

PREFACE

SPECIAL ISSUE: SELECTED PAPERS FROM THE 27TH PACIFIC ASIA CONFERENCE ON LANGUAGE, INFORMATION, AND COMPUTATION (PACLIC), NOVEMBER 22-24, 2013

Guest Editors

Shih-Ping Wang and Siaw-Fong Chung

Modern corpus linguistics started from research mainly on English (e.g., Aijmer and Altenberg 1991); nevertheless, corpus linguistics has been applied in various language studies and various linguistic fields during the past two decades. Although the use of large amounts of linguistic data is the new trend in linguistics research, in some linguistic fields, a small amount of data is still of great significance.

For example, the first paper of this volume, by **Kawai Chui, Hsiang-lin Yeh, Wen-Chun Lan, and Yu-Han Cheng**, features the analysis of morphosyntactic and concatenation preference of the verbal and adverbial expressions of PATH in Mandarin, Russian, and Spanish by studying the expressions of PATH in all these three languages. One group of learners were native speakers of the three languages, and the other group were learners of Russian and Spanish whose first language is Mandarin. This research used data from the NCCU Spoken Corpus of Mandarin, the NCCU Learner Corpus of Russian, and the NCCU Learner Corpus of Spanish. These small corpora were collected from a story-retelling of a cartoon episode that the participants had watched. The results showed that the learners' first language only affected the elementary stage of the language acquisition, but this did not hold true in

all languages. The variation across languages provides a good reference point for cross-linguistic transfer in language acquisition.

The next paper, by **Wen-hui Sah**, entitled “The Development of Coherence in Narratives: Causal Relations”, is also based on narration of a story (*Frog, Where are You?*). Thirty five-year-olds, 30 nine-year-olds, and 30 adult participants were involved in the narration of this story. In their narration, the causal network was analyzed by detection of causal statements, both overtly- and non-overtly-marked ones, in the narrative discourse. In this paper, causal statements, causal-chain events, and causal connections, all of which require careful scrutiny in the process of analysis, were analyzed separately. The results showed significant differences between adults and children, but the difference between the five- and nine-year-olds was only partial. The collected narration is a compilation of verbal description by both the children and the adults.

The collection of spoken data is a valuable resource in corpus linguistics due to the difficulty in collection and transcription procedures, and again, although the size is small, it is of great value in language acquisition research. Apart from the spoken narration, this volume also contains a paper featuring corpus annotation and another paper depicting sound profiles of initiation in French and Mandarin repairs.

The third paper, by **Hengbin Yan and Jonathan Webster**, has shown the steps in constructing the first functional corpus based on the theoretical framework of Systemic Functional Linguistics. Unlike the commonly seen part-of-speech tagging, this corpus contains annotation of process types such as ‘material’, ‘mental’, and ‘behavioral’, while each of these process types contain nuclear participants such as ‘actor/goal’, ‘senser/phenomenon’, ‘behavior/(target)’, etc. The annotation was made using a web-based collaborative tagger developed by the authors’ team, which allows the tagging of a theory-neutral annotation of syntactic, semantic, functional, and discourse information, and it enables the inclusion of multimodal data such as images, sounds, and videos. The paper also reports the challenges faced by the authors in the creation of this corpus. As mentioned above, this corpus is unlike an ordinary corpus but has multiple forms of information that can be used in the future for research on corpus linguistics.

The final paper, by **Helen Kai-yun Chen**, examines the acoustic profiles for instances of self-initiated repair, extracted from comparative speech corpora of face-to-face Mandarin and French interaction. The French data was extracted from a *Corpus of Interactional Data (CID)*, whereas the Mandarin data came from two resources – self-collected data by the author, and another corpus of conversational interaction following the format of the French CID. The author found that the speakers of the two languages differ in their repairs in spontaneous interaction. Like the first two papers in this volume, Chen's work is another piece of evidence that shows the importance of creating and analyzing speech corpora. Put simply, this volume presents four different but related papers. All provide insights into the impact of corpus linguistics on language research that deserves better attention. Last but not least, corpus-based approaches can be used for interdisciplinary cross-linguistic comparisons, and these approaches show a current trend in linguistic studies.

REFERENCE

Aijmer, K. and Altenberg, B. 1991. *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman.

Shih-ping Wang
Department of Applied Foreign Languages
National Taiwan University of Science and Technology
No. 43, Keelung Road Sec. 4
Taipei 106, Taiwan
spwang@mail.ntust.edu.tw

Siaw-Fong Chung
Department of English
National Chengchi University
No. 64, Zhi Nan Road Sec. 2
Taipei 116, Taiwan
sfchung@nccu.edu.tw