Sandbox Series: The Role of Context Maintenance and Updating in Predictive Sentence Processing

Yasemin Gokcen, David Noelle, & Rachel Ryskin
Department of Cognitive and Information Science
University of California, Merced

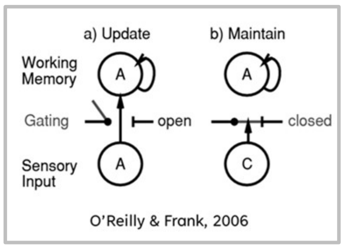
BACKGROUND

Language

- During language comprehension, humans routinely use information from preceding context to predict upcoming input (e.g., Ryskin & Nieuwland, 2023).
- The memory representation of the context is noisy (Hahn et al., 2022).

Working memory & cognitive control

- Computational models of working memory mechanisms associated with prefrontal cortex (PFC) use *gating* to capture the ability to maintain and update contextual/task-relevant information in memory (e.g., O'Reilly et al., 2002).
- The anatomy and physiology of the PFC is thought to support this mechanism (Noelle, 2012; Kriete et al., 2013).



O'Reilly & Frank, 2006

- Memory is updated when gate opens to relevant information.
- Memory is maintained when gate is closed to irrelevant information.
- Models/the brain learn when the gate should be open vs. closed

QUESTION

What neural working memory mechanisms support context maintenance and updating during predictive sentence processing?

GENERAL APPROACH AND METHODOLOGY

- Compare performance across language models with different architectures & their ability to predict neural measures of real-time language processing.

Language Models:

- Recurrent Neural Network (RNN) and Long Short-Term Memory models (Van Schijndel & Linzen, 2018) both trained on the Wikitext 2 dataset (Merity et al., 2016) and tested on Natural Stories
  - **LSTM has PFC-like gating mechanisms** (Hochreiter & Schmidhuber, 1997), RNN does not (Elman, 1990)

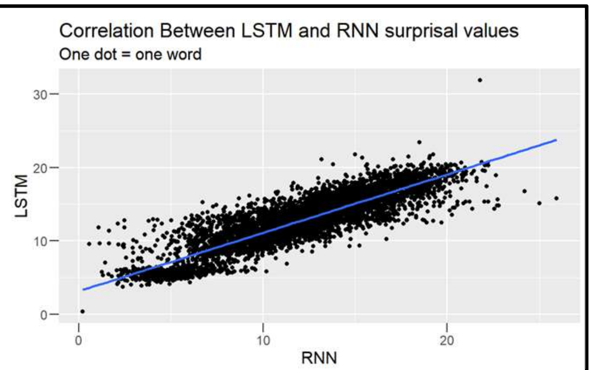
Neural activity:

- EEG passive listening task (+ comprehension questions)
  - Natural Stories corpus (Futrell et al., 2017)
  - 10 stories/narratives (~5 min each) with low frequency words/structures
  - Data collection in progress

MODEL PERFORMANCE

On Natural Stories

- Overall consistency in surprisal estimates across models
- Surprisingly, RNN seems to be producing generally lower values compared to LSTM.



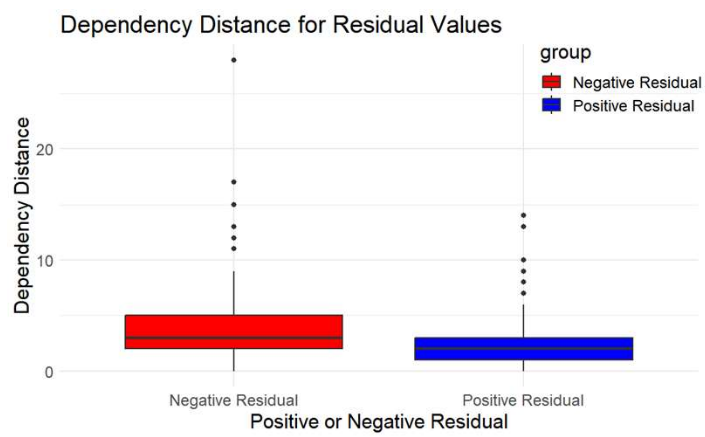
PRELIMINARY RESULTS

- Examining context for words with extreme residual values in a linear model predicting LSTM surprisal from RNN surprisal (i.e., cases where LSTM surprisal is under- or over-predicted by RNN surprisal)
- Example sentences containing words with extreme residuals:

Negative (RNN surprisal > LSTM surprisal):
"Later the same day, the Commanding General of the Eighth Air *Force* stated that, in fact, a weather balloon had been recovered by the Roswell Army Air Field personnel, rather than a "flying saucer."

Positive (LSTM surprisal > RNN surprisal):
"People were purchasing bulbs at higher and higher prices, intending to *re-sell* them for a profit."

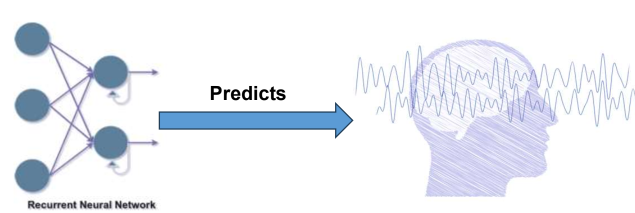
Average dependency distance for 100 most negative residuals (M = 4.38, SD = 4.32) was larger than 100 most positive residuals (M = 2.51, SD = 2.53 ($t(159.76) = 3.73$, $p < .001$, Cohen's $d = 0.53$)



Words that are more distant from the heads of the dependency are associated with higher surprisal in RNN than LSTM → Importance of gating for *maintaining* elements of context.

FUTURE WORK

- Further investigation of properties of sentence context associated with gating mechanisms (updating, maintenance)
- Predicting EEG responses during story listening from language model metrics (e.g., surprisal, residual)



REFERENCES

1. Ryskin & Nieuwland, 2023, Prediction during language comprehension: what is next?. Trends in cognitive sciences, 27(11), 1032–1052.
2. Hahn et al., 2022. A resource-rational model of human processing of recursive linguistic structure. Proceedings of the National Academy of Sciences, 119(43), e2122602119
3. O'Reilly et al., 2002, Prefrontal Cortex and Dynamic Categorization Tasks: Representational Organization and Neuromodulatory Control. Cerebral Cortex, 12(3), 246–257.
4. Noelle, 2012, On the neural basis of rule-guided behavior. Journal of Integrative Neuroscience, 11(4), 453–475.
5. Kriete et al., 2013, Indirection and symbol-like processing in the prefrontal cortex and basal ganglia. Proceedings of the National Academy of Sciences, 110(41), 16390–16395.
6. O'Reilly & Frank, 2006, Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. Neural Computation, 18(2), 283–328.
7. van Schijndel & Linzen, 2018, A Neural Model of Adaptation in Reading. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 4704–4710.
8. Merity et al., 2016, *Pointer Sentinel Mixture Models* (arXiv:1609.07843). arXiv.
9. Hochreiter & Schmidhuber, 1997, Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780
10. Elman, 1990, Finding Structure in Time. *Cognitive Science*, *14*(2), 179–211.
11. Futrell et al., 2021, The Natural Stories corpus: A reading-time corpus of English texts containing rare syntactic constructions. *Language Resources and Evaluation*, *55*(1), 63–77.