

# **Key Lexical Bundle Analysis in an EFL Teacher Corpus: A Comparative Study on Non-native and Native Teacher Talk<sup>1</sup>**

Ye-Eun Kwon  
(Kunsan National University)

## **ABSTRACT**

Lexical bundles are one of important discourse building blocks that impact speaking proficiencies. This paper investigates similarities and differences in lexical bundle uses between non-native and native English teachers in Korea. 247,398 words were collected to build the corpus for the study. To investigate different usages of lexical bundles between two corpora, key lexical bundles were extracted for structural and functional analysis. Frequency and percentage of key lexical bundles between the two corpora were compared. The findings show that non-native English teachers repeatedly used a limited number of lexical bundles inflating the total number usages. Data also show non-native English teachers heavily rely on a few uses of stance expressions. Also, more than fifty percent of lexical bundles used are clause-based. Verb-phrase based lexical bundles, meanwhile, are the most used among the phrase-based lexical bundle categories. It suggests that EFL teacher corpus possesses similar features of lexical bundles in conversations rather than academic spoken or written corpus.

**Keywords:** key lexical bundles, learner corpus, teacher talk

## **I. Introduction**

Teacher talk can be considered as a register because it is used for a particular purpose in a particular setting (Johnstone, 2002). Investigating the features of conventional expressions in this particular register can have critical implications for teacher-student communication, which can lead to more effective language learning. In

---

<sup>1</sup>The paper was extracted and revised from the author's doctoral dissertation.

order to identify the similar and dissimilar features of conventional expressions used in EFL teacher talk, the study compiled a corpus of EFL teacher talk and investigated lexical bundles that were frequently occurring sequences of words.

In particular, the paper investigated key lexical bundles only, which occur significantly more frequently in a target corpus than in a comparing corpus (Scott, 1997; Scott & Tribble, 2006). Previous studies investigated different linguistic features across different disciplines, modes, and languages by extracting key lexical bundles (Baker, 2009; Kemppanen, 2004; Ku & Yang, 1999; Mauranen, 2006; McEnery, 2009; Seales, Charteris-Black and Ziebland, 2006). Doyle and Hong (2009) examined disciplinary differences in teacher talk using the SCoRe (Singapore Corpus of Research in Education) collected from Singapore primary and secondary school classes taught in English.

Biber (1999, 2004) pioneered the study of lexical bundles and identified the differences between spoken and written discourse, academic spoken discourse and conversation, and novice and professional writing. In more recent years, learner corpus-based studies have investigated lexical bundle use (De Cock, 2004; Tribble, 2011). However, only a few studies have examined the characteristics of the learner-spoken corpus (De Cock, 2004; Lee, 2009). There has been no attempt to investigate EFL teacher talk by compiling an EFL teacher corpus or to analyze teacher talk using the unit of the lexical bundle. Therefore, studying lexical bundles in non-native English teacher and native English teacher corpora is a new attempt to identify the phraseological patterns used by EFL teachers in Korea.

This study focused on the over- or under-use of certain types of lexical bundles by comparing the key lexical bundle lists of non-native English teacher corpus (NNT) to the native English teacher corpus (NT), and discussed the similar or dissimilar uses of lexical bundles with examples. Investigating authentic EFL teacher talk by noting the frequency of lexical bundles and comparing lexical bundle lists for NNT and NT is a first step toward practical suggestions for NNTs to improve their oral proficiency and for educators to improve teacher training programs and materials.

**II. Data and Methodology**

1. Data

The EFL teacher corpus consists of two sub-corpora which are non-native English teacher corpus (NNT), and Native English teacher corpus (NT). NNT and NT include lectures from four non-native teachers at three different universities and five native teachers at two different universities respectively. This EFL teacher corpus contains 247,398 words representing approximately 60 hours of classroom teaching.

**Table 1. Composition of EFL Teacher Corpus**

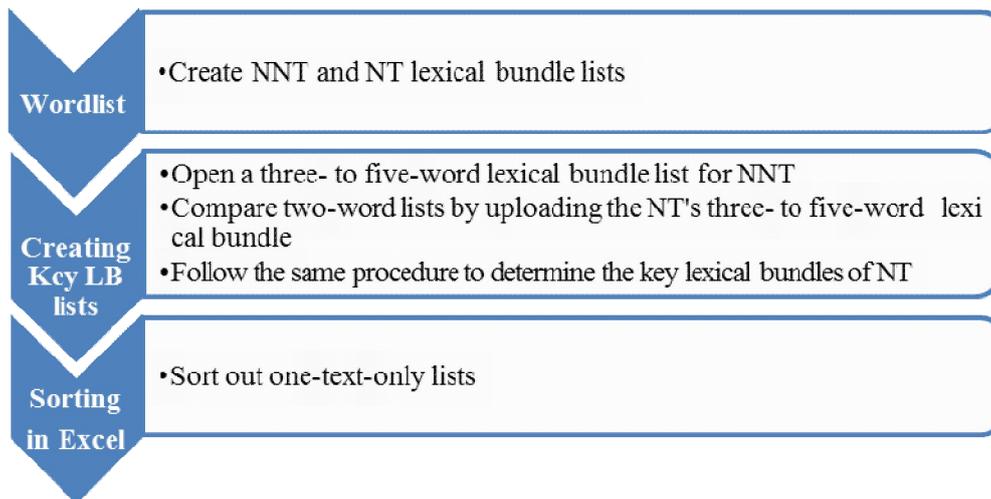
Corpus	Word	Time(min.)	Corpus	Word	Time(min.)
		1,940			1,786
NNT	123,122	(about 32 hrs)	NT	124,276	(about 30 hrs)

Lectures were audio-recorded and transcribed.

