Effects of context and proficiency on processing difficulty during L2 reading

Chen Liu, Rachel Ryskin, Zenaida Aguirre-Muñoz
Department of Cognitive and Information Science
University of California, Merced

Background

- Previous research indicates that word length, frequency, and predictability in context affect reading speed. Longer, less frequent, less predictable words are read more slowly. [1][2]
- Reading in a **second language (L2)** poses challenges that vary based on the reader's proficiency level, affecting speed and comprehension. [3]
- Word property effects on L2 versus L1 readers are similar, but the magnitudes of effects and their functional forms may differ, especially for those with lower L2 proficiency. [3]
- **Lexicon-Context Tradeoff:** Balance between reliance on word-specific properties and contextual information during reading.
- As proficiency in a second language improves, readers rely less on word frequency and more on contextual cues while reading. [3]

Current Study

- Prior studies focused on single sentences, limiting the range of word predictability.
- The current study tested the lexicon-context tradeoff in a natural reading setting using the GECO-CN corpus, which includes eye-tracking data from L2 speakers reading a novel [4]

Research aim: How does L2 proficiency affect surprisal and word frequency effects on reading time during sentence processing in natural reading contexts?

Dataset

GECO-CN corpus:

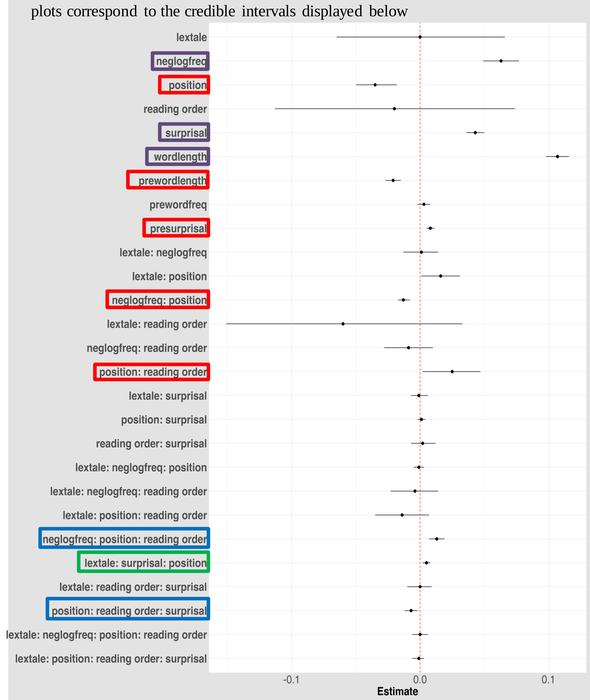
- 30 native Chinese speakers reading novel. Each participant read about 50,000 words in each language, resulting in a total of 2,500 sentences per participant.
- **Reading order:** 15 participants read Chapters 1–7 in English (L2) and Chapters 8-13 in Chinese (L1), another 15 participants read Chapters 1–7 in Chinese (L1) and Chapters 8–13 in English (L2).
- **Eye-tracking Measures**: First Fixation Duration (FFD), Single Fixation Duration (SFD), Gaze Duration (GD), Total Reading Time (TRT), and Go Past Time (GPT).
- **Proficiency Assessment:** LexTALE scores. [5]

Surprisal Values: GPT-2 surprisal [6]
Word Frequency Data: From SUBTLEX-UK [7]

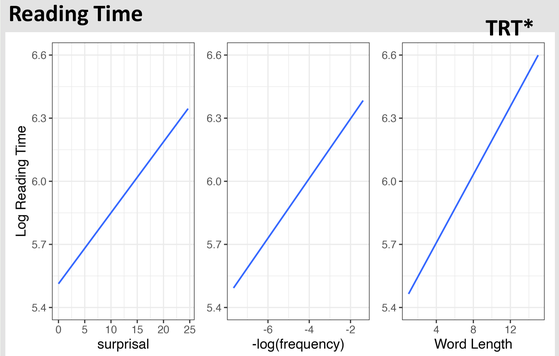
Results

Fixed Effects Coefficients with 95% Credible Intervals (CrIs) from Bayesian multilevel model of TRT

