Improving HPSG Parse Disambiguation by Enriching Syntactical Context Information

Abstract

With the advance of science and technology, Nature Language Processing plays a more important role nowadays. Natural languages are abundant of ambiguities. While ambiguity facilitates efficient communication between human speakers, it causes processing difficulties in language technology. Many applications that may benefit from deep linguistic analysis (such as Question Answering, Machine Translation, etc.) require a single parse tree as their input. The task of parse disambiguation is becoming increasingly important.

Broad coverage grammar such as HPSG or LFG has been introduced and welcomed in recent year. This kind of grammar is developed to offer a more detailed analysis for different languages. However, multiple analyses will inevitably rise in such detailed analysis, both due to the genuine ambiguity of the language and the spurious ambiguity produced by the processing. The number of correct structures that are licensed by the broad coverage grammar is often vast. In such kind of environment, how to identify the intended analysis from a handful or usually a ton of reading becomes a critical issue.

In this work, we discuss some existent methods and consider new approaches which intent to improve the parse disambiguation accuracy on HPSG. We employed a statistical model which fine-tunes the information from the output of HPSG framework in order to find valuable features on the disambiguation task. In particular, we focus on the combination of syntactic categories and the rule schemata of HPSG which were used to license the signs. The features in the combination result are more robust and active than they were when presented separately.