

Context and POS in Action: A Comparative Study of Chinese Homonym Disambiguation in Human and Language Models

Chenwei Xie^{1*}, Matthew King-Hang Ma^{1*}, Wenbo Wang¹, and William Shiyuan Wang¹
¹Research Centre for Language, Cognition and Neuroscience, Department of Language Science and Technology, HKPolyU



Department of Language Science and Technology
語言科學及技術學系

Background

- Ambiguity is pervasive in language, yet humans can resolve it using **contextual cues and part-of-speech (POS)** information (Grindrod et al. 2014).
- It remains unclear if **LLMs** are making use of these information in a **similar** or a **different** way from **humans** to generate contextualized embeddings, especially in **Chinese** where there is a lack of explicit POS marking (Wang, 1973).

Questions

- How do context similarity and POS jointly influence homonym disambiguation in humans and LLMs?
- Can model-derived metrics explain or predict human behavioral responses during homonym processing?

Results

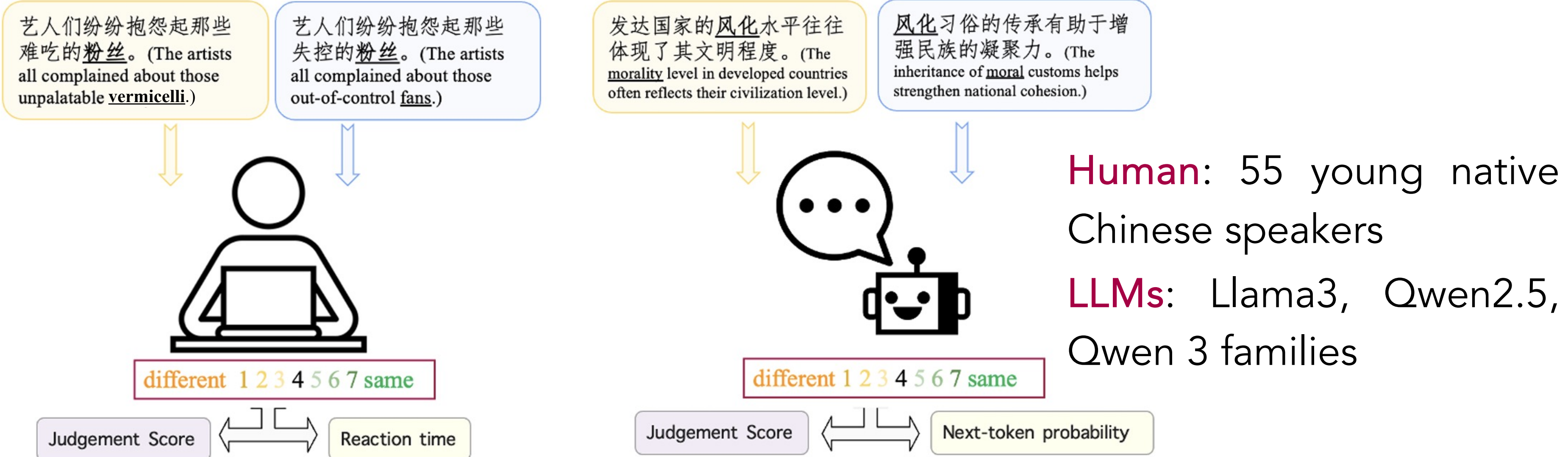
- Human performed significantly better than all included LLMs
- AoA** influenced **human** but not model responses
 - Suggest fundamental divergence between biological and transformer-based learning.
- Humans and LLMs were similarly influenced by **context similarity**
 - Contextual similarity helps when the word sense is the same but hinders when it's different
- Only **humans** leveraged **POS** information to resolve ambiguity in challenging condition (**same-context, diff-sense**)
 - May explain why LLMs performed relatively poorly.



