JELTL (Journal of English Language Teaching and Linguistics)
e-ISSN: 2502-6062, p-ISSN: 2503-1848
2021, Vol. 6(3)
www.jeltl.org
doi: https://dx.doi.org/10.21462/jeltl.v6i3.629

Functional Patterns of Lexical Bundles and Limitations in Academic Writing by Thai L2 English Learners

Attapol Khamkhien
Kasetsart University, Thailand
e-mail: faasapk@ku.ac.th

Abstract

Lexical bundles, which serve as markers of fluent and appropriate language use, are referred to as "building blocks of discourse" because they occur frequently in written texts and help readers and writers process information more quickly. They also provide important linguistic functions. However, lexical bundles can vary across genres, domains and even sections of the same work, which can be challenging for novice and non-native writers. The purposes of this study are to explore how Thai L2 undergraduate students use lexical bundles in their academic papers written in English and to compare the use of lexical bundles with that in two written corpora: the British Academic Written English (BAWE) and Cambridge Academic English (CAE). A total of 53 research reports, or approximately 615,750 words, from Thai L2 students of English language studies and applied linguistics were systematically compiled and analysed. The most frequent four-word n-grams in the corpora were then identified and their types and functions categorised. Keyword analysis was used to compare the key n-grams identified in each academic corpus. The results show that native and non-native writers use lexical bundles in rather different ways. Some patterns that did not occur in the reference corpus were overused by Thai L2 English students. The data suggest that the inclusion of phraseology in L2 writing instruction has pedagogical implications. This study can be of great pedagogical value, especially for EAP instructors, as it reveals frequent patterns in the form of a pedagogically useful list of word combinations. By extension, the data presented can be used by non-native writers or academics, especially novice Thai writers, to improve their use of phraseological patterns in writing academic research reports or writing for publication.

***Keywords**: corpus-driven approach, frequency analysis, lexical bundles, phraseology, research reports, Thai L2 learners*

1. INTRODUCTION

Non-native writers may have difficulty with conventional and academic writing styles in academic and scientific texts. Practice and understanding, which are interwoven in language acquisition, underpin native speakers' writing characterised by linguistic features in a text. It is essential for non-native and novice writers to improve the quality of their articles by learning the native writing style in a particular academic genre. To improve the quality of their work in the context of the academic community, writers need to use commonly accepted words and expressions. Thus, it is undeniable that learning common writing styles can be beneficial for high-quality research papers written in the appropriate style.

Numerous corpus analyses over the past decades have shown that language is extremely structured (Byrd & Coxhead, 2010; Chen & Baker, 2010; Cortes, 2013; Hong, 2018; Huang, 2015; Hunston, 2002; Hyland & Jiang, 2018; Le & Harrington, 2015; Römer, 2010). These studies have indicated that language used in academic texts is seen as multi-word sequences which are "ready-to-use" chunks, "stored and retrieved whole from memory at the time of use" (Wray, 2002, p. 9) rather than being created piece by piece. These sequences are particularly prevalent in certain registers. In terms of academic writing, these prefabricated components have been shown to facilitate writers' work and save processing time for readers (Nattinger & DeCarrico, 1992).

Phraseology (Cunningham, 2017; Granger & Meunier, 2008; Le & Harrington, 2015; Malá, 2020; Meunier & Granger, 2007), formulaic sequences (Pérez-Llantada, 2014; Schmitt, 2004; Wray, 2008), clusters, and n-grams are terms commonly used to refer to various types of multi-word units (Appel & Wood, 2016; Appel & Trofimovich, 2017). Lexical bundles were originally identified by Biber and colleagues (Biber et al., 1999) who described them as "the most frequently recurring sequence of words" (Biber & Barbieri, 2007, p. 264) and "important building blocks of discourse" (p. 270). The identification of lexical bundles in corpus research is largely based on a corpus-driven approach to frequency and range (Biber & Conrad, 1999; Pan, Reppen, & Biber, 2016). A sequence must occur at least 20 to 40 times per million words to be considered a lexical bundle (Biber & Barbieri, 2007; Chen & Baker, 2010; Cortes, 2004). The range of dispersion or the number of texts in which the bundle occurs is often set at 3 to 5 texts or 10% of the corpus (Hyland, 2008a). This criterion is applied to prevent idiosyncratic use by particular speakers or writers (Appel & Wood, 2016; Biber & Barbieri, 2007).

In the context of academic writing, formulaic language and phraseology provide recognisable patterns of use in word combinations for guidance. Linguistic evidence indicated by n-grams or lexical bundles is useful for academic purposes such as English writing, teaching materials, proficiency tests, and curriculum design. According to Biber and Barbieri (2007), Le and Harrington (2015) and Hyland and Jiang (2018), lexical bundles reflect genuine and unique language use based on communicative experiences in a particular discourse community. They are used to identify features of specific academic texts and to measure conventional patterns of language use.

To demonstrate their participation in the academic community, writers must successfully use lexical bundles typical of the genre and discipline (Ädel & Erman, 2012)

and adopt the linguistic forms used by expert writers to accomplish diverse rhetorical goals in research articles (Le & Harrington, 2015; Omidian, Shahriari & Siyanova-Chanturia, 2018). Writers who lack experience of the target language in a particular register may not choose the most appropriate expressions and may not be considered 'insiders' of that group (Durrant, 2017; Durrant & Mathews-Aydnl, 2011; Esfandiari & Barbary, 2017; Hyland, 2008; Wray, 2002). Unfortunately, knowledge of lexical bundles does not seem to be innate since it is far from being a 'universal language skill' (Pérez-Llantada, 2014, p. 85).

