

# Word Dependency and Word Co-occurrence as the Base for Language Analysis and Understanding

Yuji Matsumoto (Nara Institute of Science and Technology)

Traditional Natural Language analysis used to do syntactic parsing of sentences based on phrase structure grammars. Computationally and linguistically motivated grammar formalisms such as Lexical Functional Grammar, Generalized Phrase Structure Grammar, Head-driven Phrase Structure Grammar, Tree Adjoining Grammar, etc., basically use phrase structure trees for the syntactic representation of sentences. Having very clear background and some of them having a link between syntactic and semantic structures, they suffer from limitations in covering broad usages of real world language expressions. Recently, word dependency parsing attracts lots of attention because of its simplicity and robustness. The entrance of semantic analysis of sentences is the identification the predicate (exemplified as a verb, an adjective or an eventual noun) and its arguments in a sentence. Word dependency structure and some grammatical constraints work effectively for finding out predicate-argument structures coded in sentences. While semantic relationship between words and/or concepts is a key for achieving accurate language analysis, representation of semantic information (or general knowledge) has been one of the hard problems in natural language processing and artificial intelligence. Human crafted thesauri and ontologies exist, but in many cases their effect in language analysis is not perspicuous. Co-occurrence based similarity measure is now a trend in representing semantic relatedness of words and expressions. This talk will introduce recent movements of word-based analysis of natural language processing, and will introduce our linguistic annotation tools and environment developed as a part of government-funded project for constructing Balanced Corpus of Contemporary Written Japanese (BCCWJ).