The Development and Evaluation of a Corpus-based Spanish Collocation Error Detection and Revision Suggestion Tool

Lu Hui-Chuan (NCKU, Taiwan) and Cheng An Chung (University of Toledo, USA)

The topic of collocation has drawn researchers' attention for the past thirty years in the lexical area of theoretical and applied linguistics (e.g., Firth, 1957; Nattinger, & DeCarrico, 1992; Sinclair, 1991; Wray, 2000). However, available assisting tools for learning Spanish collocation are much less than those for learning English in the field (e.g., Alonso Ramos, Nishikawa, & Vincze, 2010；Boisson et al. 2013; Bolshakov, & Miranda-Jiménez, 2004; Chen, 2011). This study expanded the functions of two previously constructed corpora and furthered the application of our research products in learning Spanish collocation. It included two primary tasks: (1) the development of a modified Spanish collocation assisted learning tool and (2) the evaluation of the relative effects of the developed assisted tool for learning collocation. The research question of the study was: Do the functions of the Spanish Collocation Error Detection and Revision Suggestion Tool assist learners to improve their learning?

In this study, a corpus-based assisted tool for learning Spanish collocation was developed and evaluated. First, based on the training data compiled in two created corpora (Taiwanese Learners' Corpus of Spanish and Parallel Corpus of Spanish, English and Chinese) and an extraction tool of Spanish collocation (Spanish Collocation Tool), the developed tool was designed with two main functions: error detection and revision suggestion of Spanish collocation for learning purposes. The database of Spanish collocations was generated with machine learning and through the procedure of data processing, collocation extraction, and manual modification. The programming languages used for the interface of users' query at runtime are HTML, Perl and PHP with Jquery kit and mysql. This developed assisted tool can detect inappropriate learner uses of Spanish collocation and provide suggestion lists for them to choose to revise the collocation errors.

In the second part of the study, a pretest, a video tutorial, a posttest and a questionnaire were conducted to evaluate the effectiveness of the assisted tool. The relative effectiveness of this tool was evaluated using an experiment to compare learning outcomes and feedback from two groups of Spanish learners (experimental and contrastive groups). Research results showed that our developed tool assisted learning effectively given the progress made from pretest to posttest. Furthermore, the results of satisfaction survey on interface and usefulness demonstrated that most of the participants positively confirmed the achievement of this tool for its effectiveness in assisting in the practice of Spanish collocations.

With respect to the limitation of this research, the training data for our developed tool from the two corpora was relatively small. Therefore, the identification and detection of errors were limited to collocations in a fixed and limited range. Also, the context and the current experiment were conducted within searchable combinations. Larger amount of training data from more varieties of contexts should be included for training in the future in order to strengthen the applicability of the tool.

This study shed light in pedagogical applications of created corpora and in the learning of Spanish collocation with a corpus-based approach in the setting of multilingual acquisition.

References

Alonso Ramos, M., Nishikawa, A., & Vincze, O. (2010). DiCE in the web: An online Spanish collocation dictionary. *Lexicography in the 21st century: New challenges, new applications; Proceedings of ELex 2009, Louvain-la-Neuve, 22-24 October 2009, 7*, 369.

Boisson, J., Kao T. -H., Wu, J. -C, Yen T. -H & Chang, J. S. (2013). Linggle: a web-scale linguistic search engine for words in context. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 139-144.

Bolshakov, I. A., & Miranda-Jiménez, S. (2004). A small system storing Spanish collocations *Computational Linguistics and Intelligent Text Processing* (pp. 248-252): Springer.

Chen, H. -J. (2011). Semi-automatic retrieval of collocation errors based on native and learner corpora. *Studies in English Language and Literature, 27*, 1-14.

Firth, J. R. (1957). Modes of meaning. In J. R. Firth (Ed.), *Papers in linguistics 1934 -1951* (pp. 190-215). Oxford, UK: Oxford University Press.

Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford, UK: Oxford University Press.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, UK: Oxford University.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 21*(4), 463-489.