

Enriching and using large-scale lexico-semantic resources

Eneko Aguirre

University of the Basque Country, Spain

Abstract:

This talk presents a research line for the semantic processing of natural language. Current limitations of semantic interpretation systems stem from the pervasive ambiguity in natural language. Our working hypothesis is that a complex mixture of resources (ranging from hand-made knowledge bases, lexicons and hand-tagged corpora to raw corpora) and machine learning strategies built by different researchers need to be integrated in order to advance the state-of-the-art. An architecture using the EuroWordNet design will be presented, with some examples of integrating both hand-made resources and automatically learned relations. Finally, some promising applications on CLIR are presented, as well as the ongoing SemEval 07 / CLEF 2008 exercise on evaluating the contribution of WSD to CLIR and QA.