle for conveying our thoughts to others and inferring thoughts from their utterances. Much research in sentence processing has investigated factors that affect the relative difficulty of processing each incoming word during language comprehension, including in rich naturalistic materials [4, 8, 6, 24, 27, 23]. However, in spite of the fact that language is used to convey and infer meanings, prior research has tended to focus on lexical and/or structural determinants of comprehension difficulty. This focus has plausibly been due to the fact that lexical and syntactic properties can be accurately estimated in an automatic fashion from corpora [11] or using high-accuracy automatic incremental parsers [22, 26]. Comparable incremental semantic parsers are currently lacking. However, recent work in machine learning has found that distributed representations of word meanings — based on patterns of lexical co-occurrence — contain a substantial amount of semantic information [16, 14], and predict human responses in a range of psycholinguistic tasks [19, 2, 21, 5]. To examine the effects of semantic relationships among words on comprehension difficulty, we estimated a novel measure — incremental semantic relatedness — for three naturalistic reading time corpora: Dundee [12], UCL [7], and Natural Stories [9]. In particular, we embedded all three corpora using GloVe vectors [20] pretrained on the 840B word Common Crawl dataset, then computed the mean vector distance between the current word and all content words preceding it in the sentence. This provides a measure of a word's semantic relatedness to the words that precede it without requiring the construction of carefully normed stimuli, permitting us to evaluate semantic relatedness as a predictor of comprehension difficulty in a broad-coverage setting. Hypothesis testing was done with ablative likelihood ratio testing of linear mixed effects models, controlling for word length in characters, position in the sentence, 5-gram surprisal as computed by KenLM [11] trained on Gigaword 3 [10], and PCFG surprisal as computed by the [26] parser trained on the WSJ corpus [15] re-annotated into Generalized Categorial Grammar [18].[1] We found a significant positive effect of mean cosine distance on reading time duration in each corpus.

In summary, in line with previous work that has shown that the semantic relationship between a target word and its context affects comprehension in constructed stimuli presented in isolation [13, 17], we provide strong broad-coverage evidence of this factor, over and above linear (5-gram) and syntactic (PCFG) models of linguistic expectation. Our results are consistent with at least two (perhaps complementary) interpretations. Semantically related context might facilitate processing of the target word through spreading activation [1]. Or vector distances might approximate the surprisal values of a semantic component of the human language model, thus yielding a rough estimate of semantic surprisal. Future advances in incremental semantic parsing may permit more precise exploration of these possibilities.

[1] Random slopes for each of these by subject along with by-subject and by-word random intercepts were also included. The eye-tracking baselines (Dundee and UCL) also included saccade length and variants of the surprisal predictors accumulated over saccade regions [25]. Using 1/3 of each corpus reserved for exploratory model selection, spillover position of the baseline predictors was optimized using ordinary least squares regression. All predictors remained *in situ* except: Dundee (5-gram surprisal spillover-1), UCL (saccade length spillover-1), and Natural Stories (PCFG surprisal spillover-1). Using the exploratory set, we found the strongest main effects in spillover-1 position. We also found that accumulating semantic distance over saccade regions improves fit on the eye-tracking data. These settings were therefore used in our final evaluation.

Corpus	p	t	β-*ms*
Natural Stories	0.006	2.766	1.25
Dundee	5.59e-4	4.759	5.73
UCL	2.76e-10	7.853	16.36

Table 1: Likelihood ratio testing results for mean semantic cosine distance on Natural Stories, Dundee, and UCL. Reading times were transformed using [3] and β-ms was computed by backtransformation, and is therefore only valid at the backtransformed mean, holding all other effects at their means.

References

[1] J. R. Anderson et al. “An integrated theory of the mind”. In: *Psychological Review* 111.4 (2004), pp. 1036–1060.

[2] M. Baroni, G. Dinu, and G. Kruszewski. “Don’t count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors”. In: *Proceedings of ACL 2014*. Baltimore, Maryland, 2014, pp. 238–247. URL: `http://aclweb.org/anthology/P14-1023`.

[3] G. E. P. Box and D. R. Cox. “An Analysis of Transformations”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 26.2 (1964), pp. 211–252.

[4] V. Demberg and F. Keller. “Data from eye-tracking corpora as evidence for theories of syntactic processing complexity”. In: *Cognition* 109.2 (2008), pp. 193–210.

