Jacob EISENSTEIN  
Georgia Tech, Atlanta (Etats-Unis)

Distributed Semantics for Automatically Classifying Discourse Relations

Abstract: The coherence relations that are at the center of many models of discourse structure are fundamentally semantic in nature --- an observation that goes back to Hobbs (1978) if not earlier. While theoretical work has elucidated connections between formal semantics and discourse, at present this is of little practical use in automatic discourse parsing, due to the intractability of open-domain formal semantic analysis. Distributed compositional semantics offers an appealing alternative: the meaning of discourse arguments is captured in dense numerical vectors, which are constructed incrementally from smaller linguistic units; the compositional operations themselves can be learned so as to optimize performance on discourse parsing. But the key question is whether these vector-based representations are sufficiently expressive to capture the semantics behind discourse relations. This talk describes three projects in which distributed semantic representations yield significant improvements in discourse relation detection: rhetorical structure theory parsing, supervised PDTB relation classification, and adaptation from explicit to implicit relations in the PDTB. I also propose "structured distributed semantics" as a (nearly totally unexplored) middle ground between the expressiveness of formal semantics and the tractability of distributed representations.

Bio: http://www.cc.gatech.edu/~jeisenst/bio.html