

INTRODUCTION

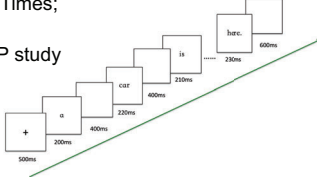
- Integrative/predictive processes play an important role in reading as readers use context in reading words and integrating them into the text meaning^{1,2,3};
- Surprisal* measures the probability of a word given the preceding context. It has been investigated using behavioral (RT, eye-tracking) and brain activity measures (fMRI, EEG);
- Previous studies found that N400 an indicator of ease of meaning processing is affected by *surprisal* in L1 readers^{1,4,5,6,7};
- However, there are no ERP studies investigating *surprisal* effects in L2 reading English.

OBJECTIVES

- Do integrative/predictive processes differ between L1 & L2 reading?
 - we examine whether the effect of *surprisal* in L2 reading is observable in N400 window as previously reported in L1 studies;
 - we examine whether L2 readers are sensitive to both lexical and syntactic information when reading a given word.
 - we compare language models of different architectures in computing surprisal. Would the models that predict L1 also predict L2?

MATERIALS & METHODS

- Participants: native English (33) & highly proficient speakers of English, who are native Chinese (35);
- Materials: 63 two-sentence passages in English from New York Times;
- Procedure: ERP study



- 7 types of surprisal was calculated for each word using language models trained on Wikipedia dataset: $S(w_{t+1}) = -\log P(w_{t+1}|w_{1:t})$

	N-gram surprisal			PCFG surprisal		RNNG surprisal	Transformer surprisal
	Lexical	Syntactic	Total	Total	Syntactic	Tree structures + word (hierarchical)	Word sequences (sequential)
Training structure	Word/POS sequences (sequential)			Tree structures + word (hierarchical)		Tree structures + word (hierarchical)	Word sequences (sequential)
Context length	N - 1			All preceding words		All preceding words	All preceding words
Primary information	Lexical	Part-of-speech	Lexical & part-of-speech	Lexical & syntactic	Syntactic	Lexical & syntactic	Lexical

RESULTS

Does surprisal affect N400 of L2 readers as in L1 readers?

- Likelihood ratio test on each type of surprisal was used to compare each surprisal regression to a null regression. Results showed a robust effect of surprisal on N400 amplitude in both groups of readers (see Table 1).

Table 1: Likelihood ratio test results for each type of surprisal

Surprisal model	L1 reader	L2 reader
PCFG syntactic surprisal	$\chi^2(1) = .905, p = .342$	$\chi^2(1) = 6.999, p = .008$
PCFG total surprisal	$\chi^2(1) = 7.493, p = .006$	$\chi^2(1) = 12.634, p < .001$
N-gram lexical surprisal	$\chi^2(1) = 15.539, p < .001$	$\chi^2(1) = 10.900, p < .001$
N-gram syntactic surprisal	$\chi^2(1) = 1.829, p = .176$	$\chi^2(1) = 4.262, p = .039$
N-gram total surprisal	$\chi^2(1) = 18.749, p < .001$	$\chi^2(1) = 14.774, p < .001$
RNNG surprisal	$\chi^2(1) = 32.410, p < .001$	$\chi^2(1) = 14.936, p < .001$
Transformer surprisal	$\chi^2(1) = 37.904, p < .001$	$\chi^2(1) = 10.563, p = .001$

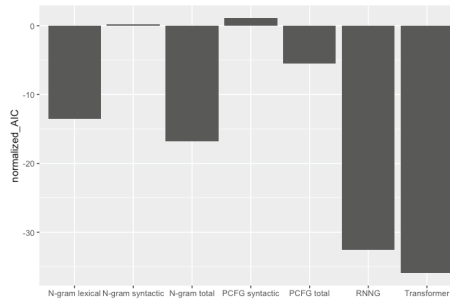
Are both groups sensitive to lexical and syntactic information when reading a word?

