Compositionality in (high-dimensional) space

Marco Baroni
Università di Trento

Abstract

Formal semanticists have developed sophisticated compositional theories of sentential meaning, paying a lot of attention to those grammatical words (determiners, logical connectives, etc.) that constitute the functional scaffolding of sentences. Corpus-based computational linguists, on the other hand, have developed powerful distributional methods to induce representations of the meaning of content-rich words (nouns, verbs, etc.), typically discarding the functional scaffolding as "stop words". Since we do not communicate by logical formulas, nor, Tarzan-style, by flat concatenation of content words, a satisfactory model of the semantics of natural language should strike a balance between the two approaches. In this talk, I will present some recent proposals that try to get the best of both worlds by adapting the classic view of compositionality as function application developed by formal semanticists to distributional models of meaning. I will present preliminary evidence of the effectiveness of these methods in scaling up to the phrasal and sentential domains, and discuss to what extent the representations of phrases and sentences we get out of compositional distributional semantics are related to what formal semanticists are trying to capture.