Bayesian surprise during incremental anticipatory processing: A re-analysis of Nieuwland et al. (2017)

Shaorong Yan[1], Gina R. Kuperberg[2,3], T. Florian Jaeger[1]
[1]University of Rochester; [2]Tufts University; [3]Massachusetts General Hospital

Background & Questions

- N400 is a negative-going ERP component that is sensitive to semantic expectation. (Kutas & Federmeier, 2011)
- Pre-nominal words reported to also elicits N400(-like) effect when incompatible with expected noun. (DeLong et al., 2005; Wicha et al., 2004)

The day was breezy so the boy went outside to fly **a** kite / **an** airplane.

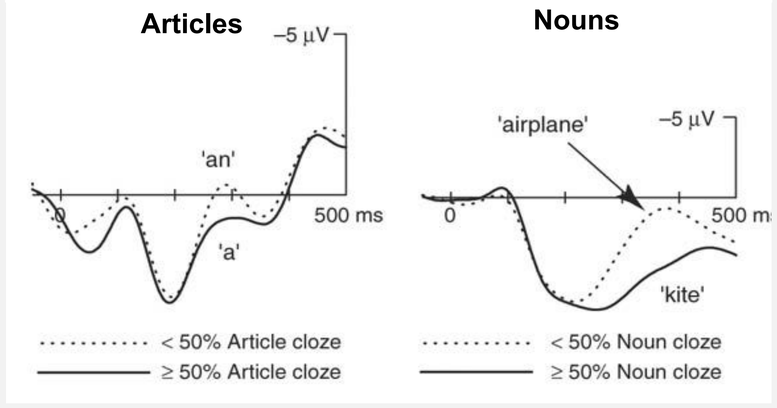

Figure 1, DeLong et al. (2005)

- Questions:
  - What underlying processes lead to pre-nominal effects?
  - What relevant measures should thus be expected to correlate with pre-nominal effects?

Approximating Semantic Updating

- When pre-nominal information is deterministically predictive, e.g. gender. Bayesian surprise (BS) equals surprisal of the pre-nominal information.

$$BS = \sum_i \text{p}\,(NOUN_i \mid masc, \text{context}) * \log \frac{\text{p}\,(NOUN_i \mid masc, \text{context})}{\text{p}\,(NOUN_i \mid \text{context})} = \cdots$$
$$= -\log \text{p}\,(\text{masc} \mid \text{context}) = \text{Surprisal of masc}$$

- When pre-nominal information is non-deterministically predictive, e.g. a/n ('an awesome kite'). Bayesian surprise does not equals surprisal.

$$BS = \sum_i \text{p}\,(NOUN_i \mid \text{'a'}, \text{context}) * \log \frac{\text{p}\,(NOUN_i \mid \text{'a'}, \text{context})}{\text{p}\,(NOUN_i \mid \text{context})} = \cdots$$
$$= -\log \text{p}\,(\text{'a'} \mid \text{context}) + \text{f}\,(\ \text{p}(\text{'a'} \mid NOUN_i, \text{context})\ ) < \text{Surprisal of 'a'}$$

Task Analysis

- Common explanation for pre-nominal effect assume predictions at different levels of language processing

