http://www.kacl.or.kr

Corpus Linguistics Research
Vol. 4 (September 2018), 99-114

ISSN 2465-812X
http://dx.doi.org/10.18659/CLR.2018.1.0.05

# *Participant-oriented Lexical Bundles in Graduate-level Math Textbooks: A Corpus-based Exploration**

Abdullah A Alasmary

*King Saud University*

Received 29 June 2018; received in revised form 19 August 2018;
accepted 15 September 2018; available online 30 September 2018

**ABSTRACT**

This research intends to investigate lexical bundles performing a participant-oriented function in mathematics. Drawing on a 5-million-word corpus of graduate-level textbooks, a total of forty-three bundles recurring at least 20 times per million words (pmw) and spreading in 10 % or more of the texts making up the corpus are retrieved and further subjected to structural and functional analyses. On a structural level, results show that the greatest number of participant-oriented bundles are clauses or clause fragments. Functionally, these patterns are used as rhetorical devices for engagement or as means for expressing the author's opinions, judgment and evaluation. Implications for language instruction and materials designing are discussed.

*Keywords:* lexical bundles, mathematical discourse, corpus analysis, recurrent word combinations

## 1. Introduction

Several corpus-based studies concur that academic speech and writing comprise a great number of recurrent lexical patterns which represent different structural forms and serve distinct discourse functions (Biber, Johansson, Leech, Conrad & Finegan, 1999; Gray & Biber, 2013; Hyland, 2012). These recurrent patterns are investigated under a wide array of concepts such as multiword constructions (Liu, 2012; Wood & Appel, 2014), formulas (Simpson-Vlach & Ellis, 2010) and clusters (Hyland, 2008a). However, the most common of these terms is that of lexical bundles (Biber & Barbieri, 2007; Biber, Conrad & Cortes, 2004; Biber et al., 1999; Cortes, 2004; Herbel-Eisenmann,

* The author extends his appreciation to the Deanship of Scientific Research at King Saud University for funding this work.

Wagner & Cortes, 2010). Bundles as such are defined as "sequences of word forms that commonly go together in natural discourse" (Biber et al., 1999: 990). Much research into lexical bundles has focused on the use of these patterns by different groups of writers, including professionals (Pan, Reppen & Biber, 2016), native and nonnative speakers of English (Ädel & Erman, 2012; Chen & Baker, 2010) and journal article authors and graduate students (Hyland, 2008b). Other research studies have explored lexical bundles as they occur in written as opposed to spoken registers (e.g., Biber & Barbieri, 2007). The register-specific nature of many lexical bundles has led researchers to examine several disciplines such as law (Breeze, 2013), biology and history (Cortes, 2004), pharmacy (Grabowski, 2015) and applied linguistics (Qin, 2014). Taken together, these studies seem to suggest that the use of lexical bundles is affected by the background of text authors, their level of expertise and the register in which they operate.

While there has been a growing interest in the study of lexical bundles in several domains, mathematics seems to receive little scrutiny. The reasons for such lack of interest are wide-ranging, the most obvious of which relates to the complex nature of mathematical register in which knowledge is constructed and disseminated using linguistic means, graphic representations and symbolic notations (O'Halloran, 2005). Drawing on the frameworks developed by Hyland (2008a, 2008b) and (Biber, 2006), this study focuses on lexical bundles recurring in a corpus of graduate-level mathematics textbooks and serving a participant-oriented function. It is hoped that this study sheds some light on how expert mathematician writers convey their judgements, opinions and attitudes in a discipline widely perceived as "logic-driven and argumentation-mediated" (Graves, Moghaddasi, & Hashim, 2014).

## 1. Lexical bundles & academic register

Textbooks are a key component of what Hyland (2009) has termed "instructional discourses", which also include university lectures and seminars. Academic texts as such are important for knowledge construction and dissemination as they constitute "one of the primary means by which the concepts and analytical methods of a discipline are acquired" (Hyland, 2009: 112). The situated characteristics of academic textbooks have been examined in a study by Biber and Conrad (2009) who noted that texts represent a written format, produced under no time and/or space constraints, taken to fulfil a combination of descriptive and explanatory purposes, aimed to educate and inform and are generally created to address specific topics. An area of interest that focuses on academic textbooks is the study of the role that language plays while laying

out facts, advancing arguments and displaying opinions in a wide range of disciplines.

