

Towards robustness in natural language processing

by Jordi Atserias

Yahoo! Research Barcelona & UPF

Abstract

The different architectures proposed for Natural Language Understanding (NLU) can be classified based on two main dimensions, namely, the level of integration of their processes and the level of integration of their data. We present in this talk an integrated architecture for robust Natural Language Understanding, exploiting constraint-based optimization techniques. Constraints allow us to integrate both processes and knowledge in the same framework. On the one hand, many forms of ambiguity can be represented in a compact and elegant manner by means of constraints. On the other hand, many natural language processes could also be represented as constraints.

The architecture proposed has been applied to two different tasks, Word Sense Disambiguation (WSD) and Semantic Role Labeling (SRL). Pursuing this goal, we have integrated several resources in a multilingual knowledge base, named Multilingual Central Repository (MCR). MCR has been built around WordNet and integrates different resources, ontologies (SUMO, Top Concept Ontology), thematic classifications (Domains), local wordnets of different languages. We will describe how each task (WSD and SRL) can be formalized in our framework, and how the relevant knowledge and processes are integrated in our framework using constraints.