Corpus-Based Analysis of Lexical Bundles: Its Potential Applications in English Language Teaching

Atsushi Mizumoto
Kansai University

Lexical bundles, or multi-word sequences, play an essential role in creating discourse in spoken and written registers. Successful use of these bundles indicates the writer is in the discourse community, where experts use the "common language." In this paper, I will first show the importance of lexical bundles in shaping the discourse by introducing some of the corpus-based research findings. I will then elaborate on the possibilities of applying corpus-based analysis of lexical bundles to English language teaching.

Definition of lexical bundles

The umbrella term "formulaic sequence," covering a wide range of phraseology, is defined as "a string of words with a meaning of function different from that of its component parts" (Wray, 2012, p. 2200). For example, phrases such as *I know what you mean, on a regular basis, at the same time, take a break* are all considered formulaic sequences. It is reported that (a) formulaic sequences are ubiquitous in language use, (b) they constitute more than 50% of spoken and written discourse, (c) they may be stored and processed as single units, and (d) they are an important component of language learning and use (Schmitt, 2004).

As an all-encompassing term, formulaic sequences are called, often interchangeably, *chunks, multiword units, clusters, idioms, collocations, formulas, prefabricated routines,* and *conventionalized forms* (e.g., Schmitt, 2004). Lexical bundles can be classified as one of those above terms, originating from corpus linguistics (Biber, Johansson, Leech, Conrad, & Finegan, 1999). According to Biber et al. (1999), "Lexical bundles are recurrent expressions, regardless of their idiomaticity, and regardless of their structural status. That is, lexical bundles are simply sequences of word forms that commonly go together in natural discourse" (p. 990). Table 1 shows some of the

examples of lexical bundles. As can be seen, those lexical bundles have special meanings such as idioms; rather, they are just a sequence of words occurring more often than others. As such, lexical bundles can be extracted automatically using a concordancer such as AntConc (Anthony, 2014) and CasualConc (Imao, 2014). Because lexical bundles are simply a given length of repetitive strings of words, they are also known as n-grams, in which "n" stands for sequences of contiguous number ("n") of words (Granger, 2014). For example, *go to, for the,* and *report on* are 2-grams, i*n order to, the use of, and in the current* are 3-grams, *at the same time, the results of the,* and *in addition to the* are 4-grams, and *as can be seen in, in the same way as,* and *the purpose of the current* are 5-grams. Figure 1 describes how we can analyze n-grams (2-gram, 3-gram, 4-gram, and 5-gram) with CasualConc.

Table 1

Examples of Lexical Bundles (4-grams)

the extent to which	as well as the
it is likely to	some of the most
on the other hand	as part of the
the role of the	as a matter of
by the end of	as a result of
on the other hand	on the basis of

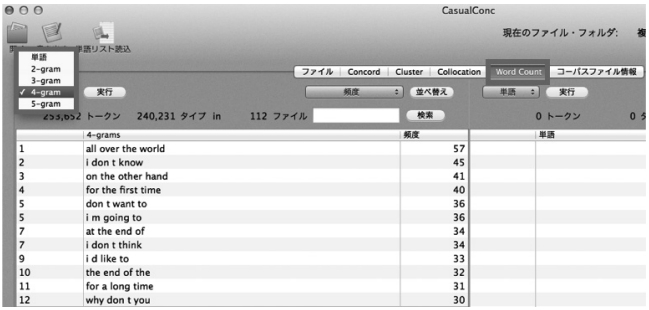

Figure 1. Choosing a type of n-grams in CasualConc.

Although lexical bundles are just a sequence of words, it is known that they serve as a building block of discourse and they show that the writer belongs to the discourse community (Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2004; Cortes, 2004; Hyland, 2008a, 2008b).

In corpus research, "4-grams are often used because they are far more common than 5-word strings and offer a clearer range of structures and functions than 3-word bundles" (Hyland, 2008b) and also "many four-word bundles hold three-word bundles in their structures" (Cortes, 2004).