In a pioneering, yet comprehensive, account of lexical bundles, Biber et al. (1999) unveiled a broad range of structurally different and functionally dissimilar bundles. In terms of the functions that these bundles serve in the discourse, authors have identified patterns associated with the speaker's or writer's stance on the argument or ideas that are being expressed in the text. Using corpus-based tools, subsequent researchers have taken a step further and suggest a classification scheme that is based on the structural attributes of the recurrent bundles. In a study to probe the distribution of lexical bundles in several disciplines, Durrant (2017) notes that stance bundles are far more common in humanities than in science and technology writings. Grouped in three categories, stance bundles are found to mark the significance of an entity, the centrality of an argument, or modality. In a series of studies, Hyland (2008a, 2008b) investigates the distribution of lexical bundles in academic domains (electrical engineering, business studies, applied linguistics, microbiology) and registers (research articles, dissertations and theses), suggesting that these bundles are functionally grouped into research-oriented, text-oriented or participant-oriented. Within participant-oriented group, bundles can either be used to mark stance (e.g., *it is possible that*) or engagement (e.g., *as can be seen).*

Researchers have long focused on the distributional, structural and functional attributes of lexical bundles in a variety of disciplines and registers. In a study to compare the presence of lexical bundles across four registers, including textbooks, Biber et al. (2004) noted that textbooks incorporated the smallest range of bundles compared with conversation, classroom teaching and academic prose. Researchers attribute the paucity of bundles in textbooks to a tendency from the part of authors to use longer statements as they are "free of the real-time production constraints of face-to-face teaching and therefore make more diverse language choices" (Biber et al., 2004). In terms of the functions that the bundles serve, stance bundles make the second most frequently used category. A framework for analyzing stance and engagement bundles was unveiled by Hyland (2005) who divided interaction in the academic discourse into stance and engagement, each of which branches out into smaller sub-categories. In another study, Biber (2006) examined the distribution of expressions fulfilling a stance function in a spoken and written registers, thus outlining a framework for analyzing and describing these expressions. According to Biber (2006: 101-102), expressions marking stance fall into three major categories: modal and semi-modal verbs, stance adverbs and complement clauses controlled by verbs. Modal and semi-modal verbs are used to convey meanings of possibility (e.g., *may*), permission (e.g., *can I ask a question*), ability (e.g., *could*), necessity (e.g., *must*), obligation (e.g., *ought to*), precision (e.g., *will*) and volition (e.g., *shall*). Stance adverbs serve to carry epistemic, attitudinal and personal senses

(e.g., *actually, amazingly*, and *generally*). The third, and by far the largest group, is that of complement clauses controlled by stance verbs (e.g., *expect, hope, worry*), stance adjectives (e.g., *certain, clear, probable*) and by stance nouns (e.g., *conclusion, fact, hope, view*).

## 2. Mathematical discourse and the lexical bundles research

While it is true that lexical bundle research in mathematics is not as robust as it is in other domains, still there are few attempts which have been made to probe the use of lexical bundles by different mathematician groups. Within ESP, Some key studies have been conducted to explore the use and distribution of vocabulary, including lexical bundles, in different mathematical contexts. Although the focus of the study by McGrath and Kuteeva (2012) was not on lexical bundles, but rather on individual lexical items, it has thus yielded some important insights into pure mathematicians' use of stance expressions in a corpus of journal articles. Stance expressions-hedges, attitude markers, boosters, self-mentions-are underused by math authors in comparison with engagement markers-reader references, directives, questions, knowledge references, and asides. Cunningham (2017) compiled a corpus of research articles in mathematics and looked for recurrent lexical frames with fillable slots. These frames are then classified into three distinct groups based on the three functions that they serve: aboutness (e.g., *the proof of the #*), coherence (e.g., *the main result of #*) and text moves (e.g., *such that the following #*).

In mathematics education, there is a key study which has attempted to investigate stance bundles. Herbel-Eisenmann et al., (2010) looked at lexical bundles in small-sized corpus of teacher-student classroom-based interactions. A total of 71 different bundles have been retrieved, nearly half of which serve a stance function. Some stance bundles signal uncertainty and imprecision (e.g., *I don't know if)* while some others encode desire (e.g., *I want you to)* and intention/prediction (e.g., *we are going to).* Although this study gives an account of lexical bundles in speech, it is important to note that professional academic writing merits similar attention. Our study is an attempt to bridge this gap, thus addressing how text authors make use of participant-oriented lexical bundles in mathematics texts.