- Largely, the majority (5 out of 7) of surprisal models suggest little differences among two groups. Surprisal models accounting for lexical and syntactic information significantly predicted N400 amplitude of both groups.
- The only difference was that unlexicalized syntactic surprisal measures of PCFG and N-gram models did not account for the significant amount of variance in the L1 readers group.

What is the best fitting model for L1 & L2 readers?

- Akaike Information Criterion (AIC) was used to find the best fitting model for each group of readers.
- L1 group: Transformer and RNNG

Figure 1: Normalized AICs for surprisal models for L1 readers



- L2 group: RNNG, N-gram total, and PCFG total

Figure 2: Normalized AICs for surprisal models for L2 readers

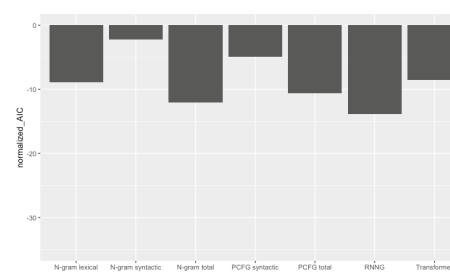


Figure 3: Grand averaged waveforms for N-gram syntactic surprisal

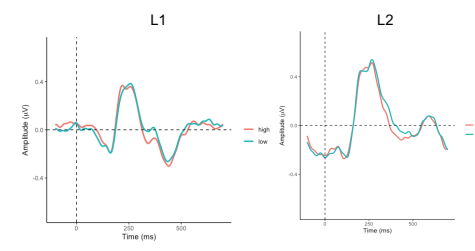
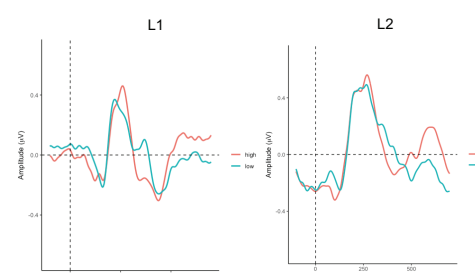


Figure 4: Grand averaged waveforms for RNNG surprisal



CONCLUSIONS

- Surprisal affects the N400 amplitude of L1 & L2 readers. Integrative/predictive processes play an important role in L1 & L2 reading: both groups register this process at around 400 ms, where semantic processing occurs.
- In both groups, the models that explain best N400 amplitude are the ones that contain both syntactic and lexical information in the calculation of surprisal (lexicalized models) in the preceding context.
- All best fitting models for L1 & L2 readers contain lexical and syntactic information (explicitly or implicitly). Overall, there is little evidence for L1-L2 differences in sensitivity to lexical or syntactic information of a word.
- However, the two groups differ in the amount of variance explained by the surprisal models: models better predict L1 data, which may reflect a greater experience in reading English for L1 than L2. It may be possible that the degree to which L1 and L2 readers rely on syntactic or lexical information differs when processing and integrating the next word.

REFERENCES

- 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.
- De Varda, A., & Marelli, M. (2022, November). The Effects of Surprisal across Languages: Results from Native and Non-native Reading. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2022* (pp. 138-144).
- Cevoli, B., Watkins, C., & Rastle, K. (2022). Prediction as a basis for skilled reading: insights from modern language models. *Royal Society Open Science*, 9(6), 211837
- Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2013). Word surprisal predicts N400 amplitude during reading.
- Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and language*, 140, 1-11.
- Goodkind, A., & Bicknell, K. (2021). Local word statistics affect reading times independently of surprisal. *arXiv preprint arXiv:2103.04469*.
- Van Schijndel, M., & Schuler, W. (2015). Hierarchic syntax improves reading time prediction. In *Proceedings of the 2015 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies* (pp. 1597-1605).

ACKNOWLEDGMENTS & CONTACTS

We want to thank NSF for funding this project (awarded to LC and CP). We also want to thank Carrie Demmans, Alona Fyshe, Daniela Teodorescu and Shannon Clark for computing surprisal values for all our stimuli.

Gaisha Oralova: oralovag@pitt.edu
Lin Chen: LINC@pitt.edu
Charles Perfetti: perfetti@pitt.edu

