Journal of Natural Language Processing Vol. 2 No. 2

Integration of Morphological and Syntactic Analysis based on LR Parsing Algorithm

Tanaka Hozumi,† Tokunaga Takenobu† and Aizawa Michio††

Morphological analysis of Japanese is very different from that of English, because no spaces are placed between words. This is also the case in many Asian languages such as Korean, Chinese, Thai and so forth. In the Indo-European family, some languages such as German have the same phenomena in forming complex noun phrases. Processing such languages requires the identification of the word boundaries in the first place. This process is often called *segmentation*. Segmentation is a very important process, since the wrong segmentation causes fatal errors in the later stages such as syntactic, semantic and contextual analysis. However, correct segmentation is not always possible only with morphological information. Syntactic, semantic and contextual information are also necessary to resolve the ambiguities in segmentation. This paper proposes a method to integrate the morphological and syntactic analysis based on LR parsing algorithm. An LR table derived from grammar rules is modified on the basis of connectabilities between two adjacent words. The modified LR table reflects both the morphological and syntactic constraints. Using the LR table and the generalized LR parsing algorithm, efficient morphological and syntactic analysis is available.

KeyWords: *Generalized LR parsing, Morphological analysis, Syntactic analysis*

† Department of Computer Science, Tokyo Institute of Technology
†† Media Technology Laboratory, Canon Inc.