

Prof. Dr. Junichi Tsuji

Faculty of Science / Graduate School of Information Science and Technology University of Tokyo, Japan

Computational Approach to Natural Language Processing

One of the miracles of human mind is how efficient it is in processing a complex symbolic system of natural language. Observation of human use of language shows that natural language is a complex system of syntactic, semantic and pragmatic constraints. Natural language is very ambiguous, compared with other artificial symbolic systems such as programming languages, logical formula, etc. However, this inherent ambiguity of natural language does not affect the efficiency and effectiveness of human mind in processing it.



In a project supported by JSPS, our group in the University of Tokyo has been involved in building a computational system that can treat language efficiently. While the main aim of the project is to develop an efficient computational model of language processing regardless whether it models human mind or not, the model we have developed seems to give us an insight how human mind or brain works. Our model is based on the following two major claims.

- According to theoretical linguistics, the system of natural language comprises constraints of various different kinds such as syntax, semantics and pragmatics that are inherently intertwined. This observation leads us to a complex system of constraints such as HPSG, LFG, LTAG, etc., the computer processing based on which is extremely slow. However, we contend that, while a processing system uses complex grammar description based on these formalisms, it needs not to handle all the constraints at once. Our model processes language in a series of different stages, while such a series of processing stages as a whole maintains the integrated nature of language that these grammar formalisms try to capture. This processing model results in a remarkably efficient computational model.
- While most of ambiguities cannot be resolved unless a processing model refer to pragmatic constraints like real world knowledge and the context in which a sentence is actually uttered, there are certain biases between preferred and less preferred interpretations. Such biases can be included in the early stages of language processing and can reduce ambiguities significantly. We combine such preferences in the form of statistics with the system of constraints. We believe that such statistical biases and the use of them in the early stages of language processing is one of the reasons why human mind is so efficient.