Research into the distinctive uses of lexical bundles by L1 and L2 learners has highlighted many important differences in academic discourse (e.g. Chen & Baker, 2010; Paquot, 2014; Pan, Reppen, & Biber, 2016). However, as issues of corpus comparability have become increasingly important in recent years (Appel & Murrey, 2020), more research is needed to better identify and validate the previously highlighted trends in language production using more comparable corpora to distinguish between L1 and L2 English writing. Despite the efforts of some studies to examine the differences between L1 and L2 learners, there still is a research gap in multiword sequences in terms of their functional discourse patterns. Furthermore, there have been very few attempts to investigate how L2 English learners apply lexical bundles to academic prose requiring careful consideration of context.

Therefore, the current study's main purpose is to use more comparable corpora of L2 English produced under analogous conditions by writers with L1 backgrounds to better understand their uses of lexical bundles in academic English texts. More specifically, this study aims to improve our knowledge of lexical bundles in learners' academic texts by investigating their usage in research reports written by Thai university students of English and applied linguistics. It is important to note that the analysed texts may contain grammatical errors since they reflect authentic language produced by Thai L2 learners. The lexical bundles are generated on a corpus-based approach that can process large language data in electronic form using a corpus tool (Salazar, 2014). Academic written texts from the BAWE and CAE corpora were used as references to compare the frequency of form and function of lexical bundles. The comparisons were made both quantitatively, by using a corpus-driven technique to identify lexical sequences in the learner and L1 corpora, and qualitatively, by classifying the identified bundles functionally and pragmatically from the perspective of the Thai learner corpus. Commensurate with these aims, the following research questions guided our investigation:

1. What are the most frequent four-word lexical bundles in research reports written in English by Thai L2 learners of English?
2. What are the similarities and differences between the four-word lexical bundles in the Thai corpus compared to those in the BAWE and CAE corpora?
3. What are the pragmatic functions of the four-word lexical bundles produced by Thai L2 learners of English?

2. LITERATURE REVIEW

2.1 Lexical Bundles in Academic Context

A lexical bundle is a recurring sequence of three or more words that occurs frequently in natural discourse, either orally or in written form (Biber et al. 1999). Research on these

chunks as fundamental parts of discourse is becoming increasingly important in English for Academic Purposes (EAP) (Altenberg, 1998; Altenberg & Eeg-Olofsson, 1990; Appel & Murrey, 2020; Appel & Wood, 2014; Biber & TracyVentura, 2007; Chen & Baker, 2010; Cunningham, 2017; Esfandiari & Barbary, 2017; Hong, 2018; Le & Harrington, 2015; Yoon & Choi, 2015). Lexical bundles associated with disciplinary variation have been studied by Cortes (2004) and Hyland (2008b). Biber, Conrad and Cortes (2004) have explored the role of lexical bundles in university teaching and textbooks. Lexical bundles have also been shown to make important contributions to linguistic proficiency (e.g., Appel & Wood, 2016; Biber & Barbieri, 2007; Shin, 2019; Staples et al., 2013).

The study of lexical bundles has gained prominence and made a significant contribution to the body of knowledge in applied linguistics. Many studies have been conducted on lexical bundles in written discourse, particularly at the university level (Biber et al., 2004; Byrd & Coxhead, 2010; Chen & Baker, 2010; Cortes, 2004, 2008; Yoon & Choi, 2015). Much attention has also been paid to general academic discourse in specific academic disciplines such as engineering (Wood & Appel, 2014), biology and history (Cortes, 2004, 2008), chemistry (Valipoor, 2010), mathematics (Cunningham, 2017), medicine (Abdollahpour & Gholami, 2018; Jalali, Moini, & Arani, 2015; Mbodj-Diop, 2016), telecommunication (Pan, Reppen, & Biber, 2016), psychology (Esfandiari & Barbary, 2017), pharmacy (Grabowski, 2015) as well as journalistic discourse (Dastjerdi & Rafiee, 2011), law (Breeze, 2013), and empirical research on second language conversation (Ädel & Erman, 2012; Chen & Baker, 2010; Hyland, 2008b; Wei & Lei, 2011). Until recently, lexical bundle knowledge has been supplemented by being linked to other theoretical ideas such as move analysis (Cortes, 2013; Kashiha, 2015; Moreno & Swales, 2018), language testing (Biber & Gray, 2013; Chen & Baker, 2016; Huang, 2015; Staples et al, 2013), and translation study (Lee, 2013).

A plethora of research has investigated the use of lexical bundles by non-native speakers of different levels across a range of registers and academic disciplines. Although there has been an increase in the use of lexical bundles by non-native speakers, previous studies indicate that their use is limited to certain bundles, leading them to overuse some expressions compared to others and making their writing appear non-native (Li & Schmitt, 2009). Some studies (e.g. Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2004; Huang, 2015; Yoon & Choi, 2015) argue that expert writers use lexical bundles in ways that are functionally different from novice writers and that non-native speakers generally have a more limited repertoire of recurrent word combinations than native speakers. Römer (2009) states that expert or professional writing is are more important than nativeness and that the distinction between novices and experts is more important than the distinction between a speaker's first language (L1) and second language (L2). Nesi and Basturkmen (2006) focused on the cohesive role of lexical bundles in a corpus of 160 university lectures and reported that the majority of frequently occurring bundles was used to signal discourse relations. Biber and Barbieri (2007) investigated the use of lexical bundles in a variety of spoken and written university registers and concluded that lexical bundles were very frequent in written discourse, in contrast to previous research which showed that bundles were much more frequent in speech than in writing. Chen and Baker (2010) used two

corpora of academic assignments from native and non-native university students to compare the use of lexical bundles. The study demonstrated that student writers are different from expert writers in that the L1 and L2 learners used more clausal bundles than did the expert counterparts, who used more phrasal bundles. Likewise, Yoon and Choi (2015) examined the use of bundles in argumentative essays produced by Korean university students and native English-speaking university students. Their findings showed that L2 students prefer using bundles in writing, whereas their native English-speaking counterparts expressed frequent use of phrasal bundles.

