The Parsing Environment for Mandarin Syntax

I-Peng Lin, Shuan-Fan Huang, Hsin-Hsi Chen & Ka-Wai Chui

林一鵬　黃宣範　陳信希 徐嘉慧

National Taiwan University

國立台灣大學

Proceedings of ROCLING I(1988)
R.O.C. Computational Linguistics Workshops I pp 211-214

中華民國第一屆計算語言學研討會論文集 211-214 頁

The Parsing Environment for Mandarin Syntax

I-Peng Lin, Shuan-fan Huang, Hsin-Hsi Chen and Ka-Wai Chui

Department of Computer Science and Information Engineering

abstract

'Syntax' is the sentence-formation component in grammar, specifying how sentences are constructed out of phrases, and phrases out of words. It is necessary to set up a parsing environment to implement the Mandarin syntax in computer. This paper thus attempts to introduce this parsing enviroment .

There are three types of sentences in total: declarative sentences, imperative sentences and questions. The declaratives can further be divided into simple, complex and compound constructions. Simple sentences are defined as those consisting of a single predicate. When there are more than one predicate within a sentence, it is called complex sentence. We classify 3 types of complex constructions. The first of which results from subcategorization. The second and the third types are the verb-reduplicated and the serial-verb constructions. Compound sentences presuppose the presence of conjunctions to link up two or more phrases or clauses. As imperative constructions are concerned, they have their basis on their declarative counterparts, yet with a number of restrictions, e.g. the optional subject has to be the second person, the construction has to be unmarked aspectually. Finally, the different types of questions in Mandarin include question-word questions, yes/no questions, choice questions, A-not-A questions, tag-questions, embedded questions. The linguistic device to generate all of them is the context-free X-bar Phrase Structure Rules.

The parsing environment for Mandarin syntax is indeed based on the data analyses of the various types of sentences discussed above. The Prolog-based bottom-up parser can even tackle the problems of movement transformation by means of three principles in the Government-Binding theory, namely the Empty Category Principle, the C-Command Principle and the Subjacency Principle. A sequence of translation rules is given to add these linguistic principles to the general grammar rules, the leftward movement grammar rules, and the rightward movement grammar rules, respectively. The empty constituent problem is solved to allow the trace to be the first element in the grammar rule body. A special data structure for extraposition list is proposed to transfer the movement information from the bottom to the top. Based upon this structure, the fastest merge algorithm is designed. Those

unnecessary *merge* predicates can be eliminated with the help of the transitive relation. Thus, the new design not only extends the original bottom-up parsing system with the movement facility, but it also preserves the parsing efficiency.