A Systemic-Functional Approach to Japanese Text Understanding

Noriko Ito, Toru Sugimoto, and Michio Sugeno

RIKEN Brain Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198 Japan
{itoh,sugimoto,msgn}@brain.riken.jp
http://www.brain.riken.jp/labs/lbis/

Abstract. We have implemented a Japanese text processing system, combining the existing parser and dictionary with the linguistic resources that we developed based on systemic functional linguistics. In this paper, we explain the text understanding algorithm of our system that utilizes the various linguistic resources in the Semiotic Base suggested by Halliday. First, we describe the structure of the SB and the linguistic resources stored in it. Then, we depict the text understanding algorithm using the SB. The process starts with morphological and dependency analyses by the non-SFL-based existing parser, followed by looking up the dictionary to enrich the input for SFL-based analysis. After mapping the pre-processing results onto systemic features, the path identification of selected features and unification based on O'Donnell are conducted with reference to the linguistic resource represented in the system networks. Consequently, we obtain graphological, lexicogrammatical, semantic and conceptual annotations of a given text.

1 Introduction

The purpose of this research is to implement a natural language processing system that follows the theoretical model of *systemic functional linguistics* (SFL). SFL aims at describing a language comprehensively and provides a unified way of modeling language use in context [1]. While SFL has been used as the basis for many natural language generation systems (e.g., [2]), less work has been done for natural language understanding systems (e.g., [3]).

Sugimoto [4] proposed the data structure of the *Semiotic Base* (SB), which stores SFL-based linguistic knowledge in computational form, and investigated how to incorporate the SB into a dialogue management model in order to enable an intelligent agent system to identify the current dialogue context and behave appropriately according to it. By elaborating their idea and combining the existing parser and dictionary with systemic resources, we implemented a Japanese text processing system that can conduct both understanding and generation of Japanese text.

In this paper, we explain the text understanding algorithm of our text processing system that utilizes the various linguistic resources in the SB. First, we describe the structure of the SB and the linguistic resources stored in it. Then, we depict the text understanding algorithm using the SB and the outputs of the process.