Making sense of meaning: enriching vector space models of semantic memory with visual information

Marco Marelli

Department of Psychology, University of Milano-Bicocca

Since the late 90s, vector space models have provided a convenient way to characterize semantic effects in language processing. These models typically build on linguistic information, inducing meaning representations through the processing of large collections of text, which arguably approximate our linguistic experience. This approach, however, has raised criticisms in the cognitive science community. In fact, according to different proposals in embodied cognition, concepts are grounded on sensorimotor experiences; a language-focused modelling approach would hence inevitably fall short in capturing human semantic memory. However, the centrality of linguistic data in vector space models is not a theoretical assumption, but rather a practical convenience. The same architectures can be readily applied to different information sources, pending data availability. In this talk I will discuss how these models can be enriched through visual information by relying on recent developments from computer vision. The resulting model not only captures human intuitions concerning images and words, but also offers new ways to investigate long standing debates, such as the impact of grounding information in semantic priming and the effect of semantic transparency in complex word processing. Moreover, when combining visual and linguistic data in a single architecture, the model provides novel explanations about the grounding of abstract concepts, showing how perceptual grounding can be reached via linguistic mediation. In summary, this approach paves the way to bring multimodality in data-driven, large-scale models of semantic memory, offering a novel, more cognitively plausible characterization of semantic effects in word processing.