A large-scale deconvolutional study of predictability and frequency effects in naturalistic reading

Cory Shain

shain.3@osu.edu

Ohio State

Question

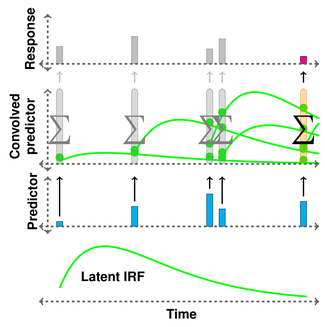
Are there distinct mechanisms for (1) predicting a word vs. (2) retrieving a word from the mental lexicon?

Theoretical Background

- **Yes:**
  - Lexical retrieval cost depends on the strength of a word's representation in memory [18, 3, 10]
  - Retrieval cost is context-independent
  - **Prediction:** Separable effects of *predictability* and *frequency*
- **No:**
  - Comprehenders incur costs for incrementally reallocating resources among possible interpretations [15, 14, 16]
  - No context-independent lexical retrieval mechanism.
  - Frequency effects are subsumed into the probability model
  - **Prediction:** No separable effects of *predictability* and *frequency*

Experimental Background

- Lots of experimental evidence for **Yes**:
  - Additive effects of corpus frequency and cloze predictability [17, 1, 8, 22]; see [21] for review.
- However:
  - Constructed stimuli may introduce task artifacts [4, 11, 2]
  - Cloze poorly differentiates low-probability words [20]
- Can be addressed by naturalistic stimuli with statistical probability estimates. However:
  - *Frequency* and *predictability* are naturally collinear [4]
  - Temporal diffusion may confound word-by-word modeling [5, 19]
- **This study:**
  - Naturalistic data address ecological validity
  - Large-scale data address collinear variables of interest
  - Deconvolutional modeling addresses diffusion of effects



Continuous-time deconvolution

Experimental Design

- Evaluation on 3 large corpora containing over 1M events total:
  - **Natural Stories**, self-paced reading [7]
  - **Dundee**, eye-tracking [13]
  - **UCL**, eye-tracking [6]
- Deconvolutional time series regression [19]
- **Controls:** Sentence position, document position, word rate, word length, saccade length, whether the previous word was fixated
- **Predictors of interest:** unigram log probability, 5-gram surprisal
- Probabilities computed by KenLM models [12] trained on the Gigaword 3 corpus [9]
- **Response:** Log-ms (go-past for eye-tracking)
- By-subject random intercepts, slopes, and impulse response parameters

Results

Comparison	*p*-value
5-gram only vs. baseline	**0.0001*****
Unigram only vs. baseline	**0.0001*****
5-gram + Unigram vs. Unigram-only	**0.0001*****
5-gram + Unigram vs. 5-gram-only	0.1440

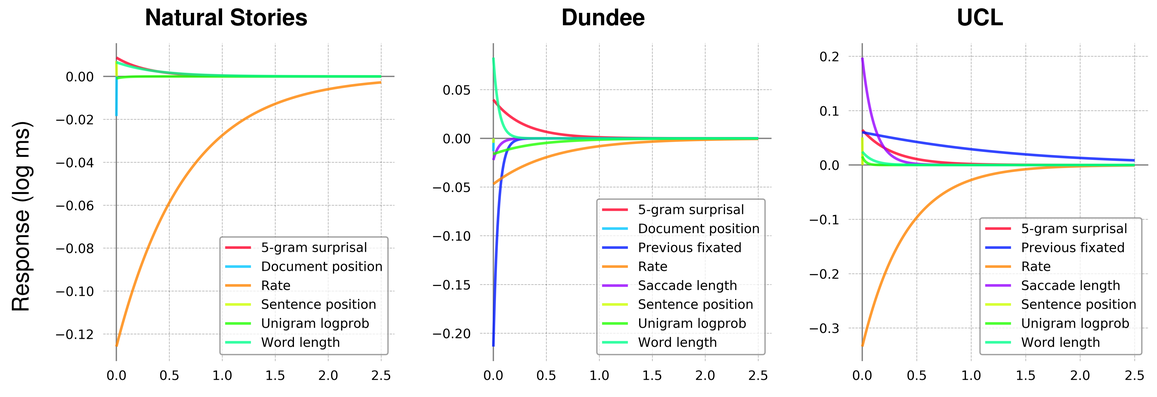
Pooled permutation testing results on out-of-sample data.