In order to identify lexical bundles, length is an essential factor. Biber et al. (1999), for example, studied lexical bundles of three to six words, and McCarthy and Carter (2006) analysed the sequences of lexical bundles of two to six words and found that it was not useful to look for lexical bundles of more than six words. However, researchers such as Ädel and Erman (2012), Appel and Trofimovich (2017), Biber et al. (2004), Chen and Baker (2010), Cortes (2004), Biber and Barbieri (2007), and Simpson-Vlach and Ellis (2010) focused only on four-word bundles in their studies because they were convinced that these were more common.

Another criterion is range or dispersion, which is sometimes prescribed in the literature to identify lexical bundles. Range means that a lexical bundle must occur in different types of text in a given register to be selected (Biber et. al., 1999). For example, lexical bundles must occur in at least 10 texts across the corpus of the study to be included in the list (Biber & Barbieri, 2007). This criterion aims to minimize possible idiosyncrasies.

2.2 Functional Classifications of Lexical Bundles

Since lexical bundles serve the function of discourse, several taxonomies have been developed. Hyland (2008a; 2008b) and Salazar (2014), for example, classified lexical bundles by function. Their taxonomy's functions relate to the meanings and purposes of language. They attempt to organise discourse according to situations or contexts. The three core categories in this taxonomy are 1) *research-oriented* bundles, which help the writer structure his or her activities and experiences in the real world; 2) *text-oriented* bundles, which are concerned with the organisation of the text and its meaning as a message or argument; and 3) *participant-oriented* bundles, which focus on the reader or writer of the text.

To illustrate, research-oriented bundles perform an ideational function according to Hyland (2008a; 2008b); expressions in this category are location (e.g., *at the beginning of*), process (e.g., *was carried out*), quantification (e.g., *a large number of*), description (e.g., *the appearance of*), grouping (e.g., *this type of*), and topic (e.g., *the currency board system*). Text-oriented bundles are combinations of words used to express textual functions. Some of the functions these expressions perform are transition (e.g., *on the other hand*), comparative (e.g., *as compared with*), inferential (e.g., *these results suggest that*), causative (e.g., *as a consequence of*), structuring (e.g., *as described earlier*), framing (e.g., *in the case of*), and objective (e.g., *to show that*). Meanwhile, participant-oriented bundles perform interpersonal functions by expressing stance (e.g., *it is likely that*) and engagement (e.g., *it should be noted that*).

Another functional taxonomy of lexical bundles is provided by Durrant (2017). His study analysed four-word bundles from four different disciplines, namely Humanities and Social Sciences, Science and Technology, Life Sciences, and Commerce. He then developed functional categories based on Hyland's (2008a; 2008b) categories and finally proposed a taxonomy with three functional categories based on the data collected in the study: research-oriented, text-oriented, and stance-oriented. The details of Durrant's (2017) classification are shown in Table 1.

Table 1. Lexical bundle classifications and their sub-categories (Durrant, 2017)

Functions	Sub-categories and functions	Examples of lexical bundles
Research-oriented	*structuring writers' activities and experiences of the world*	
	Location: indicating time/place	at the time of, in the process of
	Procedure: indicating *how* or *why* something is done/what something is for	as a means of, in a way that
	Quantification: indicating the quantity/extent of something	a large number of, the extent to which
	Description: indicating the physical properties of something	the presence of the, the shape of the
	Intangible framing attributes: indicating an abstract property of something	the existence of a, the form of the
Text-oriented	*indicating how elements of a text or its message relate to each other*	
	Transition signals: indicating relations of addition, contrast or equivalence between elements	an example of the, is defined as the
	Resultative signals: indicating inferential or causative relations between elements	does not mean that, as opposed to the
	Structuring signals: indicating stretches of discourse or referring readers to other parts of the text	as can be seen, as shown in figure
	Framing signals: indicating statements within a context or specifying their scope	for the purposes of, the context of the, the part of the
Stance	*indicating the writer's or someone else's attitude towards a statement in terms of its importance, epistemic standing or modality*	
	Centrality: indicating the importance of something	at the heart of, the importance of the
	Epistemic: indicating the certainty/uncertainty of something	can be seen as, appears to be a
	Modality: indicating the possibility of something	it is impossible to

To the author's knowledge, whereas most previous studies have paid considerable attention to the use of lexical bundles in different registers and a number of academic disciplines, only few studies focus on Thai L2 learners of English, especially the use of lexical bundles in research reports written by this group of students. Therefore, this research examined the most frequent four-word lexical bundles in English research reports written by Thai L2 learners of English. Quantitative and qualitative analyses were used to determine whether Thai learners' use of lexical bundles differs from that of expert or native writers by comparing the identified bundles with those in the reference corpora: the BAWE and CAE. Additionally, this study investigated the pragmatic functions these lexical bundles perform in

the environmental context. The results of the study are expected to inform the design and implementation of instructional EAP materials, such as academic reading and writing courses.

