Comparing lexical bundles across three advanced mathematical text types: a corpus-based genre-focused investigation

Abdullah A. Alasmary

Abstract
Although much research has been conducted on academic multi-word units across a range of academic disciplines, little attention has been given to the distribution of such units in mathematics. Drawing on three parallel corpora of doctoral dissertations, textbook chapters, and research journal articles, this study compares the use of recurrent bundles across three mathematical text types using a combination of corpus methodology and linguistic analysis. The corpus treatment of bundles is then supplemented with an analysis of their structural forms and discourse functions. A total of 291 four-word bundles that recur at least 25 times per million words, and appear in 15% or more of the texts were retrieved and their structural and functional attributes examined. Results indicate that student-produced texts tend to use fewer and less varied lexical bundles than research articles and educational texts. The three text types also show considerable variation in the use of structural forms of recurrent bundles, with the student writing incorporating more phrasal patterns, while expert texts employing clausal constructions. Functionally, the three text types exhibit striking similarities, making greater use of research- and text-oriented bundles but producing a much narrower range of participant-based bundles.

Keywords
corpus linguistics, lexical bundles, mathematical discourse, register variation, text types

Introduction
Disciplinary writing involves the use of a wide range of semantically transparent and grammatically regular multi-word units that perform identifiable discourse functions and display analyzable structural patterns (Ädel & Erman, 2012; Biber et al., 2004; Chen & Baker, 2010; Cortes, 2004; Hyland, 2008a, 2008b; Pan et al., 2016). Such patterns have been analyzed under different terms, including academic clusters (Hyland, 2008a), phrasal expressions (Biber et al., 2004; Pan et al., 2016), academic formulas (Simpson-Vlach & Ellis, 2010), multi-word constructions (D. Liu, 2012), and recurrent word combinations (Ädel & Erman, 2012). The most widely used term, however, is that of lexical bundles which are referred to as the “uninterrupted strings of three or more words that frequently recur in a register, identified empirically by running a computer program in a corpus of language texts” (Cortes, 2015, p. 205).

Although it may be tempting to argue that mathematics has a language of its own, the prevailing belief is that mathematical discourse shares common ground with other disciplines in which language plays a central role in the construction, dissemination, and interpretation of disciplinary knowledge. However, mathematics differs in two fundamental ways. First, mathematical writing is semiotic and multidimensional, implying that mathematicians draw on non-linguistic means such as formulas, graphs, symbols, and diagrams while framing mathematical relations, operations, and processes (Huang & Normandia, 2007; O’Halloran, 2005; Schleppegrell, 2007). A second important aspect of mathematical discourse lies in its intricate network of grammatical patterns such as the use of highly technical lexis, familiar expressions with unfamiliar meanings, and discipline-specific rhetorical devices, all of which seem to flummox novice writers and non-specialists. Studies carried out in the field of mathematics education and English for Specific Purposes, however, concur that language is fundamental in understanding mathematical knowledge. As Graves et al. (2014) put it, “writing in mathematics is no different from other disciplines where a sophisticated awareness is required

1King Saud University, Riyadh, Saudi Arabia

Corresponding Author:
Abdullah A. Alasmary, Department of English Language & Translation, College of Languages & Translation, King Saud University, P.O.Box 12372, Riyadh 3968, Saudi Arabia.
Email: aasmary@ksu.edu.sa

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
of communicative acts typical within that discourse community” (p. 8).

Lexical bundles have been acknowledged as both an “important building block of discourse” (Biber & Barbieri, 2007, p. 263) and a key “component of fluent linguistic production” (Hyland, 2008b, p. 4). Yet, no previous study has attempted to compare the use of such patterns in student and expert mathematics writings. By using a combination of corpus tools and linguistic analysis, this study aims primarily to compare lexical bundles in three advanced mathematical text types: doctoral dissertations, journal articles, and textbook chapters. Another important objective underlying this study is pedagogic; that is, the structural and functional description of recurrent patterns in the three text types can be utilized in informing teaching practices, helping design learning materials, and creating English for Academic Purposes resources which can be used by non-English-speaking graduate math students and early-career mathematicians whose writings lack sophistication and complexity found in texts produced by experts. To fulfill these goals, a set of inferential statistical procedures are used to determine whether the differences between the writing groups are significant or not, following Ädel and Erman’s (2012) suggestion that “future research should consider augmenting the procedures used for bundle selection with more sophisticated inferential statistics” (p. 85).

The remaining parts of this paper are organized in the following way. First, a review is given of the previous research into lexical bundles, their properties and the different structural and functional attributes of such patterns. Second, methods of creating and analyzing corpora are discussed along with procedures to refine corpus-extracted data. The third section is intended to outline the results of the study whereas the fourth section is dedicated to discussing these results. The final section presents the conclusion of the study and the implications of the results on learning and teaching disciplinary writing.

**Overview of Research on Lexical Bundles**

Lexical bundles have been the focus of intense scholarly activity during the past two decades (Biber & Barbieri, 2007; Biber et al., 1999, 2004; Bychkovska & Lee, 2017; Cortes, 2004; Pan et al., 2016). A great deal of such research has thus explored variation in lexical bundles use either across different registers or within different groups of language users. Research by Biber et al. (1999, 2004) and Biber and Barbieri (2007) seem to concur that speech and writing show marked differences with respect to both the types as well as the functions of bundles, with the former using a far greater number of bundles than the latter. Other studies took a domain-focused approach, thus unveiling recurrent bundles in disciplines such as pharmacy (Grabowski, 2015), law (Breeze, 2013), telecommunications (Pan et al., 2016), and psychology (Esfandiari & Barbary, 2017). Findings reveal that each

of these disciplines employs a distinct set of bundles with identifiable structural forms and distinct functions which can be determined based on their meanings in the context. A related type of research on lexical bundles has examined the use of such patterns across different groups of writers. Comparisons as such are conducted from the perspective of university-level students with different first-language backgrounds (Ädel & Erman, 2012), students versus experts (Cortes, 2004), renowned native authors versus nonnative academics (Pan et al., 2016), nonnative-graduate students versus native-English-speaking experts (Qin, 2014), and native-English writers versus foreign students versus professionals (Chen & Baker, 2010). While there are some discrepancies in the results, most analyses have shown that native and professional writers demonstrate knowledge of a wide range of different bundles in comparison with second language learners and less experienced writers. Cortes (2004), for example, pointed out that some bundles used by professional authors are not found in texts produced by the students, a finding that lends further credence to the existence of a gap in the use of bundles between novice writers and professional authors. As for the structural comparison of patterns, most common bundles used by immature writers are largely verb-based whereas bundles commonly attributed to experts are noun-based, a pattern that interprets the lack of complexity commonly associated with expert writing (Chen & Baker, 2010).

