TILBURG UNIVERSITY

Does multimodal pre-activation influence linguistic expectations in LLMs and humans?

Sasha Kenjeeva | Giovanni Cassani | Noortje Venhuizen | Afra Alishahi

COGSCI & AI RESEARCH CENTER

Background

- When we read, we pre-activate linguistic information of incoming words
- Rommers (2013) found that we also pre-activate the shape of the incoming word's referent
  - when expecting 'moon' in 'Armstrong landed on the…', 'tomato' is read faster than 'rice' due to shape congruency
- Visual features of referents seem to play a role when predicting words in context: also for plausible sentences?

Context (x 37)

The impatient man kept glancing at his . . .

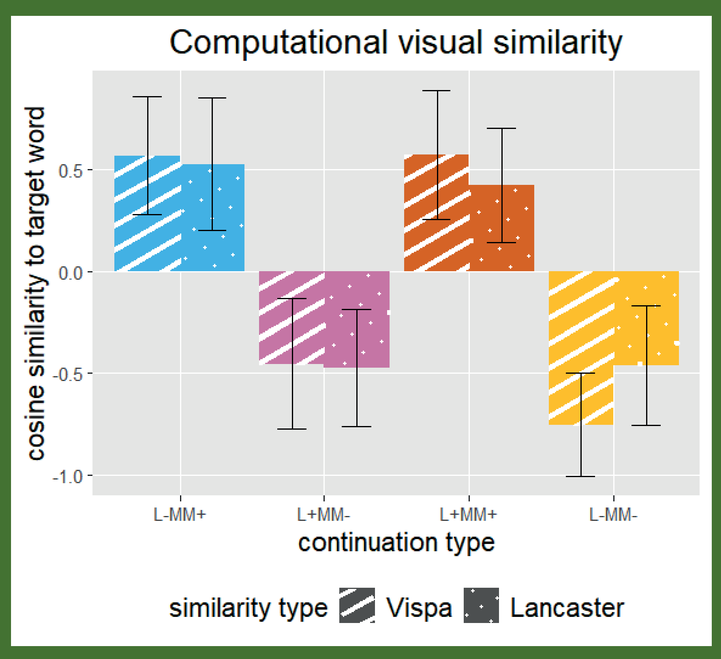
Continuation (concrete & known)

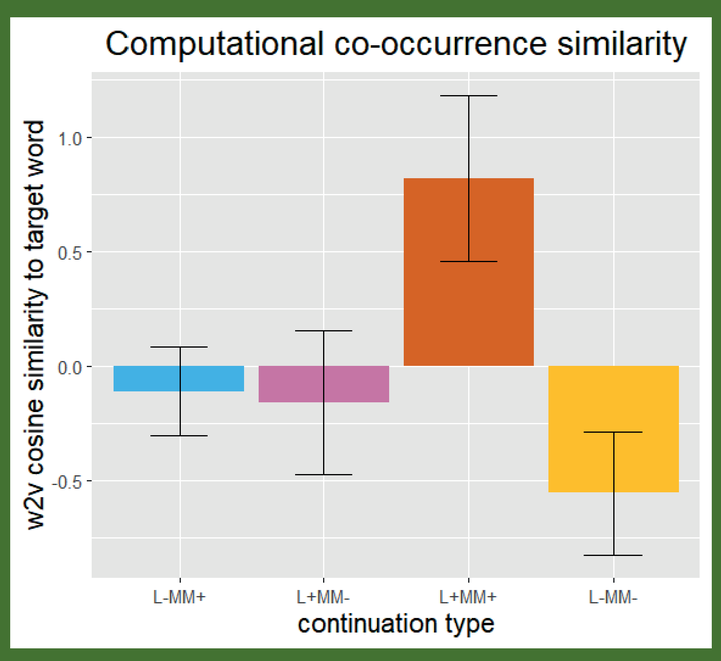
WATCH
COMPASS
WIFE
PHONE
DOG

Type

Target
L- MM+
L+MM-
L+MM+
L- MM-

Data-driven design





Norming studies

The impatient man kept glancing at his watch.

(1) Visual similarity

Which of the two referents below is more visually similar to a **watch**?

