From grounded spaces to linguistic prediction: Multimodal word meaning in context

TILBURG UNIVERSITY

Sasha Kenjeeva | Giovanni Cassani | Noortje Venhuizen | Afra Alishahi

COGSCI & AI DEPARTMENT

Motivation & RQ

- Most research on language & multimodality in isolated word meaning representations
- We process in context → are there multimodal activations when pre-activating words in context?

Context (x 37)

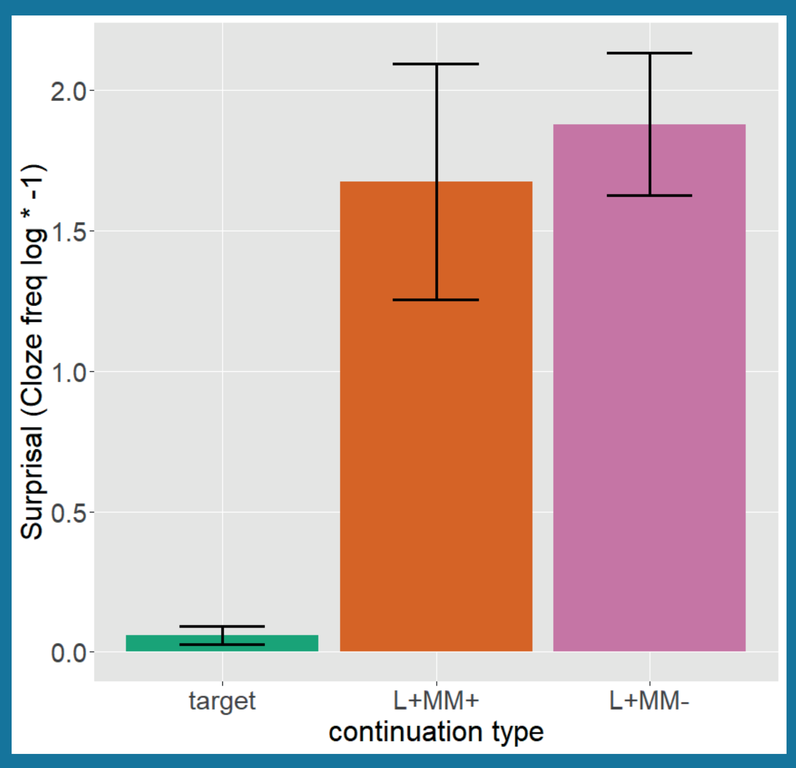
The impatient man kept looking at his . . .

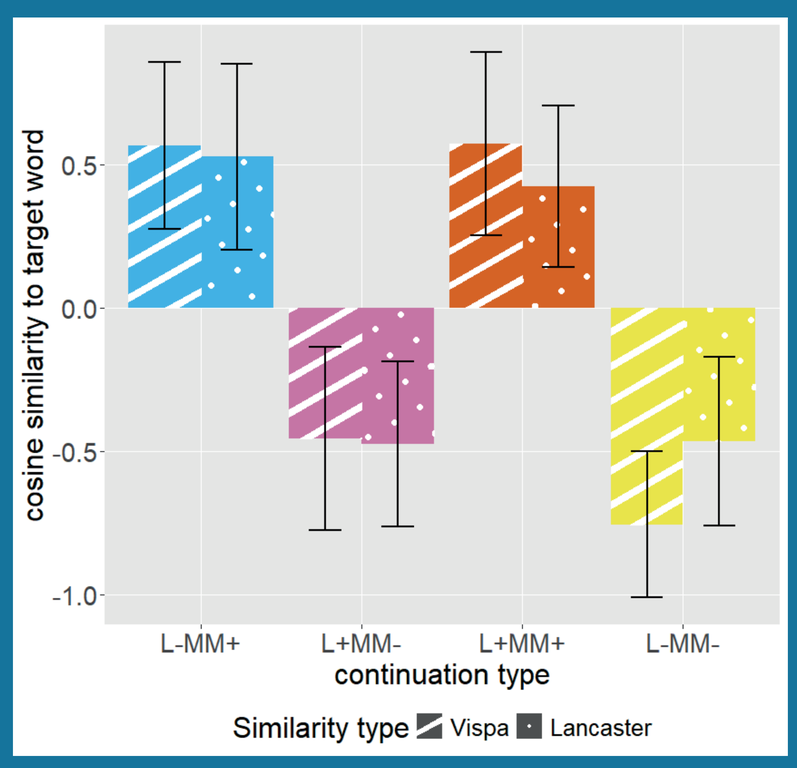
Continuation (conc & known)