A second aspect of research on lexical bundles involves applying a set of procedures while eliciting these patterns from language data. Three key criteria have thus been used, namely the length of the target bundle, frequency of occurrence, and distribution across the corpus sub-parts. In terms of length, most previous studies have extracted bundles comprising four words. Four-word bundles, C. Y. Liu and Chen (2020) point out, appear to be “pedagogically more useful due to their higher frequency, more complete syntactic structure, and a wider variety of functions they can perform” (p. 26). As with frequency of occurrence, there is no consensus among researchers on a specific frequency score beyond which a bundle is selected for analysis. Frequencies are conservatively adjusted at 40 times per million words in several studies (e.g., Biber & Barbieri, 2007; Bychkovska & Lee, 2017; Esfandiari & Barbary, 2017; Grabowski, 2015; J. Liu & Han, 2015; Pan et al., 2016). Some other researchers have used a relatively lower threshold at 20 times per million words (Cortes, 2004; Hyland, 2008a). A similar inconsistency is also noted in determining the dispersion of lexical bundles across the corpus subparts. For example, some researchers have included for analysis any lexical bundle occurring in at least 10% of texts in a corpus (Hyland, 2008b; C. Y. Liu & Chen, 2020; Shirazi zadeh & Amirfazlian, 2021). In other studies, the dispersion of a bundle is measured by setting a minimum number of two texts (Wood & Appel, 2014), three texts (Chen & Baker, 2010), and five texts (Cortes, 2004). The discrepancy in range, frequency, and
distribution values serves as an impetus for further research to create empirically-based criteria which help inform subsequent research in this area.

Functional categorization of bundles is a third major aspect of research on lexical bundles. An earlier attempt to classify bundles was carried out by Biber et al. (2004) who divided patterns into three major categories: referentials (e.g., that's one of the), stance expressions (e.g., I don't know if), and discourse organizers (e.g., let's have a look). Each category branches into subcategories. Another important classification scheme was suggested by Hyland (2008b) who, based on written academic corpora, classified lexical bundles into three major categories: research-oriented bundles, text-oriented bundles, and participant-oriented bundles. Although these two functional frameworks have been widely used as models for classifying bundles functionally, some researchers have introduced some modifications so as to account for bundles not nearly fitting either of the these main categories. Jablonkai (2010) and Breeze (2013), for example, added labels to include subject-specific bundles such as the principle of subsidiarity and the court of appeal.

Linguistic Analysis of Mathematical Discourse

The role that language plays in mathematical thinking and reasoning has captured the attention of both applied linguists and mathematics educators (e.g., Barton, 2008; Barwell, 2005; Leung, 2005; McGrath & Kuteeva, 2012; Morgan, 2005; Schleppegrell, 2007). Schleppegrell (2007) points out that the ubiquitous use of some technical vocabulary (e.g., the volume of a rectangular prism) poses a real challenge as some second language learners may confuse their mathematics-oriented meanings with what they already know about these words. Another layer of difficulty involves the use of some common grammar patterns (e.g., be and have) to highlight complex operational and relational processes in mathematics. In much the same vein, Chan (2015) demonstrated that expressions such as convert and walking speed are used differently in mathematics than in daily conversation. Some mathematical concepts denoting proportion and inference, Barton (2008) argues, carry distinct meanings in different languages, leading to confusion and misunderstanding, particularly in multilingual classrooms. Morgan’s (2005) analysis of mathematical definitions in academic and intermediate textbooks reveals that key mathematical concepts in lower-level texts appear “neither to reflect the way in which mathematical words and meanings are related in practice nor to provide any inkling of the powerful, and productive role that definitions can play in mathematics” (p. 115). The role of informal language in mathematical classroom is explored by Leung (2005) who argues that the use of informal language in some instances seems to facilitate the process of explaining core mathematical concepts. For example, the concept of factor pairs can be explained using simple, largely informal, language. Wagner and Herbel-Eisenmann (2008) analyzed the presence of the word just in a collection of classroom-based transcripts, and concluded that it serves to invite learners to participate in classroom discussions, given its collocational attributes such as just proceed, just do, and just use. In some cases, however, just appears to be used to shut down the conversation in cases such as just don’t. In a corpus-based study, Monaghan (1999) examined the collocational attributes of the word diagonal in a corpus of mathematics materials, and noted the diversity of its collocational attributes.

Grounded in Applied linguistics and English for Specific/Academic Purposes, several other studies have focused on the writing practices of members of the mathematics community. For example, McGrath and Kuteeva (2012) extracted stance and engagement markers from a corpus of journal articles in pure mathematics, and found that mathematician writers tend to employ a set of attitudinal patterns (e.g., it is easy to) and boosting devices (e.g., certainly), a conclusion that goes counter to the logic-driven, reasoning-based nature of mathematics. Investigating the introductory sections of published journal articles in mathematics, the study by Graves et al. (2014) noted that the existing model for analyzing move structure in research articles (RAs) needs to be revised, thus accounting for moves in mathematics articles which appear to demonstrate a different organizational and rhetorical model.

The study by Herbel-Eisenmann et al. (2010) is a pioneering attempt to investigate lexical bundles in a spoken corpus of math transcripts. The analysis of the corpus led to the retrieval of seventy-one bundles, nearly half of which had been identified as stance markers. Cunningham (2017) analyzed the use of discontinuous sequences with fillable slots, referred to as lexical frames, in a corpus of journal articles in mathematics, thus creating a functional taxonomy in which frames are divided into three distinct groups: aboutness (e.g., the proof of proposition), coherence (e.g., it follows from the), and moves (e.g., such that for every). Although these studies have deepened our understanding of different linguistic aspects of mathematical discourse, it is clear that this line of research needs to be extended to include other understudied genres, such as dissertations and textbooks, a gap for which this study attempts to bridge.

This study represents the first attempt to use lexical bundles as an analytical tool to explore variation in three key mathematical registers. To fulfill that goal, answers to the following questions are sought:

(a) What lexical bundles are commonly used across the three mathematical registers, namely doctoral dissertations, research articles, and educational textbooks?
(b) Which lexical bundles have the tendency to transcend register boundaries? Which bundles are register-specific?
(c) What structural and functional attributes do these bundles have?

Data and Methodology
Corpora

This study draws on three parallel corpora of written academic textbooks, peer-reviewed journal articles, and doctoral dissertations. These text types are sampled to represent the student, research and instructional discourses (Hyland, 2009). The first corpus, named TXT-BKs, includes a total of 43 textbook chapters aimed for graduate students in mathematics. There are three main criteria guiding the process of selecting textbooks: representativeness, diversity of topics and themes, and the target readership. As for representativeness, all candidate textbooks should be written by different authors, allowing for more diverse set of rhetorical and usage styles to be examined. Books chosen for analysis need also to be thematically diverse, covering different areas within mathematics. Finally, all books must be geared toward academic readership. The publication year of the textbooks is between 1999 and 2014, and the publishers are known for academically-oriented texts of worldwide reputation and readership. All authors are affiliated with academic or research institutions, and none of them is identified as a student. A wide range of topics are covered by these textbooks, including enumeration, commutative algebra, mathematical optimization, number theory, and integration theory. Given that lexical bundles are affected by the number of words in a corpus, rather than the number of texts (Cortes, 2004), creating a corpus of short chapters is more useful than of full-length textbooks.

The second corpus includes a total of 34 peer-reviewed articles that appeared in the period between 2013 and 2015. The selected articles address a wide range of mathematical topics such as algebra, geometry, approximation theory, and partial differential equation, just to name a few. Labeled JOL-ARs, the corpus is made of peer-reviewed articles drawn from four key journals: Acta Mathematica, Geometric and Functional Analysis, Communications on Pure and Applied Mathematics, and Inventions mathematicae. All these journals are sci-indexed (Web of Science) and are known for publishing high-quality papers on a range of mathematical topics. They also represent different internationally recognized publishers with a reputation of producing original research papers. To guard against journal issue or writer idiosyncrasies, a single article per issue and per author is included for analysis. All authors hold teaching positions at universities and research centers, and none of them is identified as a student.

