Recent Trends in NLP Research

Changning HUANG
Microsoft Research, China
cnhuang@microsoft.com

Abstract: Parsing is a well-known key issue in Natural Language Processing (NLP). Corpus investigation shows the following two facts: (1) CFG rules based on mono-labels, such as POS/PT, are not sufficient for parsing, and (2) CFG rules also have Zipf's distribution in corpus, i.e. the total number of rules aren't limited as expectation. The recent trends of NLP research are heavily influenced by those facts. The speaker will introduce his observation on those trends, such as multiple features and unification-based grammars, lexicalism in grammar researches, statistical language modeling and corpus-based approaches. Although the first two remedies still forward along the traditional linguistics approaches, but they explore linguistic knowledge with much more fine grain, especially lexical knowledge, than before. The last remedy falls into the so-called experientialism approaches. It is based on observable facts rather than on linguist's intuition. The mission of those approaches is to acquire linguistic knowledge with more fine grain from very large corpora automatically or semi-automatically. Because this is still the bottle-neck of NLP technology developments.