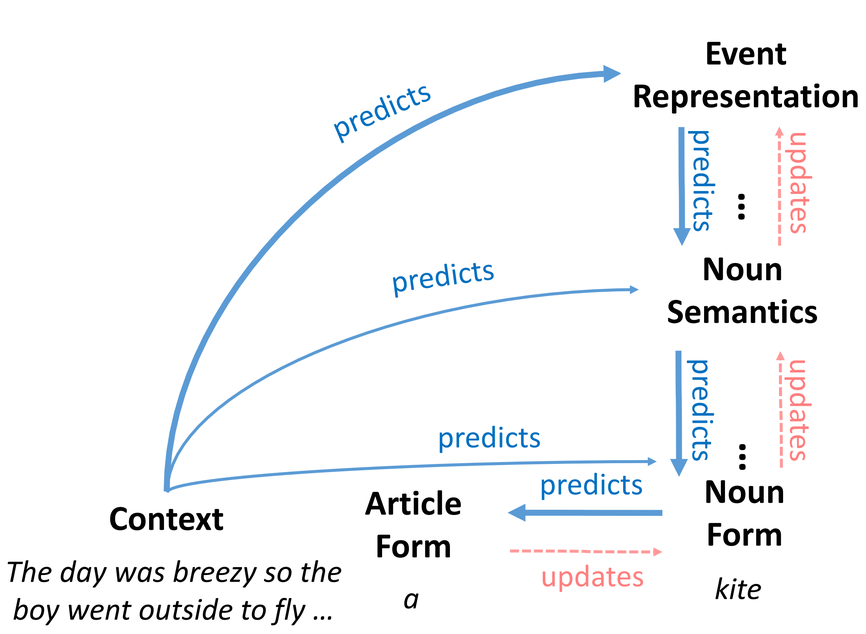

Figure 2. Possible predictive chain when the article is encountered. Predictions of the event representation cascade into predictions about possible upcoming nouns as well as form predictions of the noun, and form predictions of the article (blue arrows). When the article appears, predictions may be updated at any of these levels (red arrows).

- Unexpected information leads to large shift in predictions about the messages being conveyed (Kuperberg, 2016; Kuperberg & Jaeger, 2016; Rabovsky et al., submitted)

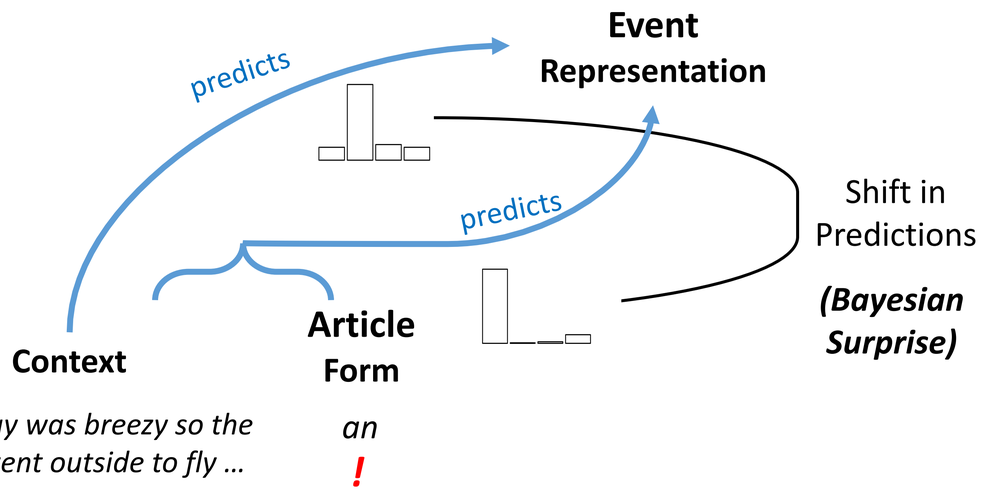

Figure 3, (Unexpected) New information leads to changes in expected event representation.

