A corpus-based investigation of the lexical bundle use by accomplished and novice mathematics writers

Abdullah Alasmary (the Department of English Language and Translation, King Saud University, Saudi Arabia)

While there has been a growing body of research into lexical bundle use in a wide range of academic disciplines, little attention is given to domains of highly theoretical and abstract nature. Drawing on a 2-million-word corpus of doctoral dissertations, textbook chapters and peer-reviewed journal articles, this research is an attempt to fill this gap, comparing the use of recurrent bundles by published authors and graduate students in mathematics. A total of 291 four-word bundles that recur 25 times per million and appear in at least 15 % of the texts were retrieved and their structural and functional attributes examined. Results show that graduate students tend to use fewer and less varied lexical bundles than textbook authors and article writers, an outcome that reinforces a pattern in the studies comparing accomplished authors and novice students in various academic disciplines. Structural analysis of the lexical bundles demonstrates that while the student writers rely more on phrasal patterns, the expert writers prefer bundles comprising clausal constructions. The analysis of the functional attributes of mathematical bundles reveals no difference between groups, with all mathematician writers relying more on research- and text-oriented bundles, but showing little interest in using participant oriented ones.

1. Introduction

Disciplinary writing constitutes a great proportion of semantically transparent and grammatically regular multiword units, which recur frequently, performing identifiable discourse functions and displaying analyzable structural patterns (Ädel & Erman, 2012; Chen & Baker, 2010; Biber, Conrad, & Cortes, 2004; Cortes, 2006, 2013; Hyland, 2008a, 2008b; Pan, Reppen, & Biber, 2016). Several terms are used to refer to such units, including academic clusters (Hyland, 2008a), phrasal expressions (Martinez & Schmitt, 2012), lexical bundles (Biber et al., 2004; Pan et al., 2016), academic formulas (Simpson-Vlach & Ellis, 2010), multi-word constructions (Liu, 2012) and recurrent word combinations (Ädel & Erman, 2012). While these terms can be placed under the umbrella of formulaic language(Wood, 2015; Wray, 2002), they are used here to refer to a special set of preassembled words whose meaning can be inferred from the meaning of their constituent parts and their identification in natural discourse involves applying some frequency and dispersion measures (Cortes, 2004).

While there is a great body of research investigating the use of lexical bundles in a variety of academic disciplines, mathematical discourse appears to be a notable exception. This may be due to the multilayered nature of the mathematical register which requires attention to other nonlinguistic means of

framing and advancing mathematical claims such as the symbolic and graphic representations accompanying a text (Huang & Normandia, 2007; O'Halloran, 2005; Schleppegrell, 2007). Another reason is the assumption that mathematics is the least language-dependent discipline, enough reason to impel researchers to examine bundles in rather supposedly language-rich contexts. Yet a third factor relates to the complex network of grammatical patterns characteristic of math discourse such as the use of highly technical lexis, patterns with special meaning, and discipline-specific rhetorical devices for describing underlying mathematical relationships, all of which seem to flummox non-specialists. Despite this obvious negligence, studies carried out in the field of mathematics education and English for Academic Purposes (EAP) concur that language plays an important role in how mathematical knowledge is constructed, disseminated and interpreted by the disciplinary members. As Graves, Moghaddasi and Hashim (2014) put it, "writing in mathematics is not different from other disciplines were a sophisticated awareness is required of communicative acts typical within that discourse community" (p.8). It is, therefore, plausible to employ linguistic tools to study corpus-derived patterns that contribute to knowledge making in a range of mathematics registers.

2. Corpora

Given that this study is comparative in nature, it was decided to create three parallel corpora comprising texts representing three academic registers: textbooks, peer-reviewed journal articles and PhD dissertations. The first corpus, named as TXT-BKs, includes chapters taken from textbooks aimed for graduate students in the field of mathematics. In order to guard against idiosyncratic use featuring some authors or texts, a decision was taken to include only one chapter from each textbook, with the final texts amounting to forty-three chapters. Each chapter is given a distinct code so as to trace the source of lexical bundles and to facilitate the process of the subsequent concordance checks.

The second corpus incorporates thirty-four peer-reviewed articles that appeared in the period between 2013 and 2015. Labelled as JOL-ARs, this corpus represents writings produced by notable authors and are published in leading, high-impact journals. The selection process of articles in this corpus was guided by three criteria. First, all journals from which articles are drawn must have a high impact factor and are indexed by the ISI Web of Knowledge. Second, only one article per issue and per author is included, so as to guard against issue or writer idiosyncrasies. Third, the content of each selected journal article must be in an electronic format so as to allow for its synthesis and analysis using automated corpus tools. Each article is cleared from appendices, references and acknowledgements.

