FOREWORD

This issue is dedicated to the field of corpus linguistics, a relatively young but well-established approach to the study of language, and particularly to researches that base on the Japanese BCCWJ corpus.

The Balanced Corpus of Contemporary Written Japanese (BCCWJ) – a one hundred million word corpus released by the National Institute of Japanese Language and Linguistics (NINJAL) in the year 2011 – has greatly advanced research in the field of Japanese language. Utilizing a corpus that is made-up of a multitude of actual language data enables researches to overcome the limitation of basing their analysis primarily on either intuition or scanty language data, making it possible to study languages in a more objective and comprehensive manner. In addition to this, analyzing data compiled by collecting learner essays or speech makes it possible to throw light on errors based on mother-tongue influence as well as acquisition patterns independent of mother-tongue influence. Furthermore, the observations and analysis based on actual language use can be used to create dictionaries and other authentic study materials.

Though the compilation of a Japanese language corpus falls far behind when compared to corpora in English, a fairly large number of Japanese corpora like the BCCWJ have been made available in recent years. Also, a corpus of ten billion words is currently being compiled at NINJAL making it possible to utilize a large-scale corpus of greater diversity in the future. Such corpora may be utilized in the following research field pertaining to Japanese language:

- Japanese language research
- Analyzing Japanese language syllabus
- Determining difficulty levels of words and passages
- Developing tools that aid text-comprehension and essay writing
- Developing tools that aid corpus-driven learning
- Automatic assessment and correction of learner errors

The thematic issue thus introduces four research studies that utilize the above mentioned corpus in the field of Japanese language education. The final research is a good example of a research of languages in contact and is also based on a corpus, though a non-electronic one.

Andrej Bekeš discusses spatio-temporal metalinguistic labels, NSM (natural semantic metalanguage), introduces a methodology based on the corpus-based approach that emphasizes function and meaning and the approach that emphasizes relation to

context, and presents his observations on metalinguistic terms for dictionaries of Japanese expressions.

Irena Srdanović introduces various tools that search and retrieve collocations and presents her analysis on collocations that utilizes Japanese corpora and corpus tools by giving actual examples of collocations consisting of adjective and noun combinations.

Kikuko Nishina and her colleagues present an overview of the three systems developed through the Hinoki Project at the Tokyo Institute of Technology, namely Natsume (collocation search system), Natane (a learner corpus search system that supports Natsume) and Nutmeg (an automatic correction, composition support system), while emphasizing the need for learner corpus.

Kristina Hmeljak Sangawa introduces exercises developed by utilizing Japanese monolingual corpora and Japanese-Slovenian parallel corpus in order to deepen the understanding of vocabulary meaning, its structure and style.

Anrei A. Avram attempts to make an overall description of the phonological, morphological, syntactic features, and vocabulary in 19th century Yokohama Pidgin Japanese by using written corpora.

Yuriko Sunakawa