The Relevance of a Cognitive Model of the Mental Lexicon to Automatic Word Sense Disambiguation

Martha Palmer and Susan Brown
University of Colorado at Boulder
Department of Linguistics
Hellems 290, 295 UCB
Boulder, CO 80309-0295, USA

Supervised word sense disambiguation requires training corpora that have been tagged with word senses, and these word senses typically come from a pre-existing sense inventory. Space limitations imposed by dictionary publishers have biased the field towards lists of discrete senses for an individual lexeme. Although some dictionaries use hierarchical entries to emphasize relations between senses, many do not. WordNet, which has been the default choice of NLP researchers for sense tagging because of its broad coverage and easy accibility, does not have hierarchical entries. Could the relations between senses that are captured by a hierarchy be useful to NLP systems? Concerns have also been raised about whether or not WordNet's word senses are unnecessarily fine-grained. WSD systems are obviously more successful in distinguishing coarse-grained senses than fine-grained ones (Navigli, 2006), but important information could be lost if fine-grained distinctions are ignored. Recent psycholinguistic evidence seems to indicate that closely related word senses may be represented in the mental lexicon much like a single sense, whereas distantly related senses may be represented more like discrete entities (Brown, 2008). These results suggest that, for the purposes of WSD, closely related word senses can be clustered together into a more general sense with little meaning loss. This talk will describe this psycholinguistic research and its current implications for automatic word sense disambiguation, as well as plans for future research and its possible impact.