

A NEW SEMANTICS: MERGING PROPOSITIONAL AND DISTRIBUTIONAL INFORMATION

Eduard Hovy (hovy@isi.edu)

Information Sciences Institute, University of Southern California
Marina del Rey, Southern California

Key words: semantic content, NLP, distributional semantics, distribution information.

Despite hundreds of years of study on semantics, theories and representations of semantic *content* — the actual meaning of the symbols used in semantic propositions — remain impoverished. The traditional extensional and intensional models of semantics are difficult to actually flesh out in practice, and no large-scale models of this kind exist. Recently, researchers in Natural Language Processing (NLP) have increasingly treated *topic signature word distributions* (also called ‘context vectors’, ‘topic models’, ‘language models’, etc.) as a de facto placeholder for semantics at various levels of granularity. This talk argues for a new kind of semantics that combines traditional symbolic logic-based proposition-style semantics (of the kind used in older NLP) with (computation-based) statistical word distribution information (what is being called Distributional Semantics in modern NLP). The core resource is a single lexico-semantic ‘lexicon’ that can be used for a variety of tasks. I show how to define such a lexicon, how to build and format it using tensors, and how to use it for various tasks. I discuss some of the recent work on composing vectors and tensors in attempts to produce statistically-based compositional semantics. Combining the two views of semantics opens many fascinating questions that beg study, including the operation of logical operators such as negation and modalities over word(sense) distributions, the nature of ontological facets required to define concepts, and the action of compositionality over statistical concepts.