# 3. Methodology

## 3.1. Criteria for bundle selection

Lexical bundles are conceptualized as sequences of uninterrupted words which are highly recurrent and widely dispersed in a corpus of naturally-occurring content. To select lexical bundles from a collection of texts, researchers have considered three key criteria: the length of the target sequence, its overall frequency and the degree at which it spreads across the texts making up the entire corpus. As for the length of the sequence, there is a tendency in lexical bundles research to focus on four-word sequences as they are more frequently-occurring than five- or six-word bundles and are more structurally and functionally robust than two- or three-word sequences (Biber et al., 1999). Frequency of occurrence is a second key criterion which ranges from 10 times (pmw) (Biber et al., 1999) to 50 times (pmw) (Breeze, 2013). In this study, bundles occurring at least 20 times (pmw) are retrieved for analysis, following the cut-off score used in some similar previous studies (Cortes, 2004; Csomay, 2013; Hyland, 2008a, 2008b). Yet a third principle guiding the process of selecting lexical bundles is that of distribution, or the extent to which a certain lexical bundle occurs across the corpus subparts. When a lexical pattern recurs frequently in a single text, this may reflect an idiosyncratic use typical of the author or the text and by applying the principle of distribution, researchers want to ensure that bundles appear in as many texts as possible. There is of course no consensus among researchers on a specific distribution threshold above which bundles can be selected for the analysis. Given that this study draws on Hyland's (2008a, 2008b) analytical framework, the decision is taken to select bundles appearing in at least four texts (approximately 10 % of texts).

## 3.2. Corpus compilation and refinement

The corpus is comprised of thirty-six textbooks aimed for graduate-level students in the domain of mathematics (see Table 1). To make the sample of texts more representative, the decision is taken to include textbooks from a broad range of mathematical topics, including algebra, number theory, geometry, calculus and typology, just to name but a few. Appendices, references and acknowledgements are removed from the texts prior to the corpus treatment. These books are then converted into plain texts, thus allowing for their extraction using *Cluster Function* in the *Wordsmith Tools* (Scott, 2016). Four-word bundles occurring 20 times per million words and appearing in at least 10 % of texts are retrieved. The overall number of bundles obtained as a result of applying frequency and distribution measures amounts to 293 patterns.

The list was given to two raters who had agreed to sort out manually the bundles that fulfil a participant-oriented function. Both raters are language instructors and have abundant experience in teaching English for academic purposes. To ensure consistency, raters were both given a detailed description of the study objectives and were introduced into key categorization schemes of stance bundles in academic register proposed by Hyland (2005) and Biber (2006). The two raters agreed on 85 % of all cases. To solve discrepancies and reach 100% agreement among raters, concordance lines of unclassified bundles are checked and functional categories are accordingly assigned. Items on the list after assigning functions amount to forty-nine. A quick look at this list reveals that these bundles are not without problems, the first of which concerns the existence of overlapped bundles. Bundles such as *is easy to see* and *easy to see that* overlap with the bundle *it is easy to*. Following the practice in previous research (e.g., Chen & Baker, 2010), a decision is taken to combine these three bundles in a single lexical pattern, thus eliminating the repeated segments and placing the dissimilar parts in brackets to indicate that other variations also exist (*it is easy to +(see that)*. A second problem emerging from the list involves the presence of a group of bundles that are almost identical except in one lexical item occurring in the same position. The lexical bundle *let x be a* has nine similar patterns in which the difference lies in the variable *x* (e.g., *let A be a, le M be a, let V be a)*. Including all these bundles will result in an inflated list, so the decision is taken to keep the most recurrent bundle and remove less recurrent variations with an asterisk marking the existence other possible variables. The final list resulting from merging overlapped bundles and removing semi-identical ones incorporates a total of forty-three items (see the Appendix).