3. RESEARCH METHODS

This study used a mixed-methods approach. It began with a thorough examination of all four-word lexical bundles. It then examined those bundles that have a complete structure and a salient meaning. Once the identified bundles were carefully selected, tentative hypotheses were formulated to investigate them further and, perhaps, reach more comprehensive conclusions (Biber, 2009). Concordance lines of these bundles were examined to determine the functional categories of the lexical bundles, which required a top-down approach for the functional categorisation procedure. The corpus and research procedures used in this research are described below.

3.1 Data Compilation

The corpus for this study consists of 53 research reports written in English by fourth-year undergraduate students enrolled in an academic English writing course between 2018 and 2020. These English majors were, at the time of the study, attending a medium-sized university in Thailand. Their English proficiency level was intermediate to upper-intermediate and the texts were research reports in the field of English language studies and applied linguistics. In total, each text has a length of about 7,000 to 8,000 words. The entire corpus comprises approximately 615,750 words.

The texts obtained for this corpus were first copied and then converted from .pdf format into an MS Word document. While cleaning the corpus, bibliography, tables, and visual representations (e.g., diagrams, graphs, images, footnotes, headers and footers), the authors' identities and formulae were excluded. The processed texts were then saved in .txt format with Unicode 8 (UTF 8) encoding and labelled for further reference in the study. Following this sorting process, the final version of the corpus contains 53 texts comprising 404,793 words or 507,135 tokens in the Sketch Engine, as this tool counts punctuation marks as tokens.

3.2 Reference Corpora

For the comprehension of frequently used bundles by Thai L2 learners of English and expert writers, two corpora were carefully selected to represent a variety of corpus sizes and English words naturally produced by native speakers or L1 English users. The reference corpora used in this study include the British Academic Written English (BAWE) corpus and the Cambridge Academic English (CAE) corpus.

The BAWE corpus was developed by Hilary Nesi and Sheena Gardner in addition to Paul Thompson and Paul Wickens from 2004 to 2007. It contains written assignments collected from students in the United Kingdom, regardless of their nationality, at Warwick University, Reading University, Oxford Brookes University, and later some at Coventry University. The assignments were all written in English and submitted electronically. Furthermore, all assignments were assessed as standard productions in their respective fields (Alsop & Nesi, 2009), as they were "merit" and "distinction" assignments. The BAWE

corpus contains 6,968,089 words or 8,336,262 tokens, which is considered a large-sized corpus. Alsop and Nesi (2009: 72) stated that the disciplines represented in the corpus are Life Sciences, Social Sciences, Arts and Humanities, and Physical Sciences. As Silva (2017) argues, the texts included in the corpus contain some metadata, such as the students' educational levels, their grades, their previous studies, their gender, and other information. Therefore, this corpus has more detailed information compared to the second corpus used in this research.

The Cambridge Academic English (CAE) corpus is a corpus of academic English comprising text samples collected from the written and spoken academic language of undergraduate and postgraduate students at various universities in the US and the UK. These universities include the University of Cambridge, the University of Manchester, Anglia Ruskin University, the University of York, the University of Birmingham, the University of Durham, the University of Bristol, the University of Lancaster, the Chinese University of Hong Kong, Kansai Gaidai University, and others. The selected texts in the corpus were taken from lectures, seminars, student presentations, journals, essays, and textbooks comprising approximately 3,163,300 words or about 3,738,300 tokens.

Table 2. Details of the Thai learner corpus and the reference corpora used in this study

Corpora	Thai learner corpus	BAWE	CAE
Tokens	404,793 words or 507,135 tokens	6,968,089 words or 8,336,262 tokens	3,163,308 words or 3,738,308 tokens
Collection period	2008 - 2020	2000s	2000s
Variety of English	British / American	British	British / American
Spoken component	0%	0%	0%
Corpus size	About 7,000 to 8,000 words of each text	1,000 to 3,000 words of each text	1,000 to 3,000 words of each text
No. of texts	53 texts	2,012 texts	600 texts
Text-types	Research report consisting of IMRD sections (100%)	Essay (43.8%), methodology recount (11.7%), critique (11.2%), case study (6.8%), explanation (6.7%), and others (19.8%)	Essay (56.4%), dissertation/thesis (14.6%), research paper (12.2%), and others (16.8%), e.g., critique, proposal, creative writing

3.3 Identification of lexical bundles

In this study, only four-word bundles were examined because they are more frequent and easier to manage (Hyland, 2012). They are the extended unit of three-word bundles in terms of structures and functions (Cortes, 2013; Hyland, 2008a; 2008b) and are more common than 5 or 6 word bundles which are much more frequent in academic prose (Biber et al., 1999). Since lexical bundles are defined by the frequency of their occurrence and the distribution (or range) of their use in the text (Biber et al. 1999, p. 992), two criteria were used to identify them in this study: frequency and range. Range is used to establish that the

lexical bundles are not idiosyncrasies of particular speakers, whereas frequency threshold is used to prove that they are systematic, not random. The cut-off frequency was per million words, and the bundles had to occur in the corpus at least 10% of the time with at least 20 frequencies (Chen & Baker, 2010; Hyland & Jiang, 2018). The Sketch Engine (Kilgarriff & al. 2004) was used to extract the bundles. The lexical bundles found in the Thai learner corpus and the BAWE and CAE corpora were compared using keyword analysis.

