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

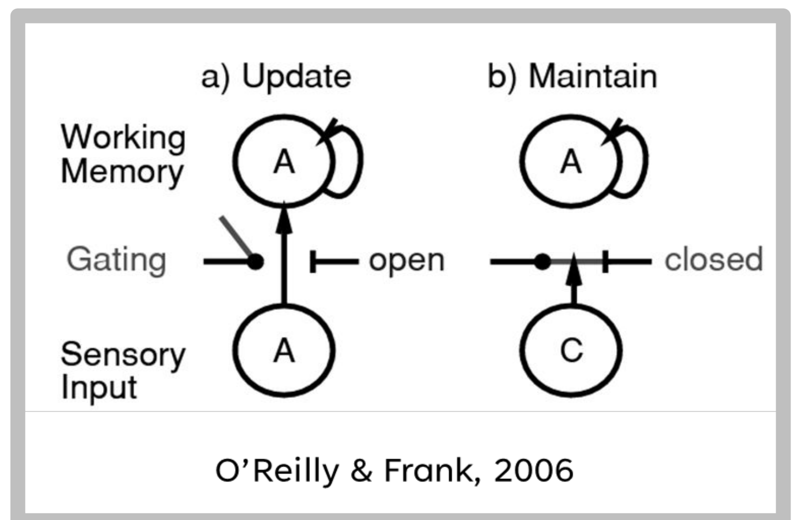
BACKGROUND

Language

- During language comprehension, humans routinely use information from preceding context to predict upcoming input[1].
- The memory representation of the context is noisy[2].
- Surprisal measures of how unexpected a word is, given context[3]
  - Lower surprisal = better word prediction performance

Working Memory & Cognitive Control

- Computational models of working memory (WM) mechanisms associated with prefrontal cortex (PFC) use *gating* to capture the ability to maintain and update contextual/task-relevant information in memory[4].
- The anatomy and physiology of the PFC is thought to support this mechanism[5,6].



O'Reilly & Frank, 2006

- Memory is updated when gate opens to relevant information.
- Memory is maintained when gate is closed to irrelevant information.
- Gate control develops via reinforcement learning.

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, as well as their ability to predict neural measures of real-time language processing.

Language Models:

- Recurrent Neural Network (RNN) and Long Short-Term Memory models[8] both trained on the Wikitext 2 dataset[9] and tested on Natural Stories
  - **LSTM has PFC-like gating mechanisms[10]**, RNN does not[11]

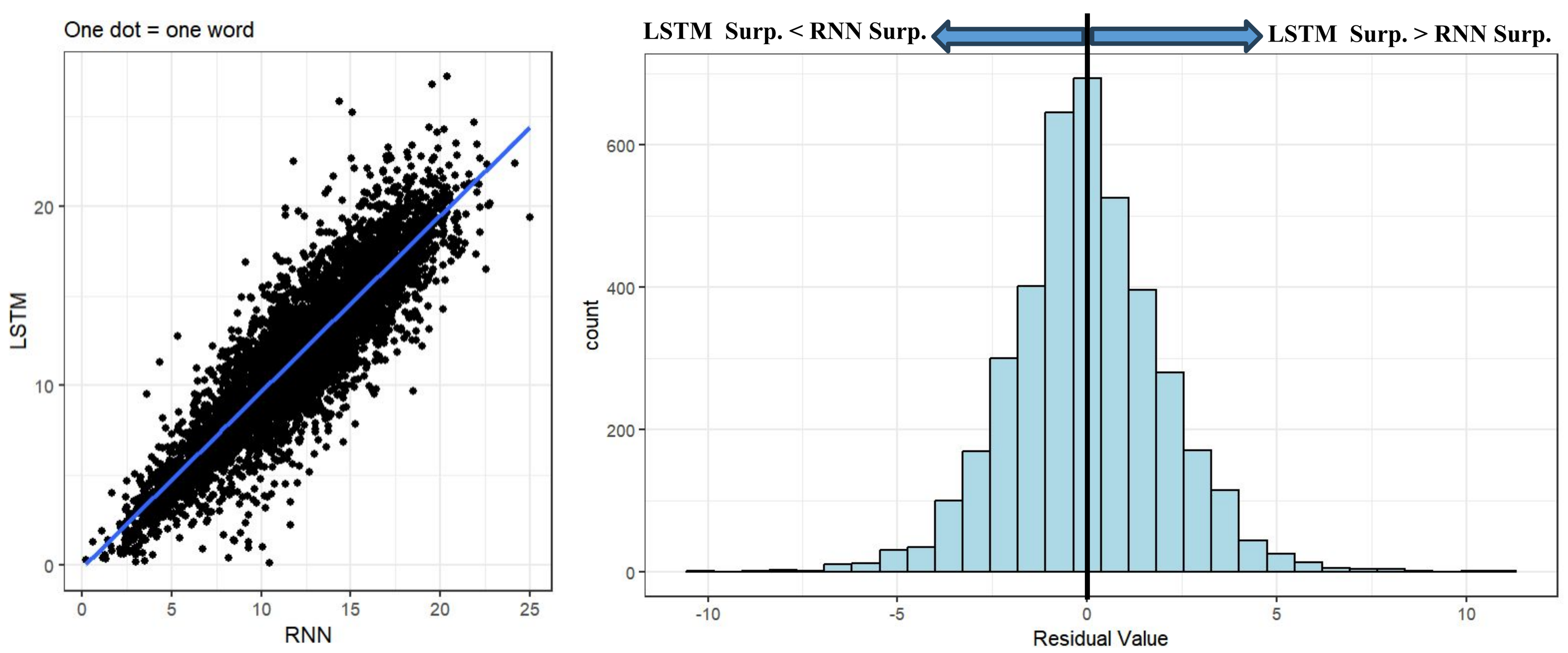
Neural activity:

- EEG passive listening task (+ comprehension questions)
  - Natural Stories corpus[12]
  - 10 stories/narratives (~5 min each) with low frequency words

COMPARATIVE SURPRISAL ACROSS MODELS

Correlation between LSTM and RNN Surprisal Values

- Overall consistency in surprisal estimates across models that are tested on Natural Stories (LSTM exhibits somewhat lower surprisal values compared to RNN)



WM DEMANDS AND RELATIVE SURPRISAL ACROSS MODELS

- According to Dependency Locality Theory[13], sentences with longer-distance dependencies place larger demands on WM:

» Integration (8 variants based on different weighting of parts of speech) & Storage

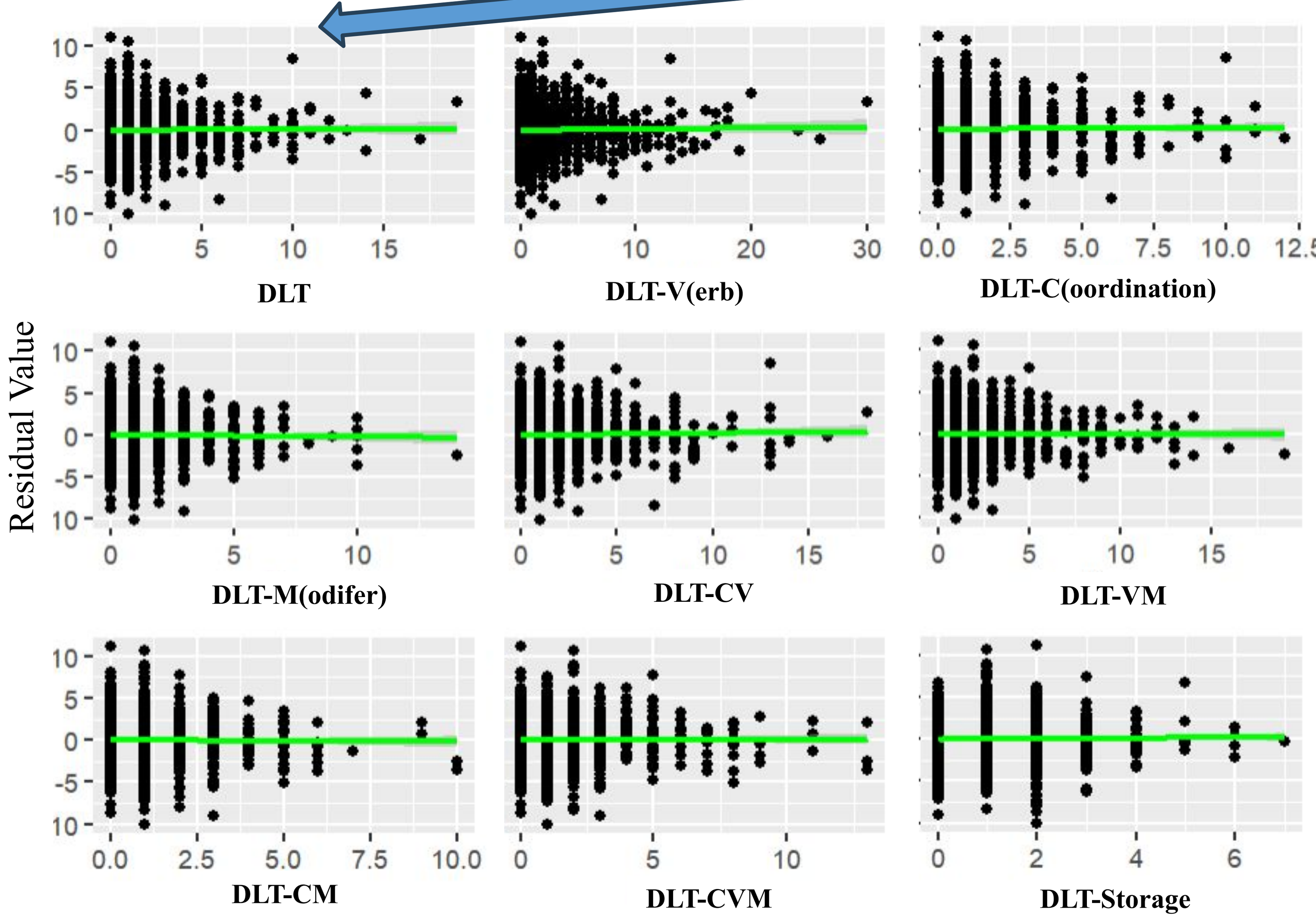
…he ***drew*** out the boar's *tongue* he had ***put*** in his hunting pouch…