**Table 1.** Corpus description

| | Descriptive Statistics |
|---|---|
| Number of Texts | 36 |
| Tokens (running word) | 4,904,419 |
| Types (distinct words) | 42,723 |
| Type/token ration (TTR) | 1.27 |
| Standardized TTR | 0.80 |

## 4. Results

The purpose of this study is two-fold: unveiling the lexical bundles which serve a participant-oriented function and assigning these bundles to distinct structural and sub-functional categories. In the following section, I will highlight the findings of this

study, exploring the forms as well as the sub-functions of the participant-oriented expressions, supplementing each pattern with illustrative examples from the corpus data.

## 4.1. Structural patterns

Drawing on the classification scheme proposed by Biber et al. (1999), all bundles are grouped into two distinct categories: phrasal and clausal (Table 2). Five bundles, representing 12 % of the total number of bundles, are phrase-based whereas thirty-eight bundles (88%) are clauses or clause fragments with a verb component. Within the first category, two bundles contain a noun with an *of*-phrase fragment while the other is headed by the gerund *using*. Another phrase-based bundle consists of preposition with a post-modifier fragment. A final structural type in this group is a bundle which begins with the conjunction *and* and is followed by a noun phrase fragment.

The largest group of participant-oriented bundles comprises a clause or clause fragment. Bundles headed by the first-plural pronoun *we* appear to dominate this group with a total of elven lexical expressions, representing almost one-third of the verb-based bundles and a quarter of all participant-oriented bundles uncovered in the study. Two similar patterns emerging from the structural analysis of bundles include anticipatory *It* structure and clause fragments with an active verb, each is represented by eight bundles. Of the eight bundles in the active voice, five bundles begin with three imperatives: *let, suppose and use.* It seems clear that *Let*-frame is productive as it comes with different mathematical variables. Bundles in the passive voice are represented by six bundles, five of which are formed using the modal verb *can.* The last structural group includes five bundle types which contain *copular be* followed by a noun phrase, adjective phrase, or *to*-infinitive clause.

**Table 2.** Distribution of structural patterns

| Structure | Sub-structures | # of structures | % of all structures | Examples |
|---|---|---|---|---|
| Phrasal | noun-based | | 6.98 | *a special case of* |
| | preposition-based | | 2.32 | *from the fact that* |
| | others | | 2.32 | *and the fact that* |
| Clausal | we + dependent clause fragment | | 25.59 | *we may assume that* |
| | anticipatory it | | 18.6 | *it is easy to* |
| | verb phrase with non-passive verb | | 18.6 | *let x be a* |
| | verb phrase with passive verb | | 13.96 | *can be written as* |
| | Verb *be* +*noun/adjective/infinitival* | | 11.63 | *is a normal group* |
| | Total | 43 | 100 | |

## 4.2. Engagement bundles

A key function of participant-oriented bundles is to allude to "the ways writers intervene to actively address readers as participants in the unfolding discourse" (Hyland, 2008b, p.18). It is clear that the greatest number of bundles in the data are initiated by the first person plural pronoun *we*. These statements contain different types of verbs and end up with either *that* clauses (e.g. *we may assume that, we will show that),* definiteness/indefiniteness markers (e.g. *we can find a, we say that a)* or infinitival form (e.g. *we need to show).* The verb *assume* occurs twice and carries a different sense in mathematics than in everyday interaction. Making an assumption in mathematics indicates that a process to elaborate on a previously carried out operation is underway, as is exemplified in the following statements:

1) Proof. Let n be the size of H. We argue by induction on n. **We may assume that** *an* is the largest component of *a*.
2) Therefore **we can assume that** in our sequence of moves, all moves of type (c) are bendings and tightenings.

Besides making assumptions, other statements initiated by the inclusive *we* serve to announce future procedures or intention (*we will show that, we are going to*), express desire or possibility *(we need to show, we can find a*), comment on an ongoing or past operation (*we see that the, we say that a, we now show that, we have seen that*) or to draw the reader's attention to a subsequent procedure or operation (*we have the following*). Given the space limitation, each group will be represented by a single example:

3) In the chapter that follows this prelude **we will show that** this construction is key to our understanding of both geometric and computational aspects of the kind of multiresolutions that can be built on X.
4) **We need to show** the following four facts.
5) To go further, we need something more. **We have seen that** the functor of local deformations is pro-representable for simple vector bundles.
6) By the same argument **we have the following** general result.