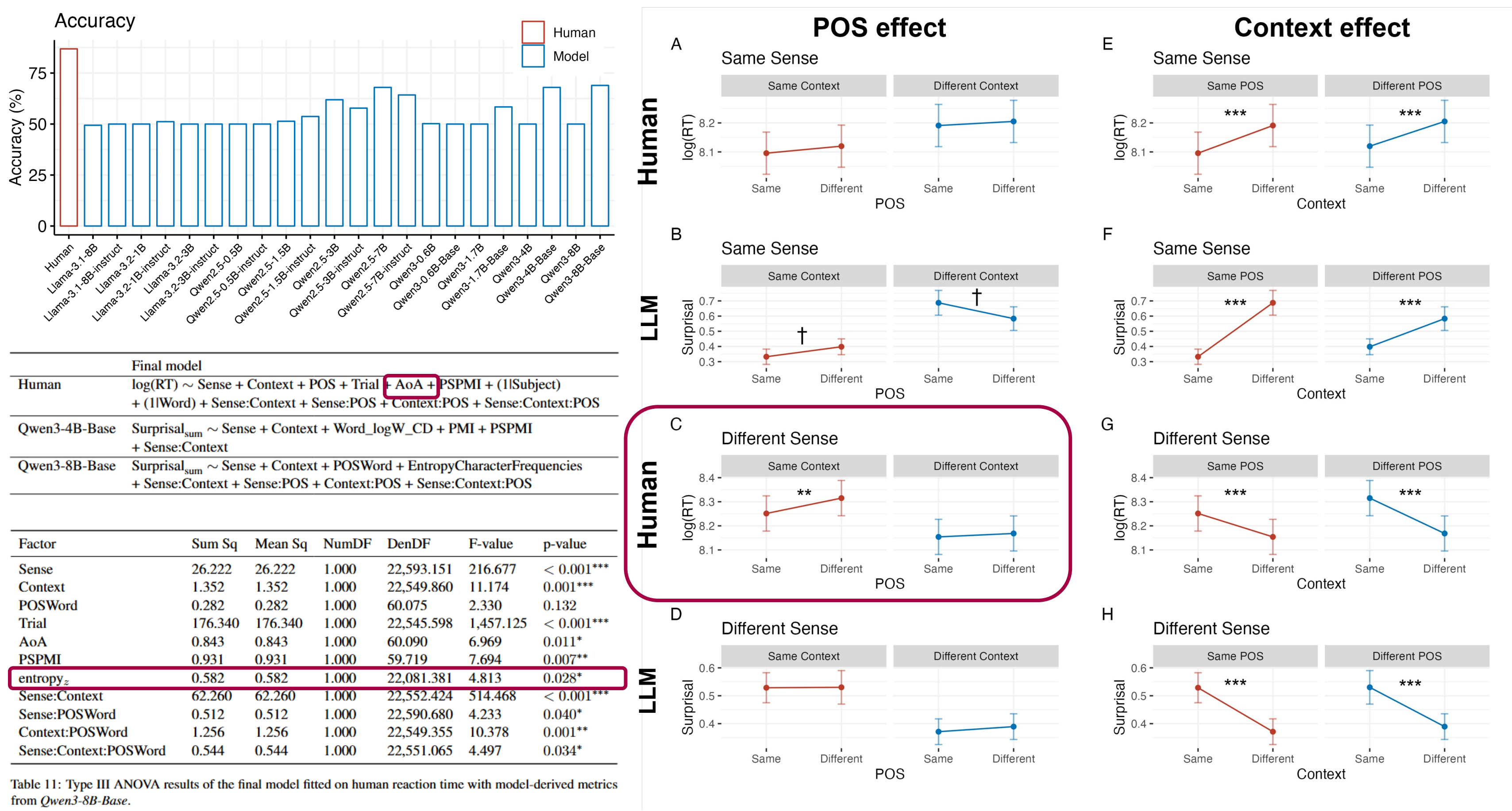
Methods

- 64** noun or verb Chinese homonyms (32 **same-POS** and 32 **different-POS**).
- 8 pairs** of sentences for each homonym (manipulating **POS** and **Context**).

粉丝 (same-POS, same-Context)	风化 (different-POS, different-Context)
艺人们纷纷抱怨起那些难吃的粉丝。 The artists all complained about those unpalatable vermicelli.	发达国家的 风化 水平往往在体地貌的形成与频繁的风化密切相关。 The level of morality in developed countries often closely related to frequent reflects their level of weathering .



- LMEMs** with human RTs and LLM surprisal values as dependent variables, fixed effects of **Context**, **POS**, **Sense**, and **psycholinguistic covariates** (AoA, PSPMI, etc.).
- Incorporate **surprisal**, **entropy**, and **angular similarity** from LLMs into human RT model.



- Higher surprisal and entropy from 4B and 8B models are associated with longer RT respectively
 - 4B model maybe better at capturing the **prediction error**, while the 8B model maybe better at capturing the **uncertainty** in the prediction space.
- Adding **angular similarity improved** the fit of models predicting human RTs
 - It captures **additional variances** that are not explained by surprisal or entropy.
 - Angular similarity became useful starting from **middle-to-late** layers, suggesting these layers may encode richest semantic information that is relevant to homonym processing.

Conclusion

- Human responses are shaped by both traditional psycholinguistic properties like **AoA**, as well as model-derived metrics of **expectation-based processing** and **semantic integration**.
- The interplay of context and POS reveal similarity and differences between humans and LLMs and highlight the present **limitations of LLMs** in capturing these nuanced processes.

References

- Grindrod, C. M., Garnett, E. O., Malyutina, S., & den Ouden, D. B. (2014). Effects of representational distance between meanings on the neural correlates of semantic ambiguity. *Brain and Language*, 139, 23–35.
- Wang, W. S.-Y. (1973). The Chinese Language. *Scientific American*, 228(2), 50–60. <https://doi.org/10.1038/scientificamerican0273-50>