The third corpus (DISS-ONs) is made up of 20 doctoral dissertations written by graduate students. The dissertations were all awarded by American, British, or Canadian universities, and were defended during the period 2005 to 2014. Topics covered in the dissertations include model theory, symmetrical algebra, quantum channels, eigenvalue problems, and nonparametric predictive influence. The dissertations were checked to make sure that no two dissertations are guided by a single academic advisor. Table 1 shows the number of texts in each corpus, the total count of all running words (tokens), and the distinct word types (types).

| Corpus                  | #texts | Tokens   | Types (distinct words) |
|-------------------------|--------|----------|------------------------|
| Dissertations (DISS-ONs)| 20     | 672,878  | 17,381                 |
| Journal articles (JOL-ARs)| 34    | 670,186  | 18,139                 |
| Textbooks sections (TXT-BKs)| 43  | 671,279  | 23,584                 |
| Total                   | 97     | 2,014,343| 59,104                 |

Bundle Identification Criteria

Length of the lexical bundle, frequency of its occurrence, and the distribution across the texts making up the corpus are the three criteria that determine the process of bundle selection and analysis. As for length, it is common practice in a broad range of previous studies to focus on four-word bundles. Cortes (2004) points out that a bundle of this length recurs more often than five- or six-word sequences, and tends to display more internal structural variation than three-word bundles. The second guiding principle is the frequency of occurrence. In the literature, frequency thresholds vary from one study to the other, ranging from 40 occurrences per million (Pan et al., 2016) to 10 per million (Biber et al., 1999). In this study, the decision is made to include bundles recurring 25 times per million words (Chen & Baker, 2010; Wood & Appel, 2014), and since the three corpora are similar in size, this cut-off threshold corresponds to a raw frequency of 17. A final key criterion concerns the degree to which lexical bundles are dispersed across the texts making up the corpus. Dispersion thresholds reported in the previous studies differed noticeably, with some studies preferring to select all bundles occurring in a minimum number of three (Wood & Appel, 2014) or five texts (Cortes, 2004) while others opting for percentages (Hyland, 2008b; C. Y. Liu & Chen, 2020). In this study and after piloting with the data, only bundles that recur in at least 15% of the texts in each corpus were selected for analysis. This moderate percentage is chosen to allow for more patterns to be analyzed while ensuring that they do not
reflect an idiosyncratic use typical to a specific text or author. After applying the aforementioned criteria, lexical bundles are chosen for analysis if they are composed of four words, recur 25 times or more in a million word, and occur in at least 15% of texts making up each corpus.

**Data Refinement**

A number of procedures were thus adopted to refine the data and reduce the number of sequences that are problematic. Given the semiotic nature of mathematics, the retrieval of some target bundles yielded patterns with excessive characters, numbers, or symbols (e.g., $x y$ is a, CID CID CID CID) which are of little interest to the linguistic analysis (McGrath & Kuteeva, 2012). To address this problem, sequences that are heavily composed of symbolic notations and figures were eliminated from the final list of bundles. A second refinement step addresses overlapping sequences. Chen and Baker (2010, p. 33) identified two cases: “complete overlap” and “complete subsumption.” The first type occurs when two four-word sequences with the same frequency and range are derived from a single five-word sequence. For instance, the bundles *it is not difficult* and *is not difficult to* were completely matched in frequency and dispersion, and were elicited from the longer sequence *it is not difficult to*. The second type involves two sequences or more in which “the occurrences of one of the bundles subsume those of the other overlapping bundle(s)” (Chen & Baker, 2010, p. 33). The expression *if and only if*, for example, occurred 245 times in the journal article corpus, while the sequence *and only if* it occurred much less frequently at 33 times per million words in the same corpus. When combined together, they form a five-word expression *If and only if* it. To address both types, sequences were merged into a single extended unit in which additional, less frequent lexical items were enclosed in brackets (e.g., *if and only if* $i + (ii)$).

To carry out more meaningful comparisons across corpora, there is an emphasis in the literature to make a distinction between the number of bundles (types) retrieved from a specific corpus, and their frequencies of occurrence (tokens; Adel & Erman, 2012; Chen & Baker, 2010). Both the bundle types and tokens were compared and contrasted in this study.

**Analytical Procedures**

WordSmith Tools 7.0 (Scott, 2016) was used to extract all four-word sequences occurring at least 17 times, a raw frequency score which was later normed to 25 occurrences per million words. Dividing bundles into functional categories was carried out according to Hyland’s (2008b) classification scheme. Bundles are thus divided into research-, text, and participant-based. Research-oriented bundles serve to help authors “structure their activities and experiences of the real world,” whereas text-oriented expressions are mainly “concerned with the organization of the text and its meaning as a message or argument” (Hyland, 2008b, p. 13). The research-oriented group consists of five subgroups, namely location, procedure, quantification, description, and topic. Bundles in the text-oriented group serve as transition signals, resultative markers, structuring devices, and framing indicators. Participant-oriented bundles, the third functional group, fall into two broad categories: stance expressions which serve to “convey the writers’ attitudes” and engagement expressions which are used to “address the readers directly” (Hyland, 2008b, p. 13).

To carry out the functional analysis of bundles, three informants were asked to do the classification based on Hyland’s (2008b) framework. The first informant is a university professor who has published papers, books, and book chapters on various mathematical issues. The second informant is an academic writing specialist who has been teaching English to non-English majors for over 10 years. The third informant is a university professor with experience in research on academic writing and register analysis. An initial analysis of functions reported a 70% agreement between raters. For example, one rater assigns the bundle *then there exists a* into framing within text-oriented bundle category while the second informant labels it as a text-oriented bundle signaling *transition*. The third informant ascertains that this bundle is a mathematically oriented expression which needs to be functionally grouped under a newly created subcategory of *existence* belonging to research-oriented category. Discrepancies as such were resolved by the informants collectively checking the concordance lines of the disputed bundles in order to find out what functions they serve. The final functional categorization is based on a 100% agreement between informants.

**Results**

**Overlapping Versus Register-Specific Bundles**

Arranged according to their normalized frequencies, the lists of the recurrent word combinations used in dissertations (DISS-ONs), research journal articles (JOL-ARs), and educational textbooks (TXT-BKs) are presented in the Appendix. Journal authors seemed to use the greatest number of bundle types (115), followed by textbook writers (113) and graduate students (63). If tokens are considered, textbook authors produced the largest number of bundle tokens (6,423), followed by article authors (6,287) and graduate students (3,358). Table 2 shows a total of 29 bundles which appear across the three text types, accounting for 46% of the overall number of bundles in DISS-ONs, 25% in JOL-ARs, and 26% in TXT-BKs. Dissertations exclusively share two bundles (*the boundary of the, it is possible to*) with journal articles and four bundles (*from the definition of, can be found in, can be used to, is the same as*) with textbooks. About 28 bundles (44%) in the DISS-ONs are register-specific, that is, they do not appear in both the journal articles and textbooks.
Professional mathematics texts, represented by JOL-ARs and TXT-BKs, are found to share a large number of bundles (31 strings), representing a proportion of 27% each. A total of 55 academic sequences, amounting to 48% of JOL-ARs bundles, do not appear in DISS-ONs or TXT-BKs. TXT-BKs includes 49 register-specific bundles, accounting for nearly half (43%) of the total bundles in the corpus.

