Neural dynamics of multimodal semantic representations

Laura Anna Ciaccio[1], Silvia Angela Mansi[1], Marco A Petilli[2], Marco Marelli[2,3], Luca Rinaldi[1,4]

[1] *Department of Brain and Behavioral Sciences, University of Pavia, Italy*
[2] *Department of Psychology, University of Milano-Bicocca, Italy*
[3] *NeuroMI, Milan Center for Neuroscience, Italy*
[4] *Cognitive Psychology Unit, IRCCS Mondino Foundation, Pavia, Italy*

Background: Recent studies have highlighted the role of both linguistic and sensory information, particularly visual information, in shaping semantic knowledge [1]. The integration of these different traces has been shown to vary dynamically depending on task constraints and stimulus modality, with this evidence being largely confined to the behavioral level [2]. The present study aims to elucidate the dynamic contribution of linguistic and visual experience to semantic representations by focusing on the time-course of neural activations, as reflected in EEG responses to word stimuli. Key questions include: i) does visual information contribute to semantic memory over and above linguistic experience? ii) do these effects have similar timing? iii) does their timing of activation flexibly vary depending on the task?

Method: The study involves analyses of EEG datasets spanning different word processing tasks: a lexical decision, a semantic decision, and a memorization task with a retrieval phase. The lexical decision and the semantic decision data were obtained from openly available EEG datasets [3, 4]. Data collection for the memorization task is close to being completed (N = 37, target N = 45). Notably, these three tasks vary in the extent to which they require conscious activation of semantic knowledge.

We leveraged computational models to obtain measures of linguistic and visual information related to each of the word stimuli. Respectively, we used distributional semantics models [5] to operationalize linguistic input, and the ViSpa vector representations [6], obtained from a deep convolutional neural network, to capture concepts'prototypical visual features. These computational measures are then combined with the EEG data at subsequent 20ms time-bins by means of Representational Similarity Analysis (RSA). These analyses allow us to assess whether and at what latencies similarity patterns of brain activation to words are aligned with, respectively, the words'linguistic similarity and the visual similarity of the corresponding concepts.

Results: Results show different activation dynamics of linguistic and visual input across the lexical decision and semantic decision datasets. For the lexical decision task, we found significant effects of linguistic similarity at early latencies (~160ms), with (weaker) effects of visual information only emerging at later latencies (i.e., after 460ms). The semantic decision task, instead, elicited early effects of visual information, starting from about 140ms with a positive peak at about 200ms, and weaker linguistic effects. For the (ongoing) memorization task, we expect the largest interplay of both types of experience, consistently with previous behavioral findings [2].

Discussion: Our results provide neural evidence for the multimodality of semantic representations. More specifically, we find that neural responses to words reflect distinct timing of activation of different experiential traces, such as linguistic and visual experience. Critically, we also show that both the extent to which such experiential traces are activated and their timing vary dynamically depending on task demands, ultimately providing evidence for a flexible semantic system.

References

[1] Ciaccio, L., & Rinaldi, L. (2025). Refining the multimodality of semantic representations. *Trends in Cognitive Sciences*, *29*(12), 1073–1074. https://doi.org/10.1016/j.tics.2025.10.014

[2] Petilli, M., Rodio, F., Gatti, D., Marelli, M., & Rinaldi, L. (2025). Multimodal prior knowledge determines false memory formation. *Journal of Experimental Psychology: General*, *155*(2), 327–343. https://doi.org/10.1037/xge0001852

[3] Dufau, S., Grainger, J., Midgley, K., & Holcomb, P. (2015). A Thousand Words Are Worth a Picture: Snapshots of Printed-Word Processing in an Event-Related Potential Megastudy. *Psychological Science*, *26*(12), 1887–1897. https://doi.org/10.1177/0956797615603934

[4] Kappenman, E., Farrens, J., Zhang, W., Stewart, A., & Luck, S. (2021). ERP CORE: An open resource for human eventrelated potential research. *NeuroImage*, *225*, 117465. https://doi.org/10.1016/j.neuroimage.2020.117465

[5] Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning Word Vectors for 157 Languages. https://doi.org/10.48550/arXiv.1802.06893

[6] Günther, F., Marelli, M., Tureski, S., & Petilli, M. (2023). ViSpa (Vision Spaces): A computer-vision-based representation system for individual images and concept prototypes, with large-scale evaluation. *Psychological Review*, *130*(4), 896–934. https://doi.org/10.1037/rev0000392