

Preface

Around the world spoken corpora have been built up for two purposes: to document endangered languages for preservation, and to document dominant languages for different applications. To bring together scholars and students to inspire nationwide and worldwide action to establish spoken corpora for the languages spoken by various ethnic groups in Taiwan for research and education, the 7th Wenshan International Symposium, held by the Department of English at National Chengchi University in Taipei on May 19, 2008, set its theme as *Spoken Corpus: Language Documentation and Applications*. The issues addressed include the establishment of spoken corpus, the use of spoken corpus in documenting and preserving languages, and applications of spoken corpus with regard to economic, social, cultural, community and humanitarian considerations.

This special issue contains revised papers presented at the 7th Wenshan International Symposium. The first article by **Shu-Chuan Tseng**, "Spoken Corpora and Analysis of Natural Speech" gives an overview of research issues investigated utilizing the spoken data. Based on the spoken corpora of Taiwan Mandarin created at Academia Sinica, this paper also examines the correlation between discourse items, prosodic features, and disfluency. The second article by **John Newman**, "Spoken Corpora: Rationale and Application" suggests reasons and introduces an array of possibilities for the construction of spoken corpora. The third article by **David Nathan**, "Minding Our Words: Audio Responsibilities in Endangered Languages Documentation and Archiving" suggests four key audio-related issues: audio quality, its accompanying symbolic data, the usage of data for practical purposes such as language revitalization, and the need for enhanced sensitivities and protocol in audio access and distribution. While surveying these issues, this paper also describes the funding, teaching, archiving and publishing activities of the Endangered Languages Project at SOAS. The fourth article by **Li-May Sung, Lily I-wen Su, Fuhui Hsieh and Zhemin Lin**, "Developing Online Corpus of Formosan Languages" introduces the elements of the construction of the NTU Corpus of Formosan Languages, its procedures, transcription conventions, database normalization, in-house system and three special features. The fifth article by **Kawai Chui and Hwei-ling Lai**, "The NCCU Corpus of Spoken Chinese: Mandarin, Hakka, and Southern Min" introduces the project of language documentation of Mandarin, Hakka, and Southern Min. Sharing a common scheme for collecting spoken data, the NCCU spoken corpus provides an open online access, and collects and archives spoken forms of various types under a simple yet user-friendly infrastructure.

Each of the papers was reviewed anonymously by at least two experts, to whom I would like to extend my greatest gratitude for their helpful evaluations. Incidentally, to ensure the double-blind reviews, the paper co-authored by Kawai Chui and myself was handled entirely by the regular editor One-Soon Her. Special thanks also extend to the editors of *Taiwan Journal of Linguistics*, One-Soon Her and Kawai Chui, for their support to the publication of this special issue.

Hwei-ling Lai
Department of English
National Chengchi University
December 22, 2008

Copyright of Taiwan Journal of Linguistics is the property of Graduate Institute of Linguistics and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.