**Structural Forms of LBs**

Contrasting the structural forms of the lexical bundles used in each register is a key purpose of the study. Dissertations (DISS-ONs) are marked by the use of more nominal and prepositional constructions and fewer verbal patterns. In contrast, textbooks and research articles seem to make more use of verb-based constructions, but tend to use comparably fewer noun and preposition-based sequences (see Figure 1). Within the verb-based bundle category, DISS-ONs include a greater number of structures comprising passive forms, copula-be + noun/adjective phrases, and verb-phrase followed by that-clause. Similarly, though to a lesser extent, three forms, namely anticipatory-it, existential-there, and active verb, are also found in the dissertation corpus. Journal authors seem to make a greater use of constructions incorporating active verbs, copula-be constructions, and verb-that clause fragments, but are less inclined to use expressions in the passive voice. In TXT-BKs, clusters involving copula-be, verb+that clause, active-voice verbs, and existential-there are prioritized while passive-voice patterns and sequences headed by anticipatory-it are underrepresented. With respect to nominal and prepositional phrase bundles, nearly one-third of the bundles used in DISS-ONs incorporate a noun phrase, the majority of which are complemented by an of-phrase fragment. Lexical bundles headed by a preposition account for one-fifth of all bundle occurrences in DISS-ONs. There is a close match in the distribution of nominal bundles in JOR-ARs and TXT-BKs, showing a tendency to use the same number of bundle types. Preposition-headed bundles appear to be used more in journal articles than in textbooks.

A final group of bundles that do not fall neatly into either of the preceding structural categories includes a set of extremely frequent, widely distributed, and clearly discipline-specific bundles such as if and only if, such that for all, and independent of the choice. The use of such patterns varies from one corpus to the other, with TXT-BKs comprising

---

**Table 2. Register-Transcending Bundles.**

| Lexical bundle                                                                 | DISS-ONs | JOL-ARs | TXT-BKs |
|--------------------------------------------------------------------------------|----------|---------|---------|
| if and only if                                                                  |          |         |         |
| on the other hand                                                               |          |         |         |
| it is clear that                                                                 |          |         |         |
| is said to be                                                                   |          |         |         |
| then there exists a                                                            |          |         |         |
| can be written as                                                                |          |         |         |
| there exists a unique                                                           |          |         |         |
| in this section we                                                              |          |         |         |
| the set of all                                                                  |          |         |         |
| with respect to the                                                             |          |         |         |
| let (a) be a                                                                    |          |         |         |
| is the set of                                                                   |          |         |         |
| it is easy to                                                                   |          |         |         |
| we have the following                                                           |          |         |         |
| the proof of theorem                                                            |          |         |         |
| the proof of the                                                                |          |         |         |
| to show that                                                                    |          |         |         |
| in the case of                                                                  |          |         |         |
| without loss of generality                                                      |          |         |         |
| in terms of the                                                                 |          |         |         |
| is the number of the existence of a                                             |          |         |         |
| as a function of                                                                |          |         |         |
| it follows from the                                                             |          |         |         |
| such that for all                                                                |          |         |         |
| completes the proof of                                                         |          |         |         |
| a special case of                                                                |          |         |         |
| there is a unique                                                              |          |         |         |
| this completes the proof                                                        |          |         |         |

---

**Figure 1.** Comparison of structural types of bundles in the three corpora (%).
the greatest number of these fragments, followed by JOL-ARs and DISS-ONs. Table 3 shows the structural distribution of bundles across the three text types.

Table 4 shows the results of a chi-square test which demonstrates that there is a significant difference in the structural distribution of lexical bundle tokens between DISS-ONs, JOL-ARs, and TXT-BKs. The Standardized Residuals (R) illustrate that DISS-ONS writers use more NP-based and PP-based bundles, and fewer VP-bundle tokens than expected, whereas journal authors use more prepositional phrase bundle tokens than expected. The picture is different when it comes to textbook authors, who rely on verb phrase bundles more significantly than on NP- or PP-bundle tokens. Another chi-square test was administered, indicating that there is no significant difference in terms of the structural distribution of bundle types across the three mathematical corpora ($\chi^2 = 12.33, df=6, p > .05$, Cramer’s $V=0.1456$).

**Functional Distribution**

Another aim of the present study is to give an account of the functions served by the set of bundles emerging from the quantitative analysis. As illustrated in Table 5, bundles are classified into three major functional categories, each of which includes a range of different sub-functions. It is clear that the largest group of bundles across the three corpora serve a research-based function. The second largest group of bundles includes patterns that fulfill a text-structuring
Bundles performing a participant-oriented function represent the smallest functional group. Table 6 shows that research-based bundles dominate DISS-ONs and TXT-BKs, displaying the same proportions of tokens (62%). With respect to JOL-ARs, bundles serving a research-based function make up the most recurrent functional category, though at a relatively lower proportion of 44% (tokens). The second most common functional type involves text-structuring devices which serve to organize the text, connect its parts, or draw the reader’s attention to certain elements appearing elsewhere within the text. Participant-based bundles rank as the most infrequently used functional type across the three corpora, with JOL-ARs using the greatest number of these bundles (22%), followed by TXT-BKs (18%) and DISS-ONs (13%). Figure 2 shows an overall comparison of bundle functions in student and expert mathematical writings.

A chi-square test shows that there are significant differences in the distribution of bundle types between the three corpora. Table 7 shows that no residuals ($R$) are greater than ±1.96, implying that the three groups do not exhibit much difference in the functional types they use.

A similar chi-square test was administered to determine whether the functional differences in the distribution of bundle tokens are statistically significant. The results show a significant difference in the distribution of functions served by the bundles within the three text types, as is illustrated in Table 8.

**Research-oriented bundles.** A close examination of the distribution of the functional types reveals a large concentration of research-oriented bundles in the three mathematical texts. These research bundles make up two-thirds of all sequences in the student and textbook corpora, and almost half of all sequences in the journal article corpus. A substantial portion of these sequences are topic-focused, thus referring to mathematical notions, relations, or processes. The most frequently used bundle in our data is the bi-conditional $\text{if and only if}$, which is inadequately labeled as a stance marker by Wood and Appel (2014). It is a key mathematical pattern used to introduce two equivalent conditions in which the truth/falsity of the first is necessary and sufficient for the truth/falsity of the second. Other bundles in this group consist of nominal

| Functional category    | Sub-function  | DISS-Ns Type | DISS-Ns Token | JOR-ARTs Type | JOR-ARTs Token | TXT-BKs Type | TXT-BKs Token |
|------------------------|---------------|--------------|---------------|---------------|---------------|--------------|---------------|
| Research-oriented      | Topic-related | 14           | 996           | 11            | 1,109         | 15           | 1,592         |
|                        | Procedures    | 10           | 393           | 10            | 523           | 14           | 689           |
|                        | Quantification| 6            | 278           | 4             | 167           | 13           | 626           |
|                        | Existence     | 5            | 263           | 10            | 452           | 14           | 694           |
|                        | Description   | 3            | 99            | 11            | 400           | 12           | 355           |
|                        | Location      | 1            | 44            | 3             | 101           | 1            | 40            |
| Total                  |               | 39           | 2,073         | 49            | 2,752         | 69           | 3,996         |
| Text-oriented          | Structuring   | 7            | 310           | 9             | 443           | 11           | 444           |
|                        | Transition    | 3            | 327           | 6             | 497           | 5            | 462           |
|                        | Framing       | 3            | 138           | 12            | 801           | 7            | 362           |
|                        | Resultative   | 2            | 57            | 11            | 406           | 1            | 31            |
| Total                  |               | 15           | 832           | 38            | 2,147         | 24           | 1,299         |
| Participant            | Stance        | 5            | 275           | 18            | 911           | 11           | 702           |
|                        | Engagement    | 4            | 181           | 10            | 477           | 9            | 429           |
| Total                  |               | 9            | 453           | 28            | 1,388         | 20           | 1,131         |