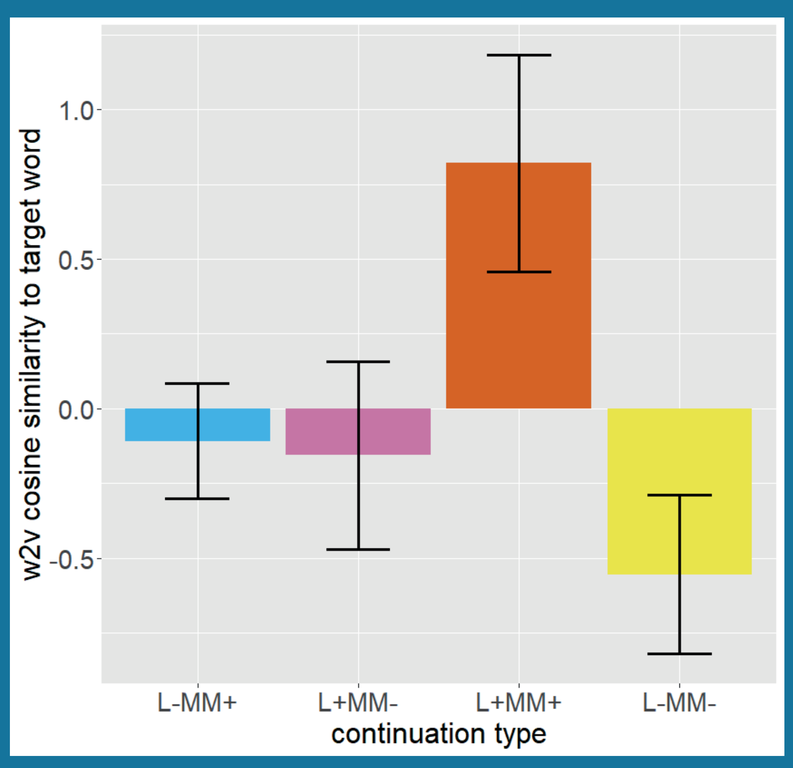
WATCH
COMPASS
WIFE
PHONE
DOG

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

Data-driven design







L = Cloze low entropy, ≥ 2 comp

MM = ViSpa & Lancaster

Extra = Google w2v

Norming studies

(1) Plausibility

The impatient man kept looking at his watch

1 2 3 4 5 6 7

(2) Visual similarity

Is a **watch** more visually similar to a **compass** or a **dog**?

compass — dog

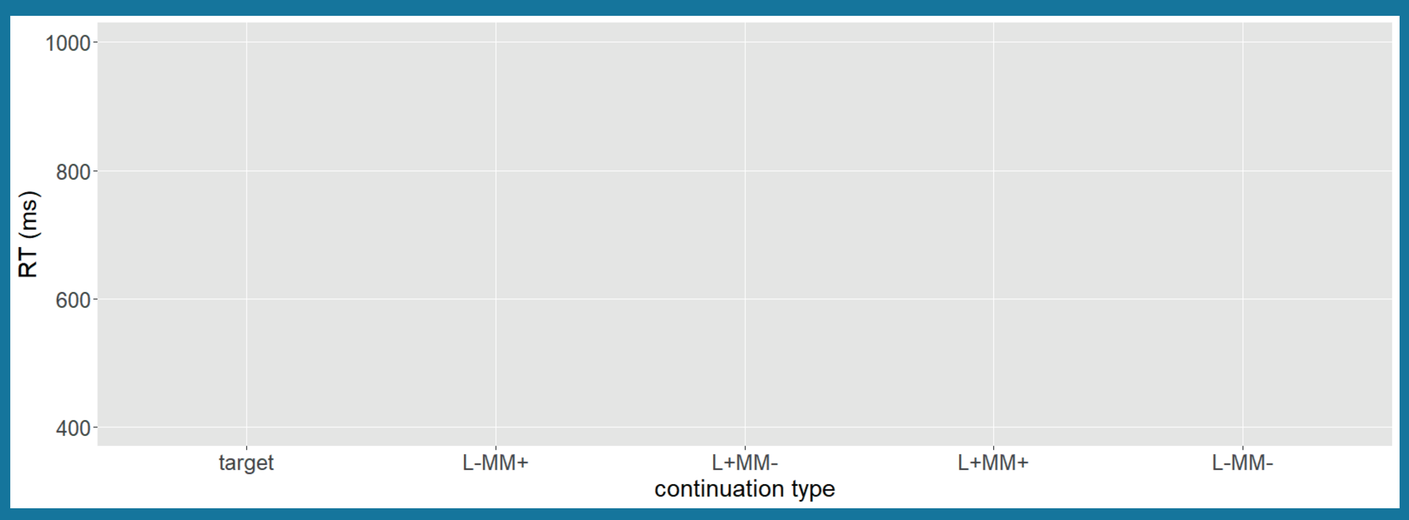
(3) Semantic relatedness

Which word is more likely to appear in the same context as **watch**?

dog — phone

To be continued…

- Self-paced reading experiment
- Do multimodal representations have an effect on reading times?
- Use unimodal and multimodal language models in the analysis



References

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms. *Behav. Res. Methods*, *52*, 1271-1291.

Günther, F., Marelli, M., Tureski, S., & Petilli, M. A. (2023). ViSpa (Vision Spaces): A computer-vision-based representation system. *Psych. Rev.*, *130*(4), 896.

Peelle, J. E., Miller, R. L., Rogers, C. S., Spehar, B., Sommers, M. S., & Van Engen, K. J. (2020). Completion norms for 3085 English sentence contexts. *Behav. Res. Methods*, *52*, 1795-1799.

Design assistance by Roxana Kenjeeva