3.4 Data Analysis

To address the research questions in this study, a frequency-based approach of the lexical bundles in the Thai learner corpus was first conducted as the unit of analysis. To reduce the list of identified bundles, those that overlap and context-dependent bundles were manually excluded from the analysis. Normalisation of the raw frequencies automatically extracted by the software was conducted for comparable reasons. The identified lexical bundles, considered 'key' bundles, were then compared with the BAWE and CAE corpora. Subsequently, the lexical bundles were manually classified into discourse functions using functional analysis in accordance with Durrant's (2017) functional classifications. He classifies lexical bundles into three categories: research-based bundles, text-oriented bundles, and stance-oriented bundles. Meanwhile, the concordance of Sketch Engine was used to determine the functional categories of lexical bundles. In doing so, the functional classification was complex, not only because categorisation involves subjectivity, but also because some bundles may serve more than one function (Khamkhien & Wharton, 2020; Liu, 2012). Therefore, three lecturers with a PhD in English and Applied Linguistics were invited to revisit the pragmatic functions of these bundles and to probe the extended context of certain bundles that appear to be multifunctional.

4. FINDINGS

4.1 Identification of Lexical Bundles

After four-grams in the three corpora were verified, the relative frequency of lexical bundles was calculated automatically and the range of every bundle was displayed to determine the distribution of bundles across corpora. Initially, the analysis revealed 2,964 four-grams in the Thai learner corpus, 31,681 four-grams in the BAWE corpus, and 13,595 four-grams in the CAE corpus. However, the bundles first identified, especially those in the Thai learner corpus, contain both functional and content n-grams (e.g., *functions of taboo words, Faculty of Liberal Arts, speech act of apology*). These cannot be considered to denote semantic or pragmatic functions and may not be useful for pedagogical purposes but rather indicate topic-specific bundles or topics of analysis. Although these bundles have structural characteristics such as prepositional phrases and noun phrases, they were discarded as they do not reflect the general use of academic language.

After the content n-grams were excluded, it turns out that 736 multiword combinations meet this criterion. From this number, further analysis focused only on the bundles that appear to be complete content phrases and have salient meanings in terms of functions and their teachability. It was found that 256 functional multiword units meet that criteria. These 256 functional multiword sequences were carefully selected on the basis of their pragmatic

functions and an intuitive-based judgment, following the model of functional classifications proposed by Durrant (2017).

Table 3 presents the top 30 lexical bundles in the Thai learner corpus, compared with the most frequent core lexical bundles identified in the BAWE and CAE corpora. The specificity of the multiword combinations in this study had to be determined by comparing the results with the frequent bundles discovered in the aforementioned corpora. To reiterate, it should be noted that the words from the BAWE and CAE corpora were taken from a multidisciplinary domain. Meanwhile, the list of lexical bundles extracted and carefully selected from the Thai learner corpus is discipline-specific and represents lexical bundles considered important and unique in applied linguistics and English language teaching. Therefore, a comparison between the three lists of identified bundles is methodologically justifiable as the lists were obtained using statistically-driven methods.

Table 3. Comparison of multiword units between Thai learner corpus, BAWE and CAE

No.	Thai learners		BAWE		CAE	
	4-grams	**Freq.**	**4-grams**	**Freq.**	**4-grams**	**Freq.**
1	**On the other hand**	117	as a result of	604	the end of the	270
2	is one of the	64	the end of the	578	the nature of the	259
3	the meaning of the	62	**On the other hand**	505	**On the other hand**	237
4	result of this study	59	in the form of	481	as well as the	231
5	the speech act of	51	as well as the	479	of the relationship between	213
6	types and functions of	51	at the same time	402	as a result of	209
7	the result of the	44	can be used to	386	the extent to which	196
8	the use of the	44	in the case of	377	at the end of	195
9	meaning used as a	42	can be seen in	363	the way in which	195
10	used as a noun	40	at the end of	361	in the context of	192
11	the perception of the	40	it is important to	358	in the form of	191
12	found that there are	40	on the other hand	334	on the other hand	190
13	the usage of address	38	the fact that the	329	can be seen in	178
14	the level of intimacy	36	it is possible to	327	can be used to	172
15	can be divided into	36	is one of the	326	in the case of	171
16	can be seen that	34	to be able to	318	at the same time	164
17	in the final position	33	that there is a	305	the ways in which	163
18	speech act of a	33	**to the fact that**	302	the fact that the	160
19	used in this study	31	the nature of the	292	in terms of the	160
20	of the use of	31	the rest of the	282	**one of the most**	148
21	The result shows that	30	**one of the most**	271	is one of the	145
22	to collect the data	30	the way in which	268	to be able to	142
23	**to the fact that**	27	a result of the	261	it is possible to	136
24	and the level of	26	can be seen that	241	**to the fact that**	128
25	it was found that	26	In the case of	234	the rest of the	127
26	the most frequently used	25	the extent to which	228	in response to the	122
27	it can been seen	24	in terms of the	225	it is important to	120
28	to the use of	24	It is important to	219	that there is a	118
29	an important role in	23	it is clear that	217	the beginning of the	114
30	**one of the most**	23	it can be seen	212	in the same way	112