(2) Co-occurrence similarity

Consider the word *watch*. Which of the words below is more likely to appear in similar sentences to *watch*?

dog — compass

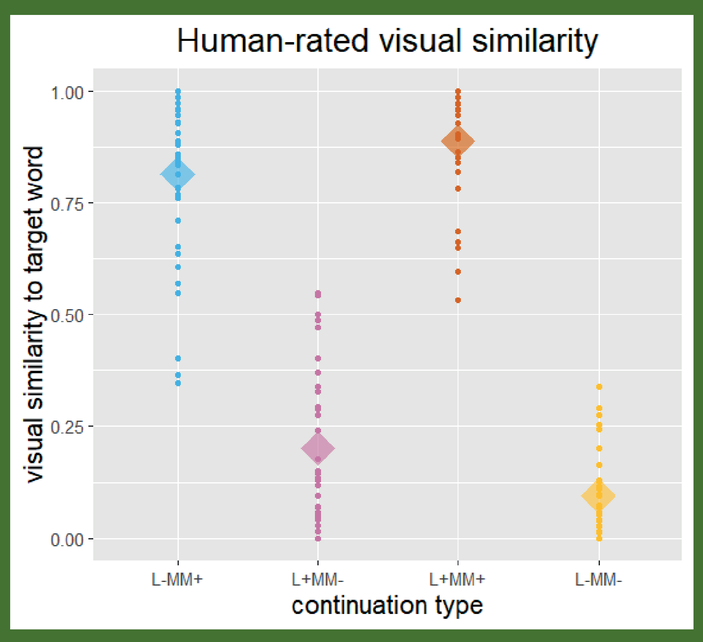
Similarity: n times label chosen/ n times label appears

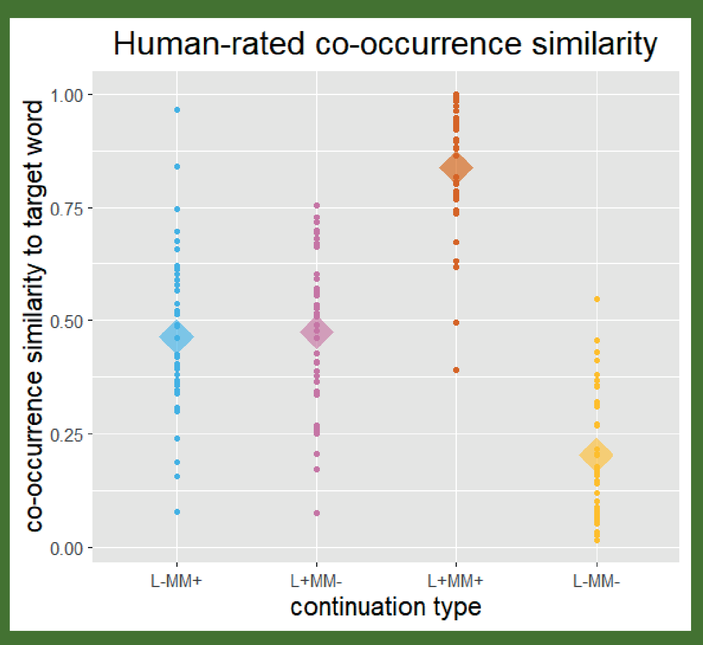
(3) Plausibility

How plausible is the situation described by the sentence?