The third corpus (DISS-ONs) is a collection of twenty PhD dissertations written by graduate students (10 males and 10 females). First language is highlighted in several studies as the cause of errors in the use of lexical bundles, so it was decided to control this variable by selecting the nonnative writings

produced by speakers whose first language is Arabic and for whom English is a foreign language. All the dissertations analyzed in this study are presented as texts upon which the Doctorate Degree was awarded to candidates who were pursuing their studies at English-only institutions and in countries where English is assigned the native language of the population. Acknowledgements, references and appendices are removed from the final list of texts in this corpus.

3. Key results

Key findings emerged as a result of the quantitative and qualitative analyses. First, expert mathematician writers exhibit control of and sensitivity to a wide range of academic clusters in comparison with graduate mathematician students who rely on a small number of highly recurrent strings. This gives further evidence to some previous findings which indicate that the successful use of lexical bundles hinges more on the academic maturity and expertise of the disciplinary writers, rather than on their native/non-native backgrounds (Haswell, 1991). Another finding is that a large number of lexical bundles in each writing group show a strong tendency to spread across registers, with the greatest proportion of the shared bundles found in the student mathematics discourse. Yet the other key finding relates to the structural and functional attributes of the bundles used by each writing groups. A great number of the lexical bundles in the student-produced texts are phrasal, while the majority of the bundles in the expert corpora are clausal, a finding that sheds doubts on the widely held assumption that as writers mature, their reliance on phrasal style become more clearly manifested (Pan et al., 2016).

Capitalizing on these results, there are several pedagogical implications for the teaching and learning of lexical bundles in the disciplines. Instructional intervention should focus not only on the forms that such lexical bundles display but also on the discourse functions that they serve. Given their perceptual saliency, the structural patterns of the academic bundles seem to be more conducive to learning and retention than the functional attributes which necessitate exposure to and familiarity with the contexts in which these bundles occur. In a form-focused situation, ESP/ EAP instructors and materials developers can use various techniques such as "highlighting" "underlining" or "bolding" to draw the attention of students and novice writers to the numerous structural types underlying bundles. The functions of the mathematical bundles can be instructionally fostered using labels discussed in this study or those alluded to by Cunningham (2017) and McGrath and Kuteeva (2012). These functional labels can be supplemented by examples derived from a specialized corpus, showing instances of occurrence as they are used by expert authors. Disciplinary textbooks and journal articles should also be incorporated into advanced ESP and EAP programs as writing models to emulate, for the reason that these texts are written by academically accomplished and profoundly experienced members whose publications undergo extensive peer-reviewing, editing and rewriting. A genre-based instruction targeting recurrent bundles could also help novice

mathematician writers accelerate transition from a student-styled use of language into more expert-styled one, a shift highlighted by Schleppegrell (2007) as posing a challenge to many aspiring novice mathematicians.

In the end, his study has shown that the mathematical discourse can be studied from a lexical bundle perspective, using a combination of quantitative and qualitative measures. It is hoped that the discussion above will enhance our understanding of how lexical bundles, defined as the most recurrent, widely dispersed n-grams in a discourse, are structurally and functionally used by mathematician writers at different levels of expertise.

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

Chen, Y.-H., & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *14*(2), 30–49. *Language Learning & Technology, 14*(2), 30-49.

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, *17*(4), 391–406.

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, *12*(1), 33–43.

Cunningham, K. J. (2017). A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes*, *25*, 71–83.

Graves, H., Moghaddasi, S., & Hashim, A. (2014). "Let G = (V, E) be a graph": Turning the abstract into the tangible in introductions in mathematics research articles. *English for Specific Purposes*, *36*, 1–11.

Haswell, R. (1991). *Gaining ground in college writing: Tales of development and interpretation*. Dallas: Southern Methodist University Press.

Huang, J., & Normandia, B. (2007). Learning the language of mathematics: a study of student writing. *International Journal of Applied Linguistics*, *17*(3), 294–318.

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 41–62.

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

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, *31*(1), 25–35.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, *33*(3), 299–320.

McGrath, L., & Kuteeva, M. (2012). Stance and engagement in pure mathematics research articles: Linking discourse features to disciplinary practices. *English for Specific Purposes*, *31*(3), 161–173.

O'Halloran, K. (2005). *Mathematical discourse: Language, Symbolism and Visual Images*. London: Continuum.

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.

Schleppegrell, M. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading and Writing Quarterly*, *23*, 139–159.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487–512.

Wood, D. (2015). *Fundamentals of Formulaic Language: An introduction*. London: Bloomsbury.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.