Table 3 demonstrates that all top 30 multiword combinations in the BAWE and CAE corpora occur more than 100 times per million words, in contrast to those in the Thai learner corpus. This is due to the number of words included in the corpus and in the analysis. However, these lexical bundles are considered key multiword combinations for research report writing and academic vocabulary. As shown in bold text in the table, out of the 30 most frequently identified bundles, the prominent bundles shared in all three corpora are: "*On the other hand*", "*to the fact that*", and "*one of the most"*. Interestingly, almost all of the identified four-word underlined bundles are shared in the BAWE and CAE corpora acting as the reference corpora. As they are only found in the BAWE and CAE corpora, this finding, to some extent, reflects Thai L2 learners' limited knowledge and use of these lexical chunks. This suggests that native and professional writers generally tend to use the same lexical bundles in academic writing, while Thai L2 learners use fewer and far less diverse lexical bundles than native speakers (Ädel & Erman, 2012). The same observation was made by Hasselgård (2019, p. 347), who examined English academic texts written by Norwegian students and found that learners tend to reuse a small number of bundles to a greater extent than native speakers.

4.2 Results from keyword analysis

The term "keywords" here refers to the way the vocabulary is generally used, compared to that in the corpora. As defined by Culpeper and Demmen (2015), the term keywords in corpus linguistics refers to a word that is "statistically characteristic of a text or set of texts" (p. 90), using a different measure of *keyness*. Keywords are used to evaluate whether the word is more frequent in the *target corpus* than in the *reference corpus*. The advantage of keyword analysis is that the study is completely insulated from the researcher's bias (p. 60) because the study does not only take the quantitative view, but it can also provide the typical patterns around grammatical keywords or the qualitative method involving both the preferred meaning of a particular discourse community and the preferred stylistic features associated with that community (Malá, 2020).

The Sketch Engine's keyword function was used to produce a list of "key" multiword units that occur unusually often in the target corpus when compared to a reference corpus. However, since the main focus is directed exclusively at the positive keyness, the Sketch Engine was configured to provide only positive results, as shown in Table 4 for the BAWE corpus and Table 5 for the CAE corpus. This is because, as Pojanapunya and Watson Todd (2018) argue, the majority of keyness studies tend to set high statistical significance thresholds, thereby excluding low-frequency items from the comparison either directly or indirectly.

Table 4. Key lexical bundles in the Thai learner corpus with a significantly different frequency from those in the BAWE corpus

No.	Key 4-grams	Frequency in core corpus	Frequency in BAWE	Keyness value
1	the result of the	85	89	14.440
2	the results of the	82	137	9.332
3	the result of this	79	37	28.828

No.	Key 4-grams	Frequency in core corpus	Frequency in BAWE	Keyness value
4	as a foreign language	71	6	81.990
5	the meaning of the	65	79	12.329
6	English as a foreign	64	3	93.537
7	is one of the	64	327	3.162
8	the results of this	48	38	17.208
9	the use of the	47	235	3.209
10	the result shows that	45	4	60.638
11	it can be seen	45	314	2.321
12	found that there are	40	4	53.975
13	the perception of the	40	13	31.208
14	can be divided into	36	63	8.412
15	results of this study	35	9	33.667
16	it was found that	35	147	3.757
17	can be seen that	34	241	2.275
18	of this study is	32	20	18.857
19	the New York Times	32	93	5.273
20	of the meaning of	31	7	33.770
21	used in this study	31	13	24.274
22	of the use of	31	56	8.050
23	of this research is	28	5	35.137
24	the results showed that	27	22	14.905
25	the total number of	27	65	6.166
26	to the fact that	27	302	1.457
27	it shows that the	26	16	17.904
28	and the level of	26	21	14.853
29	one of the most	26	374	1.140
30	the purpose of this	25	126	3.121

Table 4 depicts that a word is included in the keyword list when its frequency is either unusually high or unusually low compared to the expected frequency based on the word list from the Thai learner corpus. The keyword analysis of the two corpora revealed differences in the number of keywords. The positive values represent those words that are more frequent in the Thai L2 learner corpus, i.e. those that are overused compared to the reference corpus or native English speakers, especially the use of *"English as a foreign language", "the result shows that"*, and *"of this research is"*. These multiword combinations also belong to the genre of research writing, which is somewhat distinctive from the text genres contained in the reference corpus.

Table 5. Key lexical bundles in the Thai learner corpus with a significantly different frequency from those in the CAE corpus

No.	Key 4-grams	Frequency in core corpus	Frequency in CAE	Keyness value
1	the result of the	85	23	23.573
2	the results of the	82	77	7.533
3	the result of this	79	9	46.009
4	as a foreign language	71	51	9.630
5	the meaning of the	65	52	8.663
6	English as a foreign	64	46	9.560
7	is one of the	64	145	3.197
8	the results of this	48	36	8.998
9	the use of the	47	109	3.106
10	it can be seen	45	67	4.742
11	the perception of the	40	3	44.313
12	used as a noun	40	3	44.313
13	can be divided into	36	22	10.456
14	as a noun and	35	3	38.843
15	results of this study	35	12	16.631
16	it was found that	35	73	3.411
17	in the final position	34	3	37.749
18	of the study the	34	10	18.515
19	can be seen that	34	47	5.013
20	the New York Times	32	3	35.561
21	of this study is	32	27	7.796
22	of the meaning of	31	11	15.758
23	used in this study	31	16	11.767
24	of the use of	31	54	4.023
25	of this research is	28	9	16.497
26	the results showed that	27	15	10.821
27	the total number of	27	57	3.338
28	to the fact that	27	128	1.539
29	one of the most	26	188	1.019
30	objectives of the study	26	4	25.250

A keyword analysis of the Thai learner corpus and the CAE corpus reveals that, as indicated by the keyness value and frequency of the core corpus and the reference corpus, L2 learners overuse some words in their writing, while they do not use many necessary or key words that occur in the reference corpus. Therefore, it can be concluded that they are generally more likely to rely on the use of multiword combinations than native English speakers, especially with the use of *"the perception of the", "used as a noun", "as a noun and", and "in the final position*". In this regard, it should be noted that the result could have been influenced by the size of the corpus for this study.