"drew"= governing verb
"tongue" = direct object

Later connected to the verb "put"

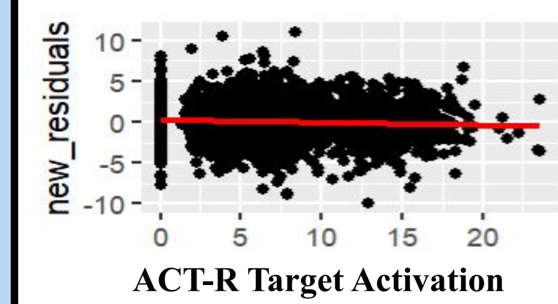
- "Drew" and "put" imply a sequence of actions.
- Intervened by "the boar's tongue he had"

Longer distance = more processing



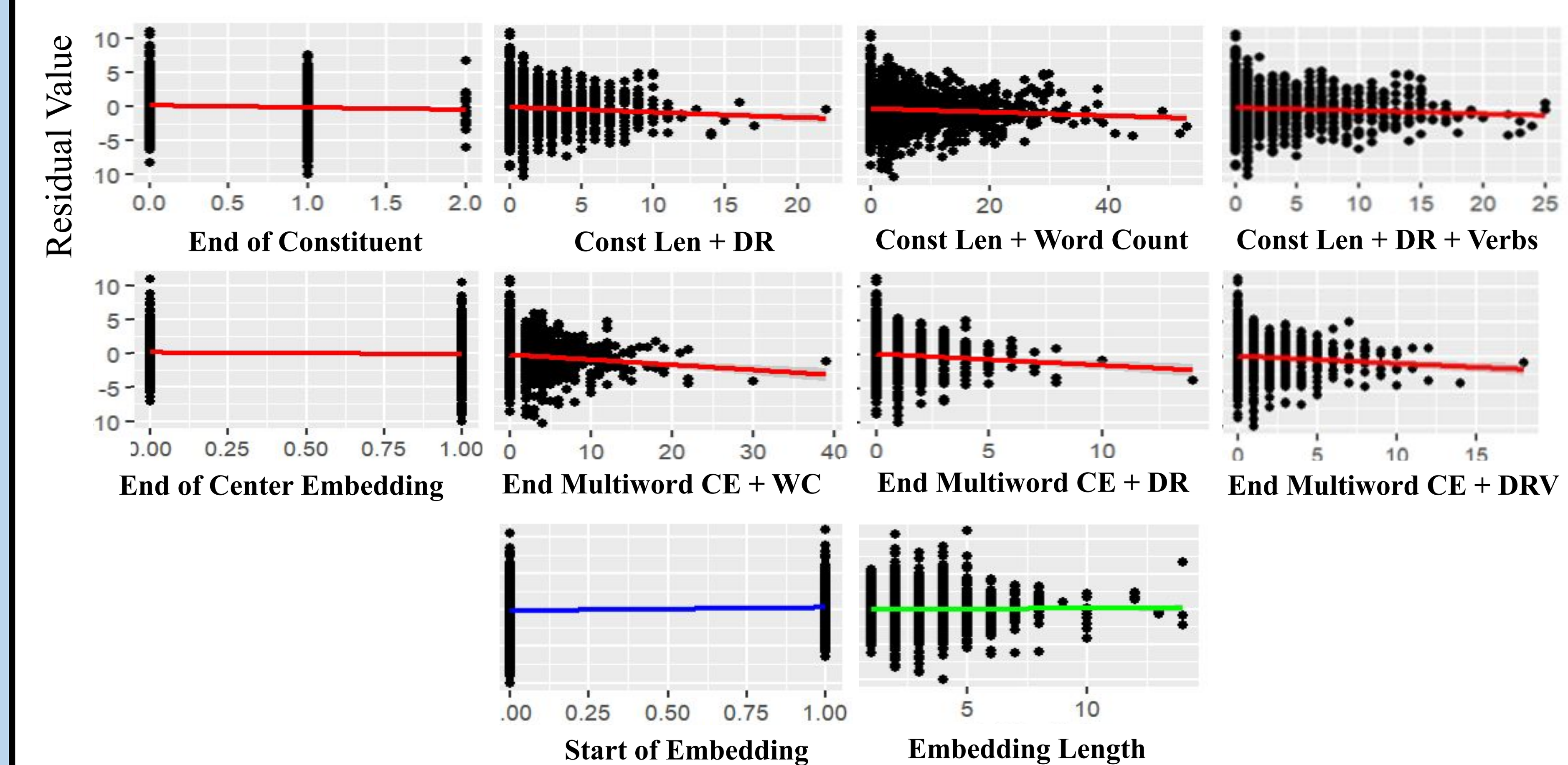
DLT integration cost (all variants) and storage show **no significant (Bonferroni corrected α = 0.0055)** correlation.

