

Evaluating the impact of relations in compound processing: a simulation study in distributional semantics

Friday, June 23, 2017 2:10 PM (20 minutes)

In most languages, words can be combined to create novel compounds that are readily understandable by speakers. Crucially, the compound meaning is not only determined by the two words, but also by the (un-expressed) relation that links them together: we have a clear intuition that *snow building* means *a building MADE OF snow*, even if we have never heard it before. In the present work, we propose a new data-driven model, CAOSS (Compounding as Abstract Operation in Semantic Space), to capture this process.

In CAOSS, word meanings are represented as vectors encoding lexical co-occurrences from a text corpus. A compositional procedure is applied to these vectors: given two constituent words u and v , their composed representation can be computed as $c=Mu+Hv$, where M and H are weight matrices estimated from corpus examples. The matrices are trained using least squares regression, having the vectors of the constituents as independent words (*car* and *wash*, *rail* and *way*) as inputs and the vectors of example compounds (*carwash*, *railway*) as outputs, so that the similarity between $Mu+Hv$ and c is maximized. Once the two weight matrices are estimated, they can be applied to any word pair in order to obtain meaning representations for untrained word combinations (i.e., productive usage of compounding).

We tested our models against behavioral results from the conceptual combination literature, and in particular the effects of relational priming and relational dominance in the processing of novel compounds. The impact of relational information, as well as its specific association with the initial constituent, are correctly predicted by the CAOSS representations.

The model simulations suggest that relational information can be learned from language experience and then applied to the processing of new word combinations. CAOSS representations are flexible and nuanced enough to emulate this procedure.

Primary author: MARELLI, Marco (University of Milano-Bicocca)

Co-authors: GAGNÉ, Christina (University of Alberta); SPALDING, Thomas (University of Alberta)

Presenter: MARELLI, Marco (University of Milano-Bicocca)

Session Classification: Contributed papers 5

Track Classification: Freely Contributed Paper