**Investigating the role of context in comprehension using topical surprisal: An fMRI study**

Shohini Bhattasali & Philip Resnik (University of Maryland)
shohini@umd.edu

**Introduction**: The notion of surprisal from information theory has been very prevalent in psycholinguistic modeling, following the work of Hale (2001) and Levy (2008). Generally, the theory of surprisal proposes that probabilistic predictions made by comprehenders yield variability in word-by-word processing difficulty: when surprisal is high, the current word is unexpected and cognitive processing effort increases accordingly. While previous work has probed the neural correlates of lexical and syntactic surprisal using computational measures (Brennan et al. 2016), to our knowledge modelling with surprisal has not been extended beyond sentences to include broader context. Xu. et al (2005) note, “important, but thus far unexplored, is the role of context within narrative, as cognitive demands evolve and brain activity changes dynamically as subjects process different narrative segments.”

In this study, our main goal is to investigate how *topical context* affects our predictions about the next word--and based on those predictions, our processing--using an analysis of fMRI timecourses collected during naturalistic language comprehension. We propose a new metric, “topical surprisal”, which we define using the weighted average of a word's probability given a topic, where weights are the (posterior) probability the context is about that topic; topics can be defined and probabilities estimated using a topic model (LDA, Blei et al., 2003). Using this metric, we can represent how expected the word is given what has been discussed previously in the context and more broadly, study the role of context in comprehending language.

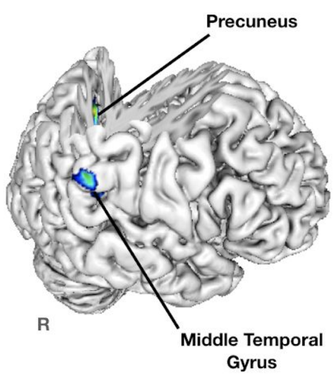
**Methods**: Participants (n=51, 32 female) listened to *The Little Prince*’s audiobook for 1 hour and 38 minutes. Participants' comprehension was confirmed through multiple-choice questions (90% accuracy, SD = 3.7%). Functional scans were acquired using multi-echo planar imaging sequence (ME-EPI) (TR=2000ms; TE's=12.8, 27.5, 43ms; FA=77 degrees; FOV=240.0mm X 240.0mm; 2X image acceleration; 33 axial slices, voxel-size 3.75 x 3.75 x 3.8mm). Preprocessing was done with AFNI16 and ME-ICAv3.2 (Kundu et al. 2012).

Using the wrapper for Mallet LDA (McCallum, 2002) in the Gensim toolkit (Rehurek & Sojka 2010), we estimated a 100-topic model using the Brown corpus (Francis & Kučera 1964), which covers diverse topics in 500 texts across 15 genres. We compute topical surprisal for each of the 6,243 non-function words in the audio sample using the paragraph containing the word as its context (Fig. 1B). In addition to this regressor, we included in the GLM analysis (SPM12) four regressors of non-interest: timestamp of each word offset, log-frequency of each word in movie subtitles (Brysbaert & New 2009), and pitch (f0) and intensity (rms) of the narrator's voice.

**Results**: We observe the largest clusters for topical surprisal in the right Precuneus and right Middle Temporal Gyrus (Fig. 1A). The whole-brain effects were FWE-corrected (T-score > 5.3).

**Conclusion**: Our results corroborate previous work on lexical access and semantic integration (Binder et al. 2009, Graves et al. 2010, Hickok & Poeppel 2007, Hagoort & Indefrey 2014) which illustrates the scope of these cognitive processes beyond the sentence level. Specifically in terms of context and discourse-level phenomena, Raposo et al. (2013) observed significant activation in the right MTG while investigating semantic processing of sentences with a preceding context. Similarly, the right Precuneus has been implicated in various language processing tasks utilizing context such as an fMRI study on narrative shifts (Whitney et al. 2009), an fMRI study contrasting sentences and narratives (Xu et al. 2005), and a PET study on processing incoherent narratives (Maguire et al. 1999). The pattern of activation for topical surprisal also differs from those reported for lexical surprisal (bilateral ATL & left IFG) and syntactic surprisal (bilateral ATL & left IPL) by Brennan et al. (2016). Thus, our results support the centrality of these two regions in processing contextual information during language comprehension and suggest topical surprisal as a cognitively plausible metric.