| Functional types       | DISS-Ns % (type) | DISS-Ns % (token) | JOR-ARTs % (type) | JOR-ARTs % (token) | TXT-BKs % (type) | TXT-BKs % (token) |
|------------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Research               | 60 (39)          | 62 (2,073)       | 43 (49)          | 44 (2,752)       | 61 (69)          | 62 (3,996)       |
| Text                   | 26 (15)          | 25 (832)         | 33 (38)          | 34 (2,147)       | 21 (24)          | 20 (1,299)       |
| Participant            | 14 (9)           | 13 (453)         | 24 (28)          | 22 (1,388)       | 18 (20)          | 18 (1,131)       |
| Total                  | 100 (63)         | 100 (3,358)      | 100 (115)        | 100 (6,287)      | 100 (113)        | 100 (6,423)      |
and prepositional patterns (e.g., the eigenvalues of the). The pervasive presence of these semantically packed and tightly structured patterns is unsurprising, given that mathematical discourse is, according to O’Halloran (2005), “technical and often involves complex taxonomies of terms in nominalized forms” (p. 78). Here are examples of topic-focused, research-oriented bundles:

1. A function \( f \) is an indefinite integral if and only if it is absolutely continuous. (DISS-ONs)
2. In principle, the proof of theorem is completely constructive. (TXT-BKs)

The next most common research-oriented sub-category is a group of bundles that help explain a mathematical procedure, such as the announcement of proof completion, the presentation of mathematical assumptions, or the delivery of the proof of a previously stated problem, axiom, or theorem.

3. This completes the proof of part (A) in the conclusion. (DISS-ONs)
4. Let \( X \) be a projective variety of dimension \( n \) over \( \mathbb{C} \) equipped with a very ample line bundle \( L \). (JOR-ARs)

Within the research-oriented group of bundles, also, we find a number of bundles that serve to label an entity as unique or special. Examples include constructions such as there is a unique, the existence of the, there exists a unique, the uniqueness of the, and then there exists a. The distribution of the existence/uniqueness bundles varies across the three corpora, with JOR-ARs and TXT-BKs incorporating a substantial
number of these expressions, whereas DISS-ONs use comparatively fewer.

5. Hence, from Corollary 1.5.4 there exists a unique semi-group morphism. (DISS-ONs)
6. If $J$ is a vector field on $H$, then there is a unique extension to a Jacobi field on $\Omega = (a, b) \times H$. (TXT-BKs)

Quantification bundles, a sub-category of the research-based group, are more prevalent in textbooks than in journal articles or dissertations. Bundles in this group provide frames for key mathematical concepts such as set, number, and sum.

7. $V$ is the set of vertices. (DISS-ONs)
8. $n$ is the number of pixels. (JOR-ARs)

The fifth bundle type within the research-oriented category incorporates a variety of sequences used to describe a mathematical attribute. Both the TXT-BKs and JOR-ARs contain more than twice the number of description bundles in DISS-ONs. Examples include the structure of the, the closure of the, and is an extension of.

9. The reader can deduce the structure of the amalgamated free products in both of these categories. (DISS-ONs)
10. $G$ is an extension of a cyclic group of order. (TXT-BKs)

The least used bundle type in the research-oriented group tends to function as a location marker. While these bundles make up the smallest category in TXT-BKs and JOR-ARs, no bundle of this type occurs in DISS-ONs corpus. Examples of location markers include sequences such as the boundary of the and in a neighborhood of:

11. Our work leaves open the question of whether the critical set coincides with the boundary of the sub-critical regime as well. (JOR-ARs)
12. If $B$ was analytic at a point on $S1$, $B$ would have to vanish at that point and hence be identically zero in a neighborhood of it, which is clearly not the case. (TXT-BKs)

**Text-oriented bundles.** The second largest functional group of sequences identified in this study serves to organize the flow of arguments, ideas, or propositions in the text. It is unsurprising to find that structuring bundles are the most employed functional type in DISS-ONs and TXT-BKs, given the extended nature of these texts and the need for their parts to be coherently and logically linked. Journal authors, in contrast, are constantly under pressure to ration language and present new ideas in the most condensed and economical manner. Examples of structuring patterns include as in the previous and at the end of:

13. The top piece is mapped as in the previous case. (JOR-ARs)
14. Another elegant proof of Sylvester’s theorem is outlined in Exercise 12 at the end of the chapter. (TXT-BKs)

Framing bundles are distributed inconsistently in the three text types, with expert mathematicians making use of a greater number of these bundles in comparison with the students.

15. A Hilbert space $H$ is an inner product space which is completed with respect to the norm defined above. (DISS-ONs)
16. In the case of $B$, the neighboring trap surface is smoother. (JOR-ARs)
The last two text-oriented functional types are transition and resultative signals. There is a substantial number of transition signals in JOL-ARs in comparison with TXT-BKs and DISS-ONs. These are mostly used to show similarities and differences between items (e.g., on the other hand, is equal to the, is the same as):

17. On the other hand, implementing the Zimmermann-Mertins method requires finding the eigenvalues. (DISS-ONs)
18. Recall that a factorial ring is the same as a unique factorization domain. (TXT-BKs)

While rarely used in TXT-BKs and DISS-ONs, resultative signals constitute the second most commonly used functional pattern in journal articles. Examples of this type include sequences such as is due to the, the main result of, and as a function of.

19. This is due to the complexity of these problems, in particular, those involving large sizes. (DISS-ONs)
20. We are now equipped to state and prove the main result of this section. (JOR-ALs)

**Participant-oriented bundles.** The participant-oriented group of bundles, the third and final functional category, consists of stance markers, which help convey the author’s views and opinions, and engagement features, which bring the reader into the discussion. Mathematical arguments are logic-driven and evidence-based, implying that the authorial voice is sometimes not commonly present. It is clear from the data that the JOR-ARs has the highest number of participant-oriented bundles in comparison with TXT-BKs and DISS-ONs.

21. It is clear that the best choice of method depends on the specific problem to be solved. (DISS-ONs)
22. Thus, to prove that it is an isomorphism, it is sufficient to show that its kernel is 0. (TXT-BKs)

Turning now to engagement bundles, we find that almost all constructions are initiated by the first-person plural pronoun we. One informant commenting on the overuse of the plural we explained that in mathematics, knowledge is viewed as collectively built and the use of the first singular pronoun I may be readily interpreted as arrogance on the part of the author.

23. We find that the region of stability for all schemes includes the negative y-axis. (DISS-ONs)
24. We see that the homology in the middle must have rank 0. (TXT-BKs)

**Discussion**

This study explores the use of four-word combinations in doctoral dissertations, peer-reviewed journal articles, and academic textbooks in the domain of mathematics. The results demonstrate that the students’ register incorporates the smallest set of lexical bundles in comparison with journal articles and textbooks. The paucity of recurrent academic clusters in the writings of student mathematicians seems to confirm a result in some previous research, suggesting that student writers tap into a restricted set of highly frequent and less varied multiword strings (Adel & Erman, 2012; Bychkovska & Lee, 2017; Chen & Baker, 2010; Cortes, 2004). The lower number of bundles in the writings of students in this study appears to be congruent with the results reported by some researchers (e.g., Hyland, 2008a; Pan et al., 2016; Qin, 2014) who found that non-native English students produced the greatest number of bundles compared with expert authors. Variation in the results may be interpreted by the sensitivity of bundles distribution to several factors, the most important of which is the target discipline (Shirazzadeh & Amirfazlian, 2021).