4.3 Functional classifications of lexical bundles

The functional analysis of the lexical bundles identified in the Thai learner corpus was carried out based on the taxonomy proposed by Durrant (2017), then adapted and further developed based on the classification of Hyland (2008a; 2008b) and the taxonomy of lexical bundles by Biber et al. (2014). Durrant's taxonomy was chosen and adapted because it accommodates the bundles found in research papers (including student texts) and in a wide range of academic written and spoken texts. The results of the functional analysis of the phraseological patterns identified in the Thai learner corpus are presented in Table 6.

Table 6. Detailed frequency and percentages of functional types in the Thai learner corpus (adapted from Durrant, 2017)

Categories	Subcategories	Number of multiword combination types	Sample multiword combination in the corpus
Research-oriented	A. Location	23 (8.98%)	in this study were, in the category if, in this study is, in this context the
	B. Procedure	51 (19.92%)	the use of the, the data were collected, by means of addressed, participants were asked to
	C. Quantification	22 (8.59%)	the total number of, that the majority of, the second half of
	D. Description	4 (1.56%)	the meaning of the, the study of the
	E. Intangible framing attributes	22 (8.59%)	the perception of the, the level of intimacy, the ability of the
Sub-total		122 (47.66%)	
Text-oriented	A. Transition signals	4 (1.56%)	on the other hand, in the same way, such as the word
	B. Resultative signals	42 (16.41%)	the result of the, the results shows that, due to the fact
	C. Structuring signals	12 (4.69%)	according to the result, can be seen in, this table shows the
	D. Framing signals	24 (9.38%)	The purpose of the, the part of the, is related to the
Sub-total		82 (32.03%)	
Stance-oriented	A. Centrality	5 (1.95%)	is one of the, an important role in, is the most important
	B. Epistemic	19 (7.42%)	can be seen that, can be said that, can be assumed that
	C. Modality	2 (0.78%)	will be able to, can be summarized that
Sub-total		26 (10.16%)	
Others		26 (10.16%)	affect the use of, do not use the, who are interested in
Total		**256 (100.00%)**	

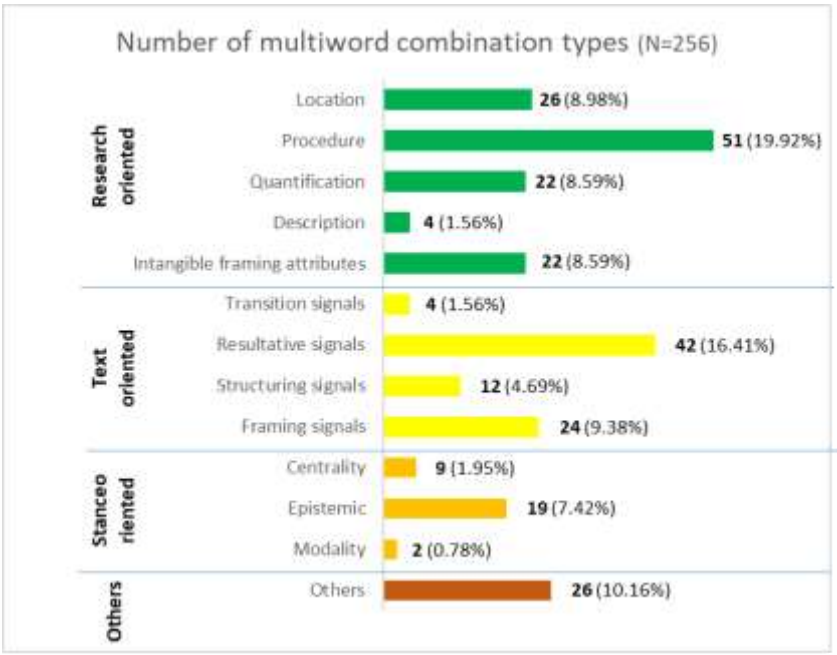

Figure 1. Number of multiword combination types found in the Thai corpus

Of the initial list of 736 lexical bundles, 256 bundles can be considered functional bundles and have semantic purposes and pragmatic functions that meet the established criteria. That is, these identified multiword clusters are structurally complete and have meaningful pragmatic functions. As can be seen in Table 6, 122 (47.66%) of the 256 lexical bundles on the list are research-oriented bundles that specify multiple attributes such as location, procedure, quantification, description, and intangible framing attributes. Meanwhile, 82 (32.03%) report research findings, thus forming the category of text-oriented bundles, 26 (10.16%) are stance-oriented, and 26 (10.16%) are classified as 'Others'. These results indicate that research report writing places more emphasis on describing situations, events and processes in research in addition to the description bundles used for reporting the quality, condition and existence of the topic under study.

These findings are not consistent with those of Hyland (2008b), in which Applied Linguistics and Business Studies relied on text-oriented functions. However, they correspond to Hyland's (2008b) results for other disciplines, namely Biology and Electrical Engineering, which relied heavily on research-oriented functions. The findings are in keeping with those of Allen (2009) and Beng and Keong (2015), who found that research-oriented lexical bundles had the highest proportion in their corpus.