Boxes indicate CrIs which do not include 0. The colored boxes indicate which plots correspond to the credible intervals displayed below



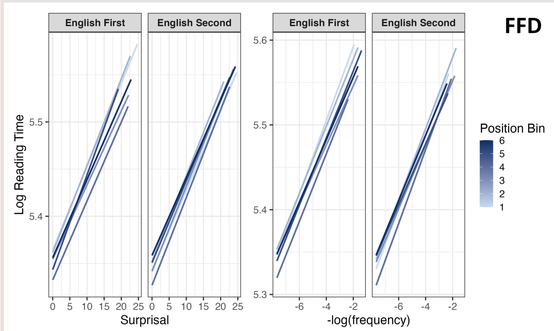
Effects of Surprisal, Frequency, and Word Length on Reading Time



- Across all models, reading times increase for words that are higher in surprisal, lower in frequency, and longer (replicating many prior findings).

- Effects of position, surprisal of the preceding word, length of the preceding word, and all 2-way interactions are inconsistent across models.

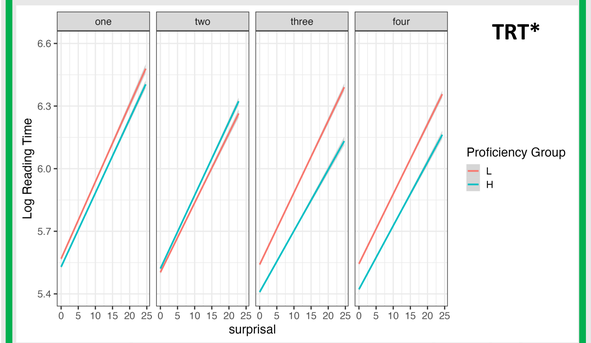
Effects of Surprisal and Frequency on Reading Time Across Sentence Positions in Two Reading Orders



*3-way interactions emerge in models of late (GD, GPT, TRT) but not early (SFD, FFD) reading measures.

- For visualization, the half of the novel that each participant read (in English) is divided into 6 position bins.
- Effects of surprisal and word frequency decrease in later position bins, especially for those who read in English first relative to those who read the second half of the novel in English.

Effect of Surprisal on Reading Time for High and Low Proficiency Readers Across Four Story Segments.



- For visualization, participants categorized into high (scores >75 on LexTale) and low proficiency groups for visualization (N=15 per group) and the novel is divided into 4 segments.
- The effect of surprisal on reading times decreases as readers advance through the story, especially for the high proficiency group.

Conclusions

- As L2 readers progress through a story, effects of frequency and surprisal on reading times decrease.
- Readers may use larger discourse context (outside of the sentence) to predict upcoming words more effectively. (Note that GPT-2 surprisal estimates do not take larger discourse into account)
- In particular, high proficiency L2 readers may use larger discourse context more effectively than low proficiency L2 readers.
- The results suggest that language proficiency affects the balance between lexical and contextual processing, consistent with a lexicon-context tradeoff.

Limitations

- LexTALE assessments only measure vocabulary size.
- Background knowledge and the sequence of language exposure (reading order) could significantly influence reading strategies, complicating the assessment of context effects in L2 reading.

References

[1] Ehrlich, S. F., & Rayner, K. (1981) Journal of verbal learning and verbal behavior
[2] Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004) European journal of cognitive psychology
[3] Berzak, Y., & Levy, R. P. (2022). Open Mind
[4] Sui, L., Dirix, N., Woumans, E., & Duyck, W. (2022) Behavior Research Methods
[5] Lemhöfer, K., & Broersma, M. (2012) Behavior research methods
[6] Misra, K. (2022) arXiv
[7] Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014) Quarterly journal of experimental psychology