2. ANALYSIS

The study attempted to locate any atypical patterns of lexical bundles usage of NNT in comparison with NT.

**Figure 1. Summary of Key Lexical Bundle Identification Steps**



After key lexical bundle lists were extracted, key lexical bundles also needed a sorting step to lessen the influence of idiosyncratic usage and different topics. After these steps, functional and structural analyses were conducted based on the taxonomy used by Simpson and Ellis (2010, p. 498-502) and Cortes (2008, p. 49).

### 3. Research Questions

- 1) What are the similarities or differences in frequency of key lexical-bundles in NNT and NT?
- 2) What are the similarities or differences in the functional and structural features of key lexical bundles in NNT and NT?

## III. Results and Discussions

The present study extracted and sorted key lexical bundles to find out similar or different uses of lexical bundles between NNT and NT. Although the size of two sub corpora is similar, there are some lexical bundles which showed a large difference in frequency. Through the analysis of functional and structural analysis of key lexical bundles of NNT and NT, different lexical bundle usages between two corpora revealed more in details.

The results of functional analysis indicated that key lexical bundles as *stance expressions* were the most, which is more than 50 percent of key lexical bundles, and as *referential expressions* were the second most. The structural analysis results showed that the *clause-based* key lexical bundles were the most, occupying about 56% in NNT and 65% in NT. Among *phrase-based* key lexical bundles, *verb-phrase* based key lexical bundles accounted for the most number, which had similar percentage of lexical bundles in conversations of previous studies (Biber & Barbieri, 2007; Biber, et al. 2004).

The top-10 key lexical bundles were discussed as over-used, exclusively-used, and under-used lexical bundles. One of the most notable differences was found in over-used lexical bundles: *you have to* vs. *want you to do* in NNT and NT, and *let's move* and *move on to* in NNT. In terms of using the verb "move," NNTs preferably employed the verb "move" with "let's," however, NTs often used imperatives simply starting with the verb "move." Examples of exclusively-used lexical bundles were *you can see*, *listen to the*, and *can you please*.

Examples of under-used lexical bundles were *what did you, do you think* and *how would you*.

Different lexical bundle preferences can suggest specific lexical bundles what NNTs need to complement to speak more fluently in teaching EFL classes. The lexical bundles that NTs preferably employed are mostly those lexical bundles that NNTs know how to use since those lexical bundles were found in NNT but with lower frequency rate. Therefore, awareness raising on these lexical bundles could enhance NNTs' proficiency into one step above.

### REFERENCES

- Baker, P. (2009). The question is, how cruel is it? Keywords, fox hunting and the house of commons. In D. Archer (Ed.), *What's in a word-list? Investigating word frequency and key word extraction*.(pp.125-136). Farnham: Ashgate.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honor of Stig Johansson* (pp. 181-190). Amsterdam: Rodopi.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
- Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, 3(1), 43-57.
- DeCock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literature* 2, 225-246.
- Doyle, P., & Hong, H. (2009). *Compiling a multi modal corpus data base of educational discourse in Singapore schools* (Report No. CRPP Project CRP20/04JD/CRP7/05AL/CRP13/05AL). Singapore: Centre for Research in Pedagogy and Practice, National Institute of Education.
- Johnstone, B. (2002). *Discourse analysis*. Oxford: Blackwell.
- Kemppanen, H. (2004). Keywords and ideology in translated history texts: A corpus-based analysis. *Across Languages and Culture*, 5(1), 89-106.

- Ku, P., & Yang, A. (1999). An analysis of key words in tourism English. *Proceedings of the 1999 English for Specific Purposes Conference* (pp.232-241). Taiwan: Ming Chuan University.
- Lee, E.-J. (2009). A corpus-based study of the Korean EFL learners' use of formulaic sequences. *Foreign Languages Education*, 16(2), 321-340.
- Mauranen, A. (2006). Speaking the discipline. In K. Hyland & B. Marina Bondi (Eds.), *Academic discourse across disciplines* (pp.271-294). Bern: Peter Lang.
- McEnery, T. (2009). Keywords and moral panics: Mary Whitehouse and media censorship. In D. Archer (Ed.), *What's in a word-list? Investigating word frequency and key word extraction* (pp.93-124). Farnham: Ashgate.
- Scott, M. (1997). PC analysis of key words— and key key words. *System*, 25(1), 1-13.
- Scott, M., & Tribble, C. (2006). Keywords and corpus analysis in language education. *Textual patterns* (pp.3-10). Amsterdam: John Benjamins.
- Seal, C. Charteris-Black, J., & Ziebland, S. (2006). Gender, cancer experience and internet use: A comparative keyword analysis of interviews and online cancer support groups. *Social Science and Medicine*, 62(10), 257-259.
- Simpson-V., R., & Ellis N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.
- Tribble, C. (2011). Revisiting apprentice texts. Using lexical bundles to investigate expert and apprentice performances in academic writing. In F. Meuniers, S. De Cock, G. Gilquin & M. Paquot (Eds.), *A taste for corpora: In honour of Sylviane Granger* (pp. 85-108). Amsterdam: John Benjamins.

## BIODATA

The author received an MA in the professional education from States University of New York, Potsdam, U. S. A. and a Ph.D. in English education from EwhaWomans University, Korea. Her professional interests include corpus linguistics and teacher education. Currently, she is a teaching professor at Kunsan National University. Email: [yeeun78@daum.net](mailto:yeeun78@daum.net).