Main result: Significant effect of *frequency* over *predictability* but not vice versa, consistent with **No**.

	Effect estimate (log-ms)							
Corpus	SentPos	Trial	Rate	WordLen	SacLen	PrevFix	Unigram	5-gram
Natural Stories	0.0098	-0.0216	-0.3069	—	—	0.0158	**-0.0018**	**0.0174**
Dundee	-0.0085	-0.0052	-0.0277	0.0068	-0.0021	-0.0178	**-0.0067**	**0.0117**
UCL	0.0524		-0.1330	0.0023	0.0221	0.0778	**0.0005**	**0.0184**

Effect estimates (integrals of impulse response functions)

Predictability effects are larger magnitude than *frequency* effects.



Estimated impulse response functions for each predictor by corpus

Conclusion

Results support **No**: no evidence of separable effects of *frequency* and *predictability*.

Finding is at odds with constructed experiments. Possible explanations:

- Frequency effects may exist in naturalistic reading but are too small to be detected.
- Constructed stimuli may introduce confounds:
  - Atypical word distributions
  - Lack of context
  - Suspension of normal communicative function of language
  - Comprehension → problem solving
- Cloze estimates may be too coarse, allowing frequency predictors to capture residual variance due to predictability

References

[1] Ashby, J., Rayner, K., and Clifton, C. *The Quarterly Journal of Experimental Psychology Section A*, 2005.

[2] Campbell, K. L. and Tyler, L. K. *Current opinion in behavioral sciences*, 2018.

[3] Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. *Psychological review*, 2001.

[4] Demberg, V. and Keller, F. *Cognition*, 2008.

[5] Erlich, K. and Rayner, K. *Journal of Verbal Learning & Verbal Behavior*, 1983.

[6] Frank, S. L., Fernandez Monsalve, I., Thompson, R. L., and Vigliocco, G. *Behavior Research Methods*, 2013.

[7] Futrell, R., Gibson, E., Tily, H. J. ., Blank, I., Vishnevetsky, A., Piantadosi, S., and Fedorenko, E. In Calzolari, N., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S., and Tokunaga, T., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 5 2018.

[8] Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., and Rayner, K. *Journal of Experimental Psychology: General*, 2011.

[9] Graff, D., Kong, J., Chen, K., and Maeda, K. English Gigaword Third Edition LDC2007T07, 2007.

[10] Harm, M. W. and Seidenberg, M. S. *Psychological review*, 2004.

[11] Hasson, U. and Honey, C. J. *NeuroImage*, 2012.

[12] Heafield, K., Pouzyrevsky, I., Clark, J. H., and Koehn, P. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 8 2013.

[13] Kennedy, A., Pynte, J., and Hill, R. In *Proceedings of the 12th European conference on eye movement*, 2003.

[14] Levy, R. *Cognition*, 2008.

[15] Norris, D. *Psychological review*, 2006.

[16] Rasmussen, N. E. and Schuler, W. *Cognitive science*, 2018.

[17] Rayner, K., Ashby, J., Pollatsek, A., and Reichle, E. D. *Journal of Experimental Psychology: Human Perception and Performance*, 2004.

[18] Seidenberg, M. S. and McClelland, J. L. *Psychological review*, 1989.

[19] Shain, C. and Schuler, W. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018.

[20] Smith, N. J. and Levy, R. *Cognition*, 2013.

[21] Staub, A. *Language and Linguistics Compass*, 2015.

[22] Staub, A. and Benatar, A. *Psychonomic Bulletin \& Review*, 2013.

To appear as: Shain, C. (2019). A large-scale study of the effects of word frequency and predictability in naturalistic reading. In *NAACL 2019*.