(A)




(B)

*During the fifty-four years that I've lived on this planet, I've only been disturbed three times. The first time was twenty-two years ago, by some scatterbrain who fell from god knows where. He made the most dreadful noise, and I made four mistakes in a sum. The second time was eleven years ago, by an attack of rheumatism. I don't get enough exercise. I don't have time to stroll about. I am a man of consequence. The third time—well, this is it! I was saying, then, five-hundred-and-one million—*

**Paragraph topic distribution: T58 (0.1651), T2 (0.0759), T55 (0.0633), (all other topics < .05)**

**T58 top-10 words: 'time', 'make', 'day', 'year', 'good', 'leave', 'give', 'long', 'people', 'man'**
**T2 top-10 words: 'child', 'school', 'teacher', 'parent', 'formula', 'number', 'achievement', 'learn', 'problem', 'study'**
**T55 top-10 words; 'strength', 'exercise', 'kate', 'foam', 'scotty', 'work', 'leg', 'roberts', 'muscle', 'weight',**

**Pr(*year*|T58) = 0.021023600687223077, Pr(*year*|T2) = 0.007018716577540107, Pr(*year*|T55) = 0.004793028322440087**

$$\text{surprisal}(\textit{year}) = -\log \sum_{\text{topic in Topics}} P(\textit{year} \mid \text{topic})\, P(\text{topic} \mid \text{context})$$

Fig. 1(A): Whole brain contrast image with significant clusters for topical surprisal after FWE voxel correction with $p < 0.05$ (B): Sample excerpt with topical surprisal example

**References**:


Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex, 19(12), 2767-2796.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993-1022.
Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W. M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. Brain and language, 157, 81-94.
Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior research methods, 41(4), 977-990.
Ferstl, E. C. (2010). Neuroimaging of text comprehension: Where are we now. Italian Journal of Linguistics, 22(1), 61-88.
Francis, W. N. & H. Kučera. (1964). A Standard Corpus of Present-Day Edited American English, for use with Digital Computers. Providence, Rhode Island: Department of Linguistics, Brown University. Revised 1971. Revised and amplified 1979.
Graves, W. W., Binder, J. R., Desai, R. H., Conant, L. L., & Seidenberg, M. S. (2010). Neural correlates of implicit and explicit combinatorial semantic processing. Neuroimage, 53(2), 638-646.
Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. Annual review of neuroscience, 37, 347-362.
Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies (pp. 1-8). Association for Computational Linguistics.
Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. Nature reviews neuroscience, 8(5), 393.
Kundu, P., Inati, S. J., Evans, J. W., Luh, W. M., & Bandettini, P. A. (2012). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *Neuroimage*, *60*(3), 1759-1770.
Levy, R. (2008). Expectation-based syntactic comprehension. Cognition, 106(3), 1126-1177.
Maguire, E. A., Frith, C. D., & Morris, R. G. M. (1999). The functional neuroanatomy of comprehension and memory: the importance of prior knowledge. Brain, 122(10), 1839-1850.
McCallum, A. K. (2002). Mallet: A machine learning for language toolkit. http://mallet. cs. umass. edu.
Raposo, A., & Marques, J. F. (2013). The contribution of fronto-parietal regions to sentence comprehension: insights from the Moses illusion. NeuroImage, 83, 431-437.
Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks.
Siebörger, F. T., Ferstl, E. C., & von Cramon, D. Y. (2007). Making sense of nonsense: an fMRI study of task induced inference processes during discourse comprehension. Brain Research, 1166, 77-91.
Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., & Kircher, T. (2009). Neural correlates of narrative shifts during auditory story comprehension. Neuroimage, 47(1), 360-366.
Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & Van den Bosch, A. (2015). Prediction during natural language comprehension. Cerebral Cortex, 26(6), 2506-2516.
Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: emergent features of word, sentence, and narrative comprehension. Neuroimage, 25(3), 1002-1015.