[5] A. Ettinger et al. “Modeling N400 amplitude using vector space models of word representation”. In: *Proceedings of the 38th annual conference of the Cognitive Science Society*. 2016, pp. 1445–1450.

[6] V. Fossum and R. Levy. “Sequential vs. Hierarchical Syntactic Models of Human Incremental Sentence Processing”. In: *Proceedings of CMCL 2012*. Association for Computational Linguistics, 2012.

[7] S. L. Frank et al. “Reading time data for evaluating broad-coverage models of English sentence processing”. In: *Behavior Research Methods* 45 (4 2013), pp. 1182–1190.

[8] S. Frank and R. Bod. “Insensitivity of the human sentence-processing system to hierarchical structure”. In: *Psychological Science* (2011).

[9] R. Futrell et al. “The Natural Stories Corpus”. In: *arXiv* 1708.05763 (2017).

[10] D. Graff and C. Cieri. *English Gigaword LDC2003T05*. Philadelphia: Linguistic Data Consortium, 2003.

[11] K. Heafield et al. “Scalable Modified Kneser-Ney Language Model Estimation”. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. 2013, pp. 690–696.

[12] A. Kennedy, J. Pynte, and R. Hill. “The Dundee Corpus”. In: *Proceedings of the 12th European conference on eye movement*. 2003.

[13] M. Kutas and S. A. Hillyard. “Reading senseless sentences: brain potentials reflect semantic incongruity”. In: *Science* 207.4427 (Jan. 1980), pp. 203–205. DOI: `10.1126/science.7350657`.

[14] O. Levy and Y. Goldberg. “Dependency-Based Word Embeddings”. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. 2014, pp. 302–308. URL: `http://aclweb.org/anthology/P/P14/P14-2050.pdf`.

[15] M. Marcus et al. “The Penn TreeBank: Annotating predicate argument structure”. In: *Proceedings of the ARPA Human Language Technology Workshop*. 1994.

[16] T. Mikolov, W.-t. Yih, and G. Zweig. “Linguistic Regularities in Continuous Space Word Representations.” In: *In Proceedings of NAACL 2013*. 2013.

[17] R. K. Morris. “Lexical and message-level sentence context effects on fixation times in reading.” In: *Journal of experimental psychology. Learning, memory, and cognition* 20 1 (1994), pp. 92–103.

[18] L. Nguyen, M. van Schijndel, and W. Schuler. “Accurate Unbounded Dependency Recovery using Generalized Categorial Grammars”. In: *Proceedings of COLING 2012*. Mumbai, India, 2012, pp. 2125–2140.

[19] M. Parviz et al. “Using language models and Latent Semantic Analysis to characterise the N400m neural response”. In: *Proceedings of the Australasian Language Technology Association Workshop*. 2011, pp. 38–46.

[20] J. Pennington, R. Socher, and C. D. Manning. “GloVe: Global Vectors for Word Representation”. In: *Proceedings of EMNLP*. 2014.

[21] F. Pereira et al. “A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data”. In: *Cognitive Neuropsychology* 33 (2016), pp. 175–190.

[22] B. Roark et al. “Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing”. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Langauge Processing* (2009), pp. 324–333.

[23] C. Shain et al. “Memory access during incremental sentence processing causes reading time latency”. In: *Proceedings of the Computational Linguistics for Linguistic Complexity Workshop*. Association for Computational Linguistics, 2016, pp. 49–58.

[24] N. J. Smith and R. Levy. “The effect of word predictability on reading time is logarithmic”. In: *Cognition* 128 (2013), pp. 302–319.

[25] M. van Schijndel. “The Influence of Syntactic Frequencies on Human Sentence Processing”. PhD thesis. The Ohio State University, 2016.

[26] M. van Schijndel, A. Exley, and W. Schuler. “A model of language processing as hierarchic sequential prediction”. In: *Topics in Cognitive Science* 5.3 (2013). Ed. by J. Hale and D. Reitter, pp. 522–540.

[27] M. van Schijndel and W. Schuler. “An Analysis of Frequency- and Memory-Based Processing Costs”. In: *Proceedings of NAACL-HLT 2013*. Association for Computational Linguistics, 2013.