Correlation between Bayesian Surprise and Surprisal in Language Use

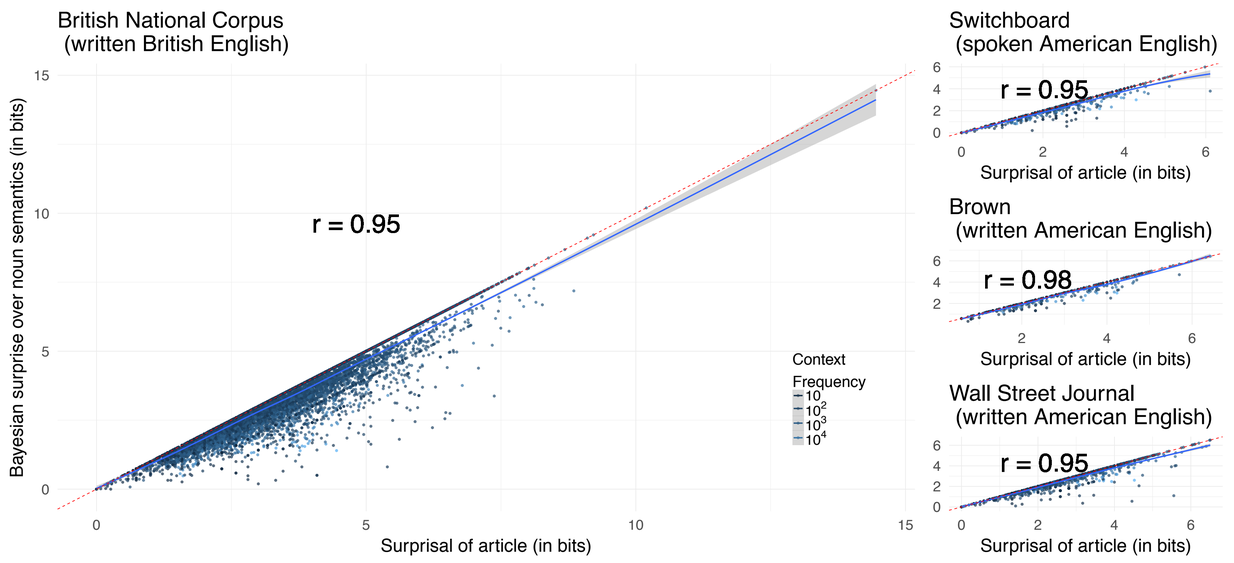

Figure 4. Correlation between the article's surprisal (log-transformed bi-gram probability) and the Bayesian surprise over the distribution of the upcoming noun incurred on the article.

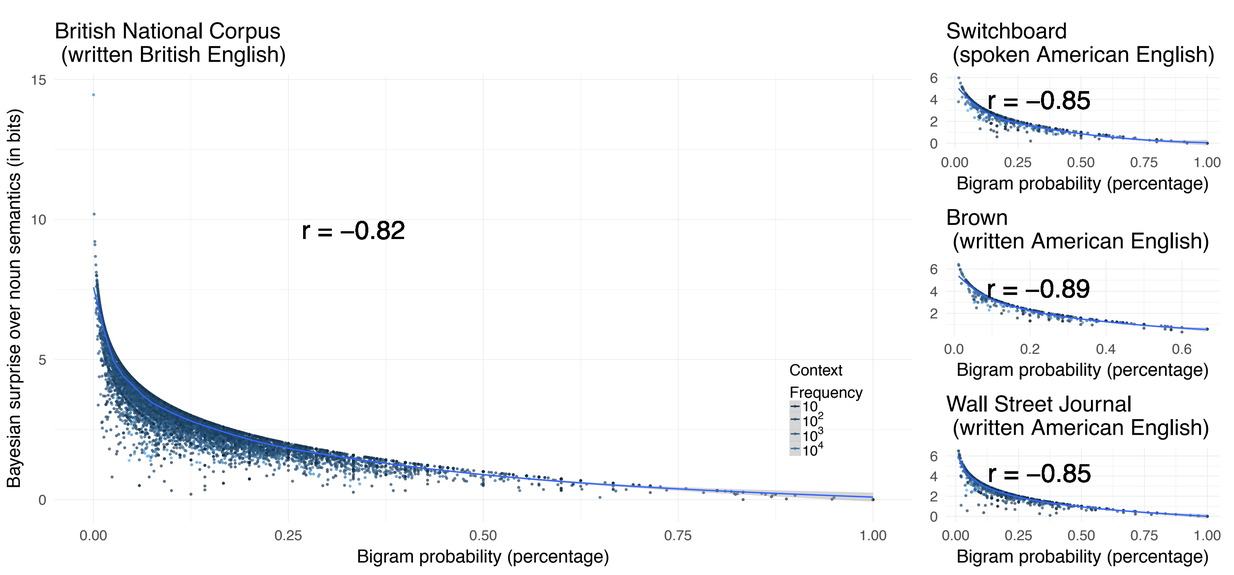

Figure 5. Correlation between the article's predictability (bi-gram probability) and the Bayesian surprise over the distribution of the upcoming noun incurred on the article.