Another purpose of this study is to determine whether the recurrent bundles are register-specific or tend to transcend register boundaries. In this study, the proportion of bundles that student-produced texts share with the other two text types is congruent with a similar proportion reported by Hyland (2008a), who found that half of the academic bundles extracted from the student registers were found in a corpus of journal articles. The high proportion of shared bundles found in the dissertation corpus may be due to the widely held assumption that students lack register awareness and are therefore more likely to revert to common, risk-free “lexical teddy bears” (Ellis, 2012). Expert mathematicians, in contrast, appear to recognize the importance of such patterns as “containers” of mathematical meanings and feel confident in using them quite frequently and abundantly.

An important result emerging from the data is that a great number of bundles uniquely occurring in dissertations reflect a general language use, not a discipline-specific one. Bundles such as the rest of the, one of the most, the use of the, is due to the and is based on the are reported in other studies on similar registers such as student writings (Adel & Erman, 2012; Bychkovska & Lee, 2017; Chen & Baker, 2010) and research articles (Pan et al., 2016). The high frequency of such bundles in the DISS-ONs may reveal that the student writers have not yet attained an expert-like mastery of the routinized conventions typical of expert-like writings. In contrast, the bundles exclusively used in JOR-Ars or TXT-BKs are indicative of both deep register awareness and broad subject-matter knowledge.

Structural comparisons reveal that each register prioritizes certain forms in a way that is distinct from the two others. Texts produced by students are dominated by nominal and prepositional phrases, a pattern typical of advanced writing (Biber et al., 2004; D. Liu, 2012; Pan et al., 2016). Authors of mathematical textbooks use bundles incorporating more verb-based constructions whereas journal articles writers appear to prefer the same construction, though to a lesser extent. Some researchers have noticed the tendency for book authors to rely less on
The three mathematical text types exhibit clear differences in their uses of certain patterns. The presence of more passive constructions in the dissertations seems to be characteristic of student science writings, as pointed out by Hyland (2008b, p. 10), who found that biology and electrical engineering corpora are replete with passive bundles in comparison with the humanities. Hyland (2008b) maintains that the impetus behind the use of passive-voice forms by hard-science writers is “to downplay the personal role of the scientist in the interpretation of data and to suggest that the results would be the same whoever conducted the research” (p. 11). A second reason for the pervasive use of passive constructions relates to the semiotic nature of mathematical discourse, which requires that the reader’s attention be implicitly drawn to the information contained in graphs, tables, diagrams, and figures accompanying the text. It should be noted that expert mathematics writers of journal articles and textbooks use passive forms significantly less frequently than student mathematicians. The scarcity of passivized forms in the expert writings in this study dovetails with findings noted by Cunningham (2017) whose final list of phrase frames does not include any passive constructions. Taken together, it seems clear the less use of such pattern is a robust indicator of mature academic writing in mathematics.

Another area where the three text types exhibit clear differences lies in the use of recurrent sequences with Copula-be, a pattern which is more prevalent in textbooks and journal articles than in dissertations. This is unsurprising, given the experts’ awareness of its function as “a nexus between a linguistic and symbolic representation in mathematics” (Veel, 1999, p. 196). The relatively infrequent use of this pattern in the dissertations is a warning sign that the student writers may have not yet developed a thorough understanding of how Copula-be forms are used in specialized texts. A similarly important tendency in the student data is the repeated use of what Ådel and Erman (2012) refer to as “relatively informal lexical choices” (p. 86). Examples include clear, easy, and possible, which are found in expressions such as it is clear that, it is easy to, and it is possible that. Expert mathematics writers, in contrast, make use of such common terms but also demonstrate awareness of and familiarity with a range of other alternative lexical fillers such as sufficient, natural, enough, and convenient. The overuse of rather common terms by student mathematicians is probably motivated by the need to facilitate the process of understanding core mathematical complexities (Leung, 2005).

The distribution of functional types is similar across the three groups of texts, with bundles fulfilling a research-oriented function ranking as the largest category in each corpus, followed by bundles serving a text-organizing function and, finally, bundles performing a participant-oriented function. This result dovetails with similar findings reported by Ådel and Erman (2012), Biber and Barbieri (2007), and Biber et al. (2004) who concluded that referential bundles, which correspond to research-oriented expressions, represent the largest functional category used in the contexts they investigated. The existence of a large number of research-oriented bundles reflects the specialized nature of mathematical discourse, which necessitates the use of a wide range of expressions to define objects, describe logical processes, reorder patterns, signal the uniqueness of entities, solve problems, and announce results. The tendency for mathematics writers to rely more on research-oriented bundles may also echo a “scientific ideology” which “emphasizes the empirical over the interpretive, minimizing the presence of the researchers and contributing to the strong claims of the sciences” (Hyland, 2008b, p. 15). The next most common functional type involves text-structuring devices, a set of bundles serving to organize the content and make it a coherent whole. In some previous studies, text-structuring bundles were ranked as the most frequently used functional type (Hyland, 2008b; Pan et al., 2016). Participant-oriented bundles, which are sometimes referred to as stance markers, are the least used functional category. The highest percentage of bundles performing a participant-oriented function is found in JOL-ARs, a result that lends credence to O’Halloran’s (2005) view that “the authors of research papers in mathematics cast a favorable impression on the results which are established” (p. 74). It is interesting to note that, unlike research or text bundles, participant markers are not strictly of a mathematical nature, an observation similar to one pointed out by Herbel-Eisenmann et al. (2010), who concluded that stance sequences elicited from a corpus of classroom interactions are not exclusively tied to mathematical content.

To conclude this part, some other important tendencies in the data merit final discussion. Expert writers make much use of quantification bundles containing some key technical concepts in comparison with graduate students whose use of such bundles is still underdeveloped. The comparatively less use of such bundles by student writers may be due to the technical nature of some expressions contained in these bundles which pose a challenge to the students for whom English is a second foreign language (Schleppegrell, 2007). In a similar way, writings produced by experts feature a range of bundles comprising common expressions which carry different meanings in mathematics than in everyday conversation. Concepts such as closure and extension, which are found in bundles the closure of the and is an extension of, serve ostensibly mathematical meanings that need to be distinguished from their non-technical use. The underuse of bundles containing common words with unfamiliar meanings is among the challenges that face foreign student mathematicians as pointed out by Chan (2015).

**Conclusion and Pedagogical Implications**

The current study has examined the use of four-word lexical bundles across three parallel corpora of PhD dissertations, journal articles, and textbooks chapters in the domain of mathematics. While dissertations are representative of student registers, journal articles, and textbooks reflect expert mathematicians. Results demonstrate that Bundles gleaned
from the three corpora are then grouped into distinct structural and functional categories with the aim of deepening the students’ understanding of expert-authorized norms and styles of writing advanced texts in mathematics discourse.