The second subcategory of engagement features include directive bundles which "instruct the reader to perform an action or to see things in a way determined by the writer" (Hyland, 2005: 184). The group contains eight recurrent patterns, beginning

with directive verbs *let, need, suppose, use* and *show.* Let-initiated statements are ubiquitous in our data which primarily serve to start off the process of proving, or disproving, a theorem. The verb *show* carries a different sense here as it is used as synonymous of the verb prove, thus alluding to the intention of the author to engage in a proof-related task.

7) **Let X be a** locally compact Hausdorff space and A a Banach algebra.
8) Thus, we only **need to show that** ρ is a homeomorphism.

A third group of engagement expressions are used to appeal to shared knowledge. This group contains four patterns, nearly all of which revolve around the word *fact:( from) the fact that the, and the fact that, does not depend on,* and *using the fact that.*

9) The first statement follows from **the fact that the** knapsack cannot contain all the items in a minimal cover.
10) For this definition to make sense, we have to verify that it **does not depend on** the choice of the curves involved.

## 4.3. Stance bundles

Now we turn to the set of bundles that serve a stance function. Stance is broadly defined as "the ways writers present themselves and convey their judgments, opinions, and commitments" (Hyland, 2005: 174). The largest group of stance bundles include extraposed complement clauses reflecting the attitude or opinion of the writer or author. These can be classified further into smaller subgroups, thus signaling the ease/difficulty of processing (*it is easy to, it is hard to, it is not hard to*), sufficiency (*it suffices to prove, it is enough to*) and intelligibility (*it is clear that*). Here are examples of each subgroup:

1) Then **it is easy to** see that there is a hook length equal to 2.
2) Multiplying an arbitrary torsion element by an appropriate power of p, **it suffices to prove** that there are no nonzero torsion elements of order prime to p.
3) **It is clear that** when the sums get more complicated, doing them by hand becomes out of the question.
4) The second group of stance expressions is composed of five strings, all of which contain the modal verb *can* and are followed by a passive construction (*can be written as, can be used to, can be found in, can be extended to* and *can be identified*

*with)*. These patterns are used to express possibility/ability meanings:

5) First let us show how the standard sphere **can be written as** a rotationally symmetric metric in all dimensions.
6) The 24 non-trivial eigenspaces of the torus **can be identified with** the 24 short roots in the root system of type F4 as follows.

In the third stance group, recurrent bundles revolve around two key adverbs: finitely and uniquely. The terms *finiteness* and *infiniteness* are key in describing the nature of items in a set or group. In a similar way, uniqueness is an important notion, denoting the existence of a certain property that is distinctive of a particular object. Here are two examples illustrative of these bundle types:

7) If I is an ideal of R and M **is a finitely generated** module such that every element of I annihilates some nonzero element of M, then there is a single nonzero element of M annihilated by all of I.
8) By a well-known result in the function theory of several complex variables, any reproducing kernel **is uniquely determined by** its values on the diagonal.

The last group incorporates stance bundles with various evaluative adjectives: special, fundamental, normal and easy. These are used to designate properties based on the evaluation and perspectives of the author.

9) As the solution set of a polynomial equation in two variables, an elliptic curve as defined here is **a special case of** a plane algebraic curve.

## 5. Discussion

This study attempts to examine lexical bundles which serve a participant-oriented function in a corpus of graduate-level mathematical texts. Using textual clues and expert insights, a total of forty-three participant-oriented bundles are then classified further according to whether they convey a stance function or are used as rhetorical devices for engaging the reader. The overall number of bundles (tokens) in this study resembles that of Herbel-Eisenmann et al., (2010) although both studies draw on corpora of different sizes. This finding gives greater credence to the idea that participant-oriented bundles can also be examined even if the corpus size is relatively small. The structures of most bundles are clausal rather than phrasal, an outcome that is congruent with

findings reported in some previous research (Biber et al., 2004; Durrant, 2017). Some researchers, however, have pointed out that the process of writing a textbook or textbook chapter is not constrained by time and space in which authors have to rely more on short phrases rather than long clauses while encoding meanings (Hyland, 2008a, 2008b).