completely implausible — completely plausible

Norming results





71% sentences plausible

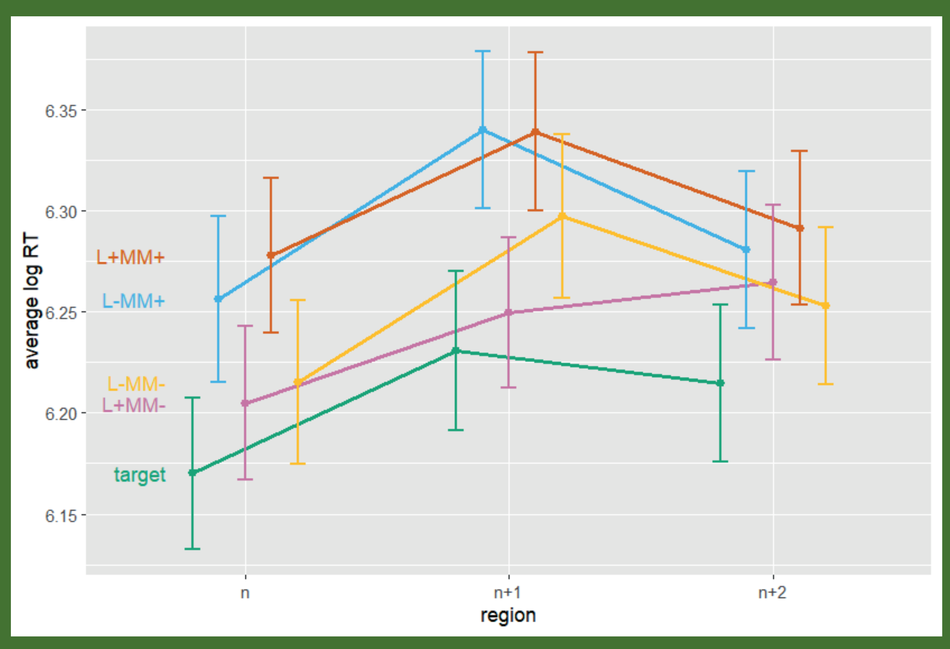
Self-paced reading study

Expectations…

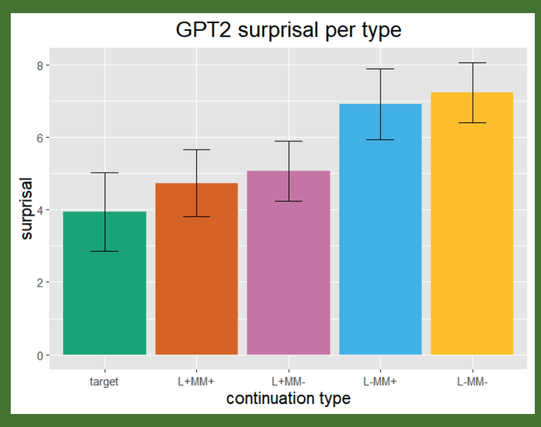
Target < L+MM+ < L+MM- ~ L- MM+ < L- MM-

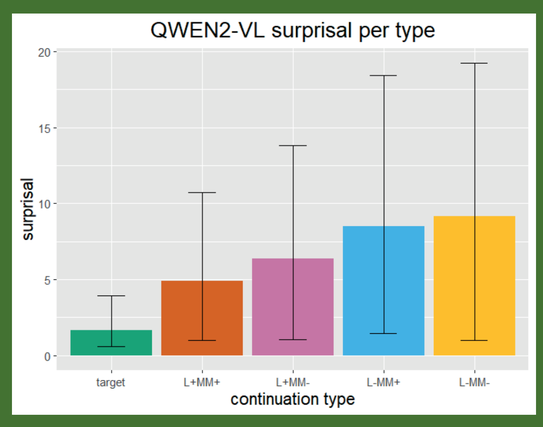
(critical prediction: L+MM- ~ L- MM+)

…Reality



Large language models (LLM and VLM)





Best model so far for human data (R2 marg.: 0.037; R2 cond.: 0.637):

log RT at n+1 ~ in Cloze (0/1) + visual sim. (PCA) + co-occurrence sim. (PCA) + log probability gpt2 + plausibility + word frequency + word length + word position + trial number + w2v sim. between verb and continuation + subordinative/infinitive sentence start + (1|subject)

LLMs follow Cloze, while visual similarity **inhibits** human RTs

References

[1] Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). *Behav. Res. Methods, 52*, 1271-1291. [2] Günther, F., Marelli, M., Tureski, S., & Petilli, M. A. (2023). *Psych. Rev., 130*(4), 896. [3] Peelle, J. E., Miller, R. L., Rogers, C. S., Spehar, B., Sommers, M. S., & Van Engen, K. J. (2020). *Behav. Res. Methods, 52*, 1795-1799. [4] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). [5] Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). *Neuropsychologia, 51*(3), 437–447. [6] Wang, P. et al. (2024). *arXiv:2409.12191.*

a.b.kenjeeva@tilburguniversity.edu