Research findings from previous studies on lexical bundles

Research on lexical bundles supports their importance in a specific discourse. Biber et al. (1999) reported that 3- and 4-word lexical bundles made up 28 percent of the conversation and 20 percent of the academic prose. It is thus possible to categorize the functions of lexical bundles according to their use in a discourse (Biber et al., 2004; Hyland, 2008b). The effects of lexical bundles can be seen in their use across genres and registers. Hyland (2008a), for instance, most frequent 4-grams in the four sub-corpora of British National Corpus Baby edition differed considerably across four genres (i.e., academic writing, imaginative writing, newspaper texts, and spontaneous conversation). Pursuing the same line of research, Mizumoto (2009) reported the use of 4-grams in 10 international academic journals in the field of applied linguistics over 10 years with a 31 million-word corpus (Figure 2) varied across journals according to their aims and scope.

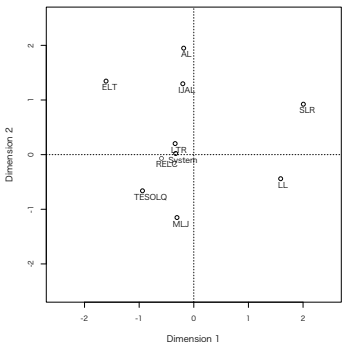

Figure 2. Distribution of journals based on their 4-gram use.

Teaching lexical bundles

As Ferris (2011) notes, "skillful incorporation of academic-sounding lexical bundles will help students sound more like they belong in the academic discourse community of which they wish to be members." Teaching lexical bundles is thus especially important for writing teachers of English for Academic/Specific Purposes (EAP/ESP). We often surmise that learners start to use those effective lexical bundles in a discourse naturally because word combinations "are so common, it might be assumed that lexical bundles are simple expressions, and that they will therefore be acquired easily" (Biber & Conrad, 1999, p. 188). However, research has suggested that that is not the case and students in the disciplines need formal instructions on the use of lexical bundles (Cortes, 2004, 2006). Materials for teaching lexical bundles effectively have not been developed so far, but by incorporating efficient and innovative teaching methods such as data-driven learning (DDL; see Boulton, 2010, for details), it would be possible to cultivate more proficient and persuasive writers in a certain discourse.

References

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer software]. Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net/

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26,* 263-286. doi:10.1016/j.esp.2006.08.003

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. 181-189). Amsterdam: Rodopi.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at. . .: Lexical bundles in university teaching and textbooks. *Applied linguistics, 25*, 371-405. doi: 10.1093/applin/25.3.371

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning, 60,* 534-572. doi:10.1111/j.1467-9922.2010.00566.x

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing:

Examples from history and biology. *English for Specific Purposes, 23,* 397-423. doi:10.1016/j.esp.2003.12.001

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education, 17,* 391-406. doi:10.1016/j.linged.2007.02.001

Ferris, D. R. (2011). *Treatment of error in second language student writing* (2nd ed.). Ann Arbor, MI: University of Michigan Press.

Granger, S. (2014). A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast, 14,* 58–72. doi:10.1075/lic.14.1.04gra

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18,* 41-62. doi: 10.1111/j.1473-4192.2008.00178.x

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27,* 4-21. doi: 10.1016/j.esp.2007.06.001

Imao, Y. (2014). CasualConc (Version 1.9.7) [Computer software]. Retrieved from https://sites.google.com/site/casualconc/Home

Mizumoto, A. (2009). Characteristic lexical bundles in a corpus of applied linguistics journals. *Paper Presented at Kansai Chapter, Language Education and Technology,* May 16, Kyoto, Japan.

Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, pricessing and use*. Amsterdam/Philadelphia: John Benjamins.

Wray, A. (2012). Formulaic sequences. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 2200-2205). Oxford: Wiley-Blackwell. doi:10.1002/9781405198431.wbeal0434