These results hold several pedagogical implications for the teaching and learning of lexical bundles in mathematics. To cultivate greater register awareness among novice mathematicians, instructional intervention should focus not only on the forms of bundles but also on their discourse functions. These functions can be instructionally fostered using labels discussed in this study or those alluded to by Cunningham (2017) and McGrath and Kuteeva (2012). The functional labels can be supplemented by examples derived from a specialized corpus through concordance lines. Excerpts from disciplinary textbooks and journal articles should also be incorporated into advanced ESP and EAP programs as writing models to emulate, for the obvious reason that these texts are written by academically accomplished and profoundly experienced members whose publications have undergone extensive revising and peer-reviewing. A register-based instruction targeting recurrent bundles could also help novice mathematician writers accelerate their transition from a student-styled use of language to a more expert-styled one, a shift highlighted by Schleppegrell (2007) as posing a real challenge to many aspiring novice mathematicians.

Although the study has adhered to principles of corpus methodology and linguistic analysis, it is important to allude to two potential limitations. While this study could not give a full account of other influencing factors such as the writing experience of the expert groups or the situational characteristics of texts as distinct registers, some researchers may find these variables essential for interpreting tendencies in the data. In this study, interpreting variation in bundle use is informed by novice versus expert dichotomy. Book chapters and journal articles serve distinct communicative purposes and display different structural and functional characteristics. However, results show that they share nearly the exact number of bundle “tokens” and “types.” In much the same vein, dissertations and book chapters share a common ground; that is, their authors are not under “space pressure” to encode meanings using densely packed language similar to what journal articles authors do. Though they are similar in production circumstances, both dissertations and book chapters employ a set of patterns that are different in number as well as in forms and functions. So examining these patterns from a strict register-based perspective may overlook an important fact which demonstrates that these three texts are produced by two different groups of writers: novices and experts. To account for factors influencing bundle use, this study has taken a different approach, thus examining authors’ choices and patterns of use while alluding to places where we find that differences are attributed to register variation. Yet another limitation concerns relying solely on corpus data while explaining the differences in language use. Conducting structured interviews with writers would have clarified many tendencies in the data. Despite these limitations, the present study brings to light a careful discussion of recurrent lexical bundles in mathematics, a field of study that has remained underexplored.

The present study has shown that mathematical discourse, though semiotic and highly symbolic, can be investigated using a combination of corpus methodology and linguistic analysis. It is hoped that future research will build on the current study and expand the horizon by looking into other mathematical registers and/or writing groups. Findings gleaned from such studies will be indispensable for designing ESP material, improving current pedagogical practices, and creating discipline-based writing programs that cater for the needs of a growing number of novice students and researchers in the domain of mathematics.

### Appendix

| Sequences in ART-ALs | Sequences in TXT-BKs | Sequences in DISS-ONs |
|----------------------|----------------------|----------------------|
| 1. if and only if (a) (52) | 1. if and only if (a) (82) | if and only if (a) (100) |
| 2. the proof of theorem (239) | 2. on the other hand (214) | on the other hand (273) |
| 3. it is easy to + (check) (198) | 3. it is clear that (152) | from the definition of (203) |
| 4. with respect to the (186) | 4. the set of all (150) | it is clear that (132) |
| 5. (as) + in the proof of (179) | 5. let (x) be a (144) | is said to be (98) |
| 6. we may assume that (167) | 6. it is easy to + (check that) (132) | then there exists a (82) |
| 7. on the other hand (161) | 7. (as) + in the proof of (115) | can be written as (82) |
| 8. the proof of the + (lemma) (158) | 8. we may assume that (110) | there exists a unique (82) |
| 9. (is) + easy to see that (125) | 9. is said to be (106) | in this section we (79) |
| 10. in the case of (107) | 10. the proof of the (104) | the set of all (76) |
| 11. in this section we (101) | 11. an open subset of (101) | with respect to the (68) |
| 12. in terms of the (95) | 12. that there is a (99) | let (a) be a (62) |
| 13. let (x) be a (89) | 13. the proof of theorem (95) | the eigenvalues of (61) |
| 14. this completes the proof (86) | 14. with respect to the (94) | is the set of (58) |
| 15. we have the following (86) | 15. there exists a (89) | it is easy to (53) |
| 16. in the sense of (83) | 16. let (x) and (y) + (be) (79) | can be found in (52) |

(continued)
17. the proof of lemma (83) 17. in this section we (77) 17. we have the following (49)
18. the set of all (82) 18. in the sense that (76) 18. the proof of theorem (47)
19. the proof of proposition (80) 19. (is) easy to see that (74) 19. the length of the (46)
20. in the sense that (76) 20. the fundamental theorem of (65) 20. the proof of (46)
21. of the proof of (73) 21. then there is a (64) 21. the structure of the (46)
22. without loss of generality + (that) (70) 22. is the set of (61) 22. to show that the (46)
23. then there exists a (67) 23. (it) suffices to show that (61) eigenvalues of the matrix (44)
24. completes the proof of (67) 24. that there exists a (61) the boundary of the (44)
25. be the set of (61) 25. this completes the proof (59) in the case of (43)
26. with the property that (61) 26. a finite number of (57) 26. it is enough to (54)
27. is equal to the (58) 27. to show that the (57) in this chapter we (37)
28. to show that the (58) 28. there is a unique (55) without loss of generality (36)
29. the existence of a (57) 29. if there exists a (54) be a Hilbert space (34)
30. the fact that the (57) 30. in terms of the (54) can be considered as (34)
31. that there exists a (55) 31. the fact that the (54) can be used to (34)
32. it is clear that (58) 32. with the property that (52) in terms of the (33)
33. it is enough to (54) 33. be the set of (51) is the number of (33)
34. the main result of (51) 34. is the same as (51) the existence of a (33)
35. in the case when (49) 35. we have the following (51) we find that the (31)
36. is given by the (49) 36. can be written as (49) as a function of (30)
37. is independent of the (49) 37. is equal to the (49) in the following theorem (30)
38. there exists a constant (49) 38. completes the proof of (48) it follows from the (30)
39. we can assume that (49) 39. it follows that the (48) the use of the (30)
40. we say that a (49) 40. there exists a unique (45) it is known that (28)
41. that there is a (48) 41. a vector space over (43) one of the most (28)
42. then there is a (48) 42. and the fact that (43) such that for all (28)
43. is contained in the (46) 43. proof of the theorem (43) the analysis of the (28)
44. such that for any (46) 44. show that there is (43) upper bound for the (28)
45. does not depend on (45) 45. the following are equivalent (43) completes the proof of (27)
46. in a neighborhood of (45) 46. from the definition of (42) in a similar way (27)
47. it is sufficient to (45) 47. is equivalent to the (42) is based on the (27)
48. the case of the (45) 48. it follows from the (42) is due to the (27)
49. is a consequence of (42) 49. show that there exists (42) is the same as (27)
50. is the union of (42) 50. in a neighborhood of (40) of the differential equation(27)
51. it is natural to (42) 51. (less) + than or equal to (40) we can show that (27)
52. (it) suffices to show that (42) 52. the existence of a (40) a special case of (25)
53. the image of the (42) 53. without loss of generality (40) Adjoint operator on a (25)
54. can be written as (40) 54. an example of a (39) an inner product space (25)
55. it follows that the (40) 55. does not depend on (39) it is possible to (25)
56. the closure of the (40) 56. is an automorphism of (39) lower and upper bounds (25)
57. as a function of (39) 57. is a basis for (37) lower bound of the (25)
58. on the choice of (39) 58. we see that the (37) the eigenvalues of a (25)
59. the case of a (39) 59. we will show that (37) the order of the (25)
60. the union of the (39) 60. at least one of (36) there is a unique (25)
61. in view of the (37) 61. can be used to (36) this completes the proof (25)
62. it follows from the (37) 62. is a set of (36) we note that the (25)
63. not depend on the (37) 63. is a vector space (36)
64. we see that the (37) 64. is an isomorphism of (36)
65. the restriction of the (34) 65. is the number of (36)
66. in the definition of (36) 66. the definition of the (36) (continued)
Appendix. (continued)

