

Exploiting coreferential information in NPCMJ for L2 reading of Japanese texts

Kei Yoshimoto and Akiko Takahashi

Tohoku University and Miyagi University of Education

NINJAL Parsed Corpus for Modern Japanese (NPCMJ) is a treebank with annotations of both syntactic and semantic information for modern Japanese sentences. The advantage of NPCMJ over other Japanese corpora developed so far lies in representing the dependencies between all words and phrases in each sentence by utilizing predicate logic formulas automatically produced from syntactic information. Furthermore, omitted NPs with obligatory grammatical roles (subjects and objects), which appear frequently in Japanese texts, are annotated explicitly as zero pronouns. These two characteristics of NPCMJ make it possible to obtain correct information on case NPs directly related to verbs and adjectives even for zero pronouns, controlled subject and object in complex sentences, and sentences with unbounded dependency (e.g. sentences with relativization and topicalization).

It has been pointed out that extensive reading, a rapid reading task assigned to learners of a second language (L2) without requiring complete understanding of the vocabulary and sentence structures in the text, is effective in enhancing the learners' text comprehension ability. However, most real texts are inappropriate for this task, since they contain too difficult vocabulary and too long and complicated sentences for L2 learners to understand with ease. Unlike English, Japanese has yet few graded readers, rewritten texts with easy-to-understand vocabulary and sentence structures.

The main reason that prevents L2 learners from comprehending original texts is the difficulty in relating verbs and adjectives to their case NPs that are omitted (zero-pronominalized), controlled by the subject/object in the matrix clause, or in sentences with unbounded dependency. As explained above, NPCMJ allows us to automatically get information on case structures for verbs and adjectives in the text. By using this information, we can develop a system that displays case information of all verbs and adjectives in the text. This might be expectedly as effective as graded readers in enhancing L2 learners' reading ability with extensive reading.

We conduct an experiment to determine whether such assistance actually facilitates L2 learners' reading activities. We propose a hypothesis that a CALL reading material with visually displayed case structures is as effective as one with rewritten simple sentence structures and is more effective than an original text in enhancing L2 learners' comprehension. To prove this, we give intermediate learners of Japanese three kinds of reading materials, i.e., an original text, a text with displayed case information of verbs and adjectives, and a structurally simplified text. After the reading task, questions are given to the subjects on implicit case roles and the content of the text. We examine the hypothesis by analyzing the test result.

Example: Three kinds of text assigned to L2 learners

Original text:

はしけはひどく揺れた。踊子はやはり唇をきつと閉じたまま一方を見つめていた。私が縄梯子につかまろうとして振り返った時、さようならを言おうとしたが、それもよして、もう一ぺんうなずいて見せた。

(The lighter pitched violently. The dancer stared fixedly ahead, her lips pressed tight together. As I started up the rope ladder to the ship I looked back. She wanted to say good-by, but she only nodded again.)

Text with case structures.

はしけはひどく揺れた。踊子はやはり唇をきつと閉じたまま一方を[踊り子が]見つめていた。私が縄梯子につかまろうとして[私が]振り返った時、さようならを[踊り子が]言おうとしたが、それも[踊り子が]よして、もう一ぺん[踊り子が]うなずいて見せた。

Rewritten text:

はしけはひどく揺れた。踊子はやはり唇をきつと閉じたまま一方を見つめていた。私は縄梯子につかまろうとして振り返った。踊子は私にさようならを言おうとしたが、それもよして、もう一ぺんうなずいて見せた。