It seems clear that mathematician textbook authors have exhibited a preference for a wide range of different participant-oriented lexical patterns, the greatest number of which are used to convey epistemic stance towards the proposition that is being made. Most stance bundles are initiated by the anticipatory-*It*, thus expressing meanings associated with possibility and ease of processing. Three more patterns have emerged from the data which have not been recognized in the previous research on lexical bundles (Durrant, 2017). The bundles *it suffices to show*, *it is convenient that* and *it suffices to prove* are quite frequent in our data and are thus used to signal sufficiency and conveniency, two key notions in mathematical register. Furthermore, stance expressions are utilized in passive constructions to mark ability. The presence of passive forms in our data seems to go against the finding reported in Cunningham's (2017) study which found no passive patterns in a corpus of mathematical journal articles. However, it should be noted that Cunningham's study focuses on frames with slots to be filled with lexical items and does not include for analysis uninterrupted strings of words. Future researchers into participant-oriented bundles need to be aware of the fact that studying mathematics discourse using a frame-based approach may not yield comparable functional patterns. A final group of stance bundles are used to transfer the author's attitudes or certainty regarding the argument in the text. The greater number of attitudinal patterns is unsurprising, given the importance of such patterns in constructing and disseminating knowledge (Biber, 2004). The list of patterns conveying certainty meanings revolve around nouns (e.g., *the fact that the*) whereas patterns signaling an attitudinal position are expressed using adjectives (e.g., *a special case of*) and adverbs (e.g., *is uniquely determined by*). Although mathematics is seen as a discipline of highly abstract nature, the occurrence of attitudinal patterns seems to suggest that mathematics writers exhibit a desire to project themselves more into the context of discussion.

Turning to engagement bundles, it seems clear that textbook authors show an interest in engaging the reader while laying out arguments and opinions. The first group of engagement bundles are headed by the inclusive plural pronoun *we*. The use of this pattern is quite common in academic writing, as is pointed out by Hyland (2005). This pattern is not confined to written registers, as is maintained by Herbel-Eisenmann et al. (2010) who report that the use of such structural form is quite frequent in their corpus of oral classroom interactions. We-patterns are followed by modal verbs which are used to signal intention, desire and/or possibility, as is the case in the bundles *we are going to, we will show that* and *we can assume that*. The final engagement

pattern involves the use of the directive verb *let*. *Let*-initiated sequences are very ubiquitous which are functionally used to alert the reader of an eminent mathematical operation.

## 6. Implications & Conclusion

There seems to be a consensus among researchers that recurrent word combinations are prevalent in academic prose and that their presence is associated with distinct functions. Given this prominent status, language instructors and materials designers need to draw the learners' attention to these patterns. One way to do so is to present these patterns explicitly to learners in a carefully selected, corpus-informed context. Easy-to-use corpora are now accessible to professionals as well as to amateurs who might consider building their own materials in a customized way. Software programs that extract frequent patterns from a text of group of texts are also available on several platforms. Another approach is to draw the learners' attention to patterns commonly used by experts and mature writers. Cortes (2004) notices that the bundles used by novice learners are markedly different from those used by professionals. The novice-expert gap can be bridged by means of highlighting expert-authorized norms of expressions in a way that make them easily noticed by immature writers. Another way to foster bundle learning is to use corpus-derived examples of various structures and functions. It is rather unwise to concentrate on on the structural forms of the target bundle and ignore its discourse function.

In conclusion, this study is an attempt to synthesize and analyze participant-oriented bundles in a corpus of mathematical texts. The analysis of data has yielded a total of forty-three different bundles that serve as a means for engaging the reader or signaling the author's stance. This research can be extended in various ways. Lexical bundles serving other functions such as text-structuring or research-oriented can similarly be examined for recurrent patterns of use. Other types of bundles such as collocations and phrases merit similar attention. Corpus-extracted bundles can also be given to a panel of experts to determine their pedagogical use. Taken together, these approaches to the study of recurrent patterns will deepen our understanding of how mathematical discourse is constructed, disseminated and interpreted.

## 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. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, *5*(2), 97-116.

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. (2009). *Register, Genre, and Style*. Cambridge: Cambridge University Press.

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

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

Breeze, R. (2013). Lexical bundles across four legal genres. *Intnational Journal of Corpus Linguistics*, *18*(2), 229-253.

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

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

Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied Linguistics*, *34*(3), 369-388.

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.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university Students' writing: Mapping the territories. *Applied Linguistics*, *38*(2), 165-193.

Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes*, *38*, 23-33.