Bayesian Surprise

- Bayesian surprise: Relative entropy or Kullback-Leibler divergence between a prior and posterior distribution
- Measures how much update occurs in the generative model after encountering new information
- More unexpected pre-nominal information leads to larger change in expected event representation (see Figure 3 above)
- Offers an unified account of both N400 effect on content word (nouns) and corresponding effect on pre-nominal information

Surprisal and Pre-nominal N400

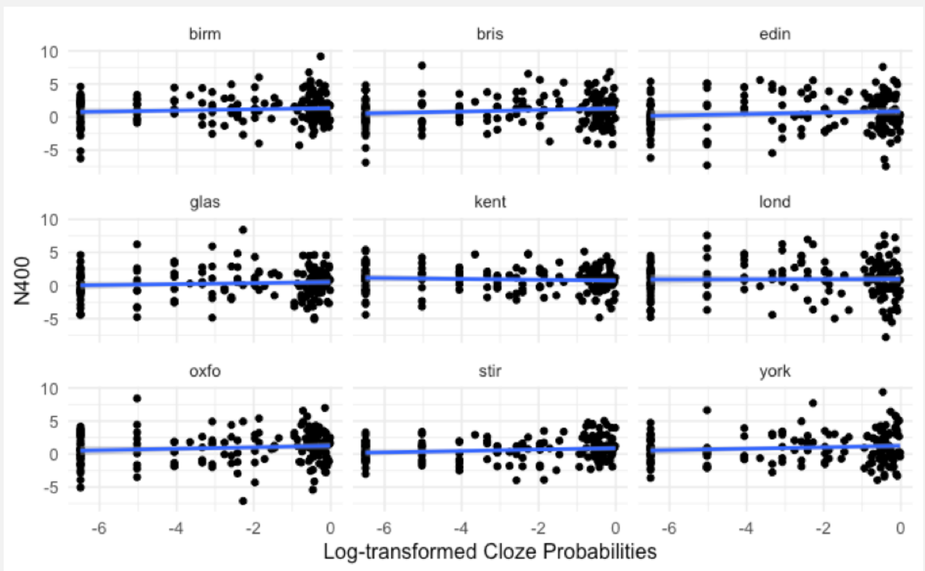

Figure 6. Re-analysis of the ERP data shared by Nieuwland et al. (2017). Surprisal of the article is a better linear predictor for N400 amplitude on the article ($p < 0.016$) than cloze probabilities of the articles ($p < 0.13$). The same holds for surprisal and N400 evoked by the noun (data not shown).

References:

1. Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647.

2. DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, *8*(8), 1117-1121.

3. Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of cognitive neuroscience*, *16*(7), 1272-1288.

4. Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, cognition and neuroscience*, *31*(5), 602-616.

5. Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, cognition and neuroscience*, *31*(1), 32-59.

6. Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2017). I Like Coffee With Cream And Dog? Change In An Implicit Probabilistic Representation Captures Meaning Processing In The Brain. *bioRxiv*, 138149.

7. Nieuwland, M., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., ... & Mézière, D. (2017). Limits on prediction in language comprehension: A multi-lab failure to replicate evidence for probabilistic pre-activation of phonology. *bioRxiv*, 111807.

For more details, see our draft on *BioRxiv*

Yan, S., Kuperberg, G. R., & Jaeger, T. F. (2017). Prediction (Or Not) During Language Processing. A Commentary On Nieuwland et al. (2017) And Delong et al. (2005). *bioRxiv*, 143750.

Link: https://www.biorxiv.org/content/early/2017/05/30/143750

For downloading the poster