
The world in words: Training computational models on linguistic data, and what it tells us about grounded cognition

Marco Marelli¹

¹ University of Milano-Bicocca, Italy

The relation between language, as a symbolic, abstract system realized within a speaker, and the external world, as a source of situated, embodied experience, has been a long matter of debate. Recent empirical work suggests that the inter-relation between the two might be deeper than previously thought, with language acting as a compressed version of the experienced world, and by this fulfilling the purpose of efficient communication. In support of this view, I will discuss a series of case studies employing language-trained computational models.

First, I will show how a model trained on simple lexical co-occurrences can reproduce perceptual distributions to the point of predicting visual processing in non-human primates, i.e. agents that are devoid of language. Indeed, activations in the ventral visual stream of macaques watching images are aligned with computational predictions about those images' labels. Second, I will describe how even pseudowords can be associated with sensory intuitions in human speakers, and evoke sensory-related effects in lexical decision. Such a mechanism builds on subtle systematic associations between sub-lexical patterns and semantics, as emerging from text-trained models once their lexicalist assumptions are relaxed. Third, I will show how, differently from their more dated counterparts, frontier large language models largely align with human sensorimotor intuitions. Such an alignment is observed even for systems that are not provided multimodal information during training.

In conclusions, statistical processing of linguistic data can reproduce perceptual intuitions and processing at a level that challenges traditional characterizations of language as an abstract, disembodied system of symbols. Language and grounded experience emerge as deeply intertwined, establishing the former as a main vehicle of experience-by-proxy and speaking for an integrated, non-modular view of cognition.