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.

Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, *18*(1), 109-136.

Herbel-Eisenmann, B., Wagner, D. & Cortes, V. (2010). Lexical bundle analysis in mathematics classroom discourse: The significance of stance. *Educational Studies in Mathematics*, *75*(1), 23-42.

Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173-192.

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.

Hyland, K. (2009). *Academic Discourse: English in a Global Context*. London: Continuum.

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, *32*, 150-169.

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.

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.

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System*, *42*(1), 220-231.

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

Wood, D. C. & Appel, R. (2014). Multiword constructions in first year business and engineering university textbooks and EAP textbooks. *Journal of English for Academic Purposes*, *15*, 1-13.

Scott, M. (2016). *Wordsmith Tools7*. Stroud: Lexical Analysis Software.

## Appendix

| No. | word | freq. | % | texts | % |
|---|---|---|---|---|---|
| 1 | we may assume that | 735 | 0.01 | 34 | 94.44 |
| 2 | it is easy to + (see that) | 707 | 0.01 | 35 | 97.22 |
| 3 | let x* be a | 623 | 0.01 | 28 | 77.78 |
| 4 | it is clear that | 524 | | 35 | 97.22 |
| 5 | (it) +suffices to show that | 337 | | 30 | 83.33 |
| 6 | can be written as | 252 | | 32 | 88.89 |
| 7 | (from) +the fact that the | 249 | | 35 | 97.22 |
| 8 | we see that the | 243 | | 29 | 80.56 |
| 9 | is easy to check +(that) | 194 | | 24 | 66.67 |
| 10 | we have the following | 188 | | 29 | 80.56 |
| 11 | does not depend on | 183 | | 24 | 66.67 |
| 12 | we can find a | 171 | | 23 | 63.89 |
| 13 | we can assume that | 169 | | 17 | 47.22 |
| 14 | it is enough to | 159 | | 27 | 75.00 |
| 15 | we will show that | 159 | | 23 | 63.89 |
| 16 | and the fact that | 157 | | 29 | 80.56 |
| 17 | we say that a | 152 | | 29 | 80.56 |
| 18 | a special case of | 148 | | 31 | 86.11 |
| 19 | it suffices to prove | 146 | | 22 | 61.11 |
| 20 | can be used to | 145 | | 31 | 86.11 |
| 21 | let x and y | 145 | | 19 | 52.78 |
| 22 | are only finitely many | 142 | | 19 | 52.78 |
| 23 | need to show that | 135 | | 26 | 72.22 |
| 24 | can be found in | 133 | | 27 | 75.00 |
| 25 | (it)+ is not hard to | 133 | | 16 | 44.44 |
| 26 | it is possible to | 130 | | 26 | 72.22 |
| 27 | the fundamental theorem of | 130 | | 19 | 52.78 |
| 28 | is to show that | 124 | | 32 | 88.89 |
| 29 | we need to show | 122 | | 28 | 77.78 |
| 30 | suppose m is a | 119 | | 6 | 16.67 |
| 31 | use the fact that | 119 | | 29 | 80.56 |
| 32 | using the fact that | 118 | | 26 | 72.22 |
| 33 | is uniquely determined by | 116 | | 26 | 72.22 |
| 34 | we now show that | 116 | | 18 | 50.00 |

| No. | word | freq. | % | texts | % |
|---|---|---|---|---|---|
| 35 | can be extended to | 114 | | 29 | 80.56 |
| 36 | we are going to | 114 | | 18 | 50.00 |
| 37 | can be identified with | 108 | | 21 | 58.33 |
| 38 | we have seen that | 106 | | 26 | 72.22 |
| 39 | it is convenient to | 101 | | 25 | 69.44 |
| 40 | let e be an | 101 | | 14 | 38.89 |
| 41 | to show that if | 101 | | 28 | 77.78 |
| 42 | is a finitely generated | 100 | | 13 | 36.11 |
| 43 | is a normal subgroup | 100 | | 14 | 38.89 |

**variables A, M, V, R, S,E,* Γ, D also exist

## ◆ About the author ◆

**Abdullah A Alasmary**

Assistant professor
Department of English Language & Translation
College of Languages and Translation,
King Saud University
Riyadh 12372-7101, Saudi Arabia
aasmary@ksu.edu.sa