WM DEMANDS AND RELATIVE SURPRISAL ACROSS MODELS



ACT-R[15] provides a measure of WM retrieval costs and does not consider storage/maintenance. LSTM produces lower relative surprisal as ACT-R target activation increases. This relationship suggests that a model with gating more closely aligns with WM performance as implemented in ACT-R.

- Left-Corner Parsing[14] examines incremental sentence comprehension based on lexical and grammatical word matches;
  - Constituents (i.e., phrases) and embeddings (i.e., phrase within a phrase in a sentence)
  - involves markers for phrase endings, distance metrics, and memory depth tracking.



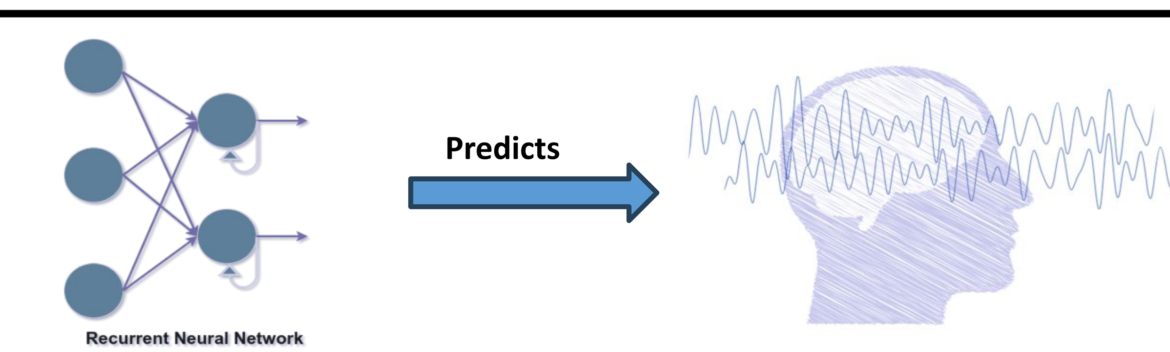
Predictors related to the costs of maintaining constituents and center embeddings in WM are all **significantly negatively correlated with Residuals (Bonferroni corrected α = 0.005).** Embedding depth is **not significantly correlated.** Start of Embedding is **significant and positive**.

This shows a relationship between the size of constituent and embedding and the utility of gating. LSTM does relatively better, compared to RNN, as these measures of WM demand increase.

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. Levy (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.
4. O'Reilly et al. (2002). Prefrontal cortex and dynamic categorization tasks: Representational organization and neuromodulatory control. *Cerebral Cortex*, 12(3), 246–257.
5. Noelle (2012). On the neural basis of rule-guided behavior. *Journal of Integrative Neuroscience*, 11(4), 453–475.
6. 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.
7. 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.
8. van Schijndel & Linzen (2018). A neural model of adaptation in reading. *Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing*, 4704–4710.
9. Merity et al. (2016). *Pointer Sentinel Mixture Models* (arXiv:1609.07843).
10. Hochreiter & Schmidhuber (1997) Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
11. Elman (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211.
12. 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.
13. Gibson (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, Language, Brain*, 2000, 95-126.