Cross-cultural and paradigmatic influences on lexical bundles in English academic writing

Duygu Candarli (University of Manchester, UK)

Lexical bundles, which are defined as "the most frequently recurring sequence of words" (Biber et al., 1999, p. 90), play a central role and fulfil a wide range of important functions in English academic writing. The main discourse functions of lexical bundles can be summarised as follows (Biber, Conrad, & Cortes, 2004): they present arguments and frame attributes through referential expressions (e.g. 'in the context of'), organise texts through discourse organisers (e.g. 'on the other hand'), and convey attitudes, certainty and hedging through stance expressions (e.g. 'it is possible to'). Lexical bundles reflect how academic writing is framed by a particular discourse community, reveal disciplinary membership and enhance effective communication in academic writing (Cortes, 2004; Hyland, 2008). Uncovering preferred ways of organisation and meaning-making, lexical bundles can also shed light on cultural preferences in academic writing (Hyland, 2008).

A large and growing body of literature has investigated lexical bundles in published English academic writing (e.g. Ädel & Erman, 2012; Biber, 2009; Cortes, 2004; Hyland, 2008). Most of these studies focused on four-word lexical bundles and investigated their discourse functions in academic prose. The general trend that emerges from the study of lexical bundles is reliance on noun and prepositional phrases which primarily serve as referential expressions in published English academic writing. In a recent cross-cultural study, Pan, Reppen and Biber (2016) examined lexical bundles in telecommunication research articles written in English by L1 English and L1 Chinese academic writers and found that lexical bundles in the two groups differed considerably in terms of functional and structural categories. Little attention has been paid to the use of lexical bundles in research articles written in English by L1 Turkish academic writers and paradigmatic influences in research on lexical bundles. A research paradigm can be defined as "a shared belief system or set of principles on what problems are to be investigated and how to investigate them" (Cohen, Manion, & Morrison, 2013, p. 13). While there is a multiplicity of research paradigms in education research, quantitative and qualitative paradigms are considered the two major paradigms in the discipline of education (Cohen et al., 2013; Creswell, 2009). Quantitative and qualitative paradigms are held to govern the methodological procedures and influence the discoursal and rhetorical conventions of research output (Cohen et al., 2013; Hu & Cao, 2015). Hu and Cao (2015) found that the quantitative research articles in the disciplines of applied linguistics, psychology, and education made more frequent use of hedges than the qualitative research articles. Based on this finding, it was hypothesised that there would be cross-cultural and cross-paradigmatic influences on the frequencies of stance expressions and discourse organisers in English academic writing.

The aim of the present study is to identify the discourse functions of lexical bundles and investigate cross-cultural and cross-paradigmatic influences on the frequency of each functional category of lexical bundles in English academic writing in the discipline of education. The two specific research questions addressed are as follows:

1. To what extent, if any, is there cross-cultural variation in the discourse functions of lexical bundles between English research articles published in international journals and English research articles written by L1 Turkish academic writers?

2. To what extent, if any, is there paradigmatic difference in the discourse functions of lexical bundles between the quantitative and qualitative research articles in the discipline of education?

Efforts were made to construct corpora which are similar in terms of genre, size, subject matter, and date. Accordingly, the empirical research articles that were published in SSCI-indexed journals in the discipline of education were extracted from the period of 2011-2015 to compile research articles that were published in international English-medium journals and those that were written in English by L1 Turkish academic writers. At the next stage, the research articles were coded into research methodology paradigms, as in Hu and Cao's (2015) study, according to Creswell's (2009) descriptions of quantitative, qualitative and mixed methods research paradigms in education research. When the articles did not fit neatly into these three categories, they were coded as other. Finally, a balanced number of quantitative and qualitative research articles were selected to include in the four comparable corpora (c. 1 million each).

In order to extract lexical bundles, the frequency cut-off point was set to 20 times per million words for four-word lexical bundles, and the cut-off point for range was kept at five percent of the texts, i.e. at least in seven texts in each corpus. Lexical bundles were extracted by using WordSmith Tools Version 6.0. Four-word lexical bundles were examined in terms of their discourse functions. After coding all the discourse functions of lexical bundles, a script in the Python language was written to record the token frequencies of each discoursal category of lexical bundles (i.e. referential expressions, discourse organisers, and stance expressions) for each file, which allowed me to use inferential statistics in this study. A series of two-way ANOVAs were performed in order to compare cross-cultural (English in international journals vs L2 English of L1 Turkish academic writers) and paradigmatic (quantitative vs qualitative) influences on the normalised frequencies (per 1000 words per text) of referential expressions, discourse organisers, and stance expressions. White's adjustment, which provides a correction for the heterogeneity of variance was used for two-way ANOVAs in R (R Core Team, 2016). When a statistically significant interaction was found, simple effects analysis was conducted in order to describe the nature of the interaction. When a significant main effect of cross-cultural influence was detected, the post hoc Games-Howell test was conducted to make further comparisons.

The results revealed a significant interaction between cross-culturality and paradigm in the frequencies of discourse organisers and stance expressions, and significant effects of cross-culturality and paradigm in the frequencies of referential expressions. Paradigmatic influences were greatest for stance expressions in that the quantitative research articles published in international journals contained more stance expressions than the qualitative research articles. On the other hand, no statistically significant difference in the frequency of stance expressions was found between the quantitative research articles and qualitative research articles written in English by L1 Turkish academic writers. An opposite trend emerged for the frequencies of discourse organisers in that the quantitative research articles written in English by L1 Turkish academic writers included significantly more discourse organisers than the qualitative

research articles written in English by L1 Turkish academic writers. No statistically significant difference in the frequencies of discourse organisers was found between the quantitative and qualitative research articles published in international journals. The key qualitative similarities and differences in the use of lexical bundles across four corpora will also be discussed. The results of this study have important implications for academic writing practices and teaching academic writing in English.

References

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, *31*(2), 81–92.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*(3), 275–311.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ... : Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371–405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, UK: Pearson.

Cohen, L., Manion, L., & Morrison, K. (2013). *Research methods in education* (7th ed.). London: Routledge.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*(4), 397–423.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Los Angeles, CA: SAGE.

Hu, G., & Cao, F. (2015). Disciplinary and paradigmatic influences on interactional metadiscourse in research articles. *English for Specific Purposes*, *39*, 12–25.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4–21.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, *21*, 60–71.

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: https://www.R-project.org.