| Sequences in ART-ALs | Sequences in TXT-BKs | Sequences in DISS-ONs |
|----------------------|----------------------|----------------------|
| 75. and the fact that (33) | 75. is an extension of (33) | 75. is an extension of (33) |
| 76. in such a way + (that) (33) | 76. such that for all (33) | 76. such that for all (33) |
| 77. in the case where (33) | 77. then there is an (33) | 77. then there is an (33) |
| 78. by the definition of (31) | 78. a special case of (31) | 78. a special case of (31) |
| 79. is of the form (31) | 79. as a function of (31) | 79. as a function of (31) |
| 80. is said to be (31) | 80. is given by the (31) | 80. is given by the (31) |
| 81. it suffices to prove (31) | 81. is one of the (31) | 81. is one of the (31) |
| 82. such that for all (31) | 82. there exists a constant (31) | 82. there exists a constant (31) |
| 83. in the same way (30) | 83. this follows from the (31) | 83. this follows from the (31) |
| 84. is the number of (30) | 84. can be extended to (30) | 84. can be extended to (30) |
| 85. it turns out that (30) | 85. is an element of (30) | 85. is an element of (30) |
| 86. the proof of this (30) | 86. is contained in the (30) | 86. is contained in the (30) |
| 87. we would like to (30) | 87. of the set of (30) | 87. of the set of (30) |
| 88. a consequence of the (28) | 88. the choice of the (30) | 88. the choice of the (30) |
| 89. gives rise to a (28) | 89. the collection of all (30) | 89. the collection of all (30) |
| 90. if there exists a (28) | 90. the sum of the (30) | 90. the sum of the (30) |
| 91. in the context of (28) | 91. we must show that (30) | 91. we must show that (30) |
| 92. is invariant under the (28) | 92. at the end of (28) | 92. at the end of (28) |
| 93. is the set of (28) | 93. if there is a (28) | 93. if there is a (28) |
| 94. of the choice of (28) | 94. in the sense of (28) | 94. in the sense of (28) |
| 95. the boundary of the (28) | 95. is a closed subspace (28) | 95. is a closed subspace (28) |
| 96. the right hand side (28) | 96. it is sufficient to (28) | 96. it is sufficient to (28) |
| 97. there is a unique (28) | 97. that there is an (28) | 97. that there is an (28) |
| 98. we will show that (28) | 98. be a sequence of (27) | 98. be a sequence of (27) |
| 99. a special case of (27) | 99. can be found in (27) | 99. can be found in (27) |
| 100. as a consequence of (27) | 100. let (m) be the (27) | 100. let (m) be the (27) |
| 101. assume without loss of (27) | 101. such that for every (27) | 101. such that for every (27) |
| 102. from the fact that (27) | 102. that the set of (27) | 102. that the set of (27) |
| 103. in order to prove (27) | 103. has the property that (25) | 103. has the property that (25) |
| 104. independent of the choice (27) | 104. independent of the choice (25) | 104. independent of the choice (25) |
| 105. (it) + is not difficult to (27) | 105. is independent of the (25) | 105. is independent of the (25) |
| 106. it is possible to (27) | 106. (we) + need to show that (25) | 106. (we) + need to show that (25) |
| 107. it will be convenient (27) | 107. of continuous functions on (25) | 107. of continuous functions on (25) |
| 108. of the set of (27) | 108. that there are no (25) | 108. that there are no (25) |
| 109. we assume that the (27) | 109. this shows that the (25) | 109. this shows that the (25) |
| 110. we need to show (27) | 110. to see that the (25) | 110. to see that the (25) |
| 111. as in the previous (25) | 111. to show that there (25) | 111. to show that there (25) |
| 112. is to show that (25) | 112. we now show that (25) | 112. we now show that (25) |
| 113. may assume that the (25) | 113. we obtain the following (25) | 113. we obtain the following (25) |
| 114. of this section is (25) | 114. we need to show (27) | 114. we need to show (27) |
| 115. we can find a (25) | 115. it will be convenient (27) | 115. it will be convenient (27) |

*Italicized sequences occur in all three lists.
**Frequency of occurrence (per million words).
***Bolded sequences occur in TXT-BKs and JOR-ARs.
****Bolded, italicized sequences occur in DISS-ONs and TXT-BKs or in DISS-ONs and JOR-ALs.
†Other lexical frames are possible.

Declaration of Conflicting Interests
The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The Author would like to extend appreciation to the Deanship of Scientific Research at King Saud University for funding this research.

ORCID iD
Abdullah A. Alasmary https://orcid.org/0000-0002-8025-5702

References
Ådel, 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.
Barton, B. (2008). The language of mathematics: Telling mathematical tales. Springer.
Barwell, R. (2005). Ambiguity in the mathematics Classroom. *Language and Education*, 19(2), 117–125.

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., & 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*. Pearson.

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

Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, 30, 38–52.

Chan, S. (2015). Linguistic challenges in the mathematical register for EFL learners: Linguistic and multimodal strategies to help learners tackle mathematics word problems. *International Journal of Bilingual Education and Bilingualism*, 18(3), 306–318.

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*, 25(4), 397–423.

Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum: Origins and functional analysis development. In V. Cortes & E. Csomay (Eds.), *Corpus-based research in applied linguistics: Studies in honor of Doug Biber* (pp. 197–216). John Benjamins.

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.

Ellis, N. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17–44.

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes, 29*, 21–42.

Grabowski, L. (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.

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.

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.

Hyland, K. (2009). *Academic discourse: English in a global context*. Continuum.

Jablonski, R. (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes, 29*(4), 253–267.

Leung, C. (2005). Mathematical vocabulary: Fixers of knowledge or points of exploration? *Language and Education, 19*(2), 126–134.

Liu, C. Y., & Chen, H. J. H. (2020). Analyzing the functions of lexical bundles in undergraduate academic lectures for pedagogical use. *English for Specific Purposes, 58*, 122–137.

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.

Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes, 39*, 1–11.

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.

Morgan, C. (2005). Words, definitions and concepts in discourses of mathematics, teaching and learning. *Language and Education, 19*(2), 102–116.

Monaghan, F. (1999). Judging a word by the company it keeps: The use of concordancing software to explore aspects of the mathematical register. *Language and Education, 13*(1), 59–70.

O’Halloran, K. (2005). *Mathematical discourse: Language, symbolism and visual images*. 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.

Schleppegrell, M. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading and Writing Quarterly, 23*(2), 139–159.

Scott, M. (2016). *WordSmith Tools version 7*. Lexical Analysis Software.

Shirazizzadeh, M., & Amirfazlian, R. (2021). Lexical bundles in these, articles and textbooks of applied linguistics: Investigating interdisciplinarity uniformity and variation. *Journal of English for Academic Purposes, 49*, 100946. https://doi.org/10.1016/j.jeap.2020.100946

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

Veel, R. (1999). *Language, knowledge and authority in school mathematics*. In F. Christie (Ed.), *Pedagogy and the shaping of consciousness: Linguistics and social processes* (pp. 185–216). Continuum.

Wagner, D., & Herbel-Eisenmann, B. (2008). “Just don’t”: The suppression and invitation of dialogue in mathematics classrooms. *Educational Studies in Mathematics, 67*(2), 143–157.

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