This discussion reflects what Biber et al. (2004) pointed out, that there is a strong association between the form and function of lexical bundles in academic genres and discourses. Based on the analysis of university registers, their results suggest that the patterns are register-specific, meaning that certain registers frequently use lexical bundles. Durrant (2017) also suggests that a number of key differences in the use of lexical bundles may be due to academic disciplinary variation. In addition, it should be noted that the list presented above is not intended to be a definitive enumeration of the functional types of

student bundles, but rather a means of interpreting the present lists of identified bundles in their original context.

Furthermore, among the research-oriented function, "*procedure"* was the most frequently occurring subcategory, while "*description"* was the least frequently occurring subcategory. The most frequently occurring subcategory under the text-oriented function was "*resultative signals"*, and the least frequently occurring subcategory was "*transition signals"*. Finally, both stance-oriented and 'other' functions occurred with equal frequency.
In order to provide a complete picture of the pragmatic functions of lexical bundles identified in the Thai learner corpus, concordance lines were used to look at some bundles in each category in order to understand how they are applied to the environmental context. The following sections present an assessment of concordance lines in the Thai learner corpus. It should be noted that all examples presented hereafter are authentic examples from the corpus built for the present study.

4.3.1 Research-oriented bundles

Research-oriented bundles, also known as 'referential expressions' in Biber et al.'s (2004) taxonomy, helps the writer structure their research activities and methods and real-world experiences. As mentioned earlier, this category is dominated by lexical bundles that provide descriptions or explanations, whether they are objects, models, equipment or research materials. In addition to the description function, bundles with the procedure function are also frequently observed (51 bundles).

The other research-oriented bundles, namely location, quantification, grouping and topic, occur only infrequently. Even though their number is small, the bundles still contribute to the accuracy of the research process by indicating the location (23 or 8.98%) and research procedures (51 or 19.92%), determining size and number (22 or 8.59%), describing (4 or 1.56%), and indicating the intangible framing attributes (22 or 8.59%). In this regard, the location function is often realised in the form of prepositional-based bundles. The following instances, numbered (1) to (4), illustrate some realisations of research-oriented bundles.

(1) In order to scrutinize the characteristics of Thai Pidgin English, the selected informants *in this study were* originally Thais who were only woman merchants habituating regularly on KhaoSan road more than 5 years and were from 25 to 45 years old.

[TH 10-2018]

(2) This research aims to study the extent and characteristics of English-Thai code-mixing from news on the LINE Today official account based on the classification framework of Ho's (2007) and to study *the perception of the* sample group with different disciplinary backgrounds on how different perceptions were made to this phenomenon.

[TH 30-2020]

The extensive use of research-oriented bundles in research articles suggests that this type of genre places more emphasis on the research practice and methods, including procedures and equipment used as well as research objects.

(3) Chiravate (2011) in this research study, it was examined what measures the use of politeness strategy of Thai EFL learners' different from L1 and what the evidence of L1 influences *the use of the* politeness strategy of the learner in making requests.

[TH 29-2020]

(4) *The data were collected* through Reid's questionnaires (1987) that consisted of 6 learning styles, totaling 30 items.

[TH 12-2019]

4.3.2 Text-oriented bundles

The function of text-oriented bundles is also known as discourse organizers according to Biber et al.'s (2004) classification. It is concerned with the organisation of texts and their meaning as a message or an argument (Hyland, 2008a; 2008b). There are two subcategories that dominantly appear in this bundle, namely resultative signals (42 or 16.41%) and framing signals (24 or 9.38%). The resultative signals are mainly used to mark inferential or causative relations to the reporting of the study results. The following instances (5) and (6) are examples of multiword combinations used as resultative signal bundles in context.

(5) *The result of the* study showed that although an English advanced group could do better in multiple-choice tests, both groups of learners mostly made errors in the semi-controlled and free-writing test because of the mother-tongue interference or L1 interference, and the limited or misunderstanding collocational knowledge of EFL learners.

[TH 22-2020]

(6) Hence, it is worth to mention that language learning students who feel anxious when they have to speak cannot successfully accomplish the oral tasks *due to the fact that* their anxiety probably might influence their ability.

[TH 3-2019]

Framing signals deal with statements within a context and specify their scope or limiting conditions, as shown in (7) to (9). In addition, structuring functions (12 or 4.69%) and transitional signals (4 or 1.56%) also occur frequently in research papers. The framing function is associated with the conditioning of arguments by indicating conditional boundaries. Likewise, the structuring function refers to reflexive text markers that structure the text, arrange sequences, or guide the reader to a particular place in the text. Prepositional-based bundles are also mostly used for this function.

(7) The instrument was initially developed for *the purpose of the* investigation of requests and apologies (Blum-Kulka & Olshtain 1984) which popularized the DCT as the most efficient data collection tool of its time.

[TH 29-2020]

(8) Nevertheless, *according to the result* of current study analysis, this can conclude that The New York Times writers used syntactic features in terms of simple sentences and phrases.

[TH 28-2020]

(9) *According to the result*, the conventional indirect level preferred to use the conventional indirect strategies more than other strategies, this point to you that Thai people are very concerned about seniority.

[TH 6-2019]

From the description above, it can be seen that the bundle with this function is very helpful for writers to produce a unified and integrated idea. With these multiword units, writers can convey the interpretation of their data and readers will find it easier to understand articles through structured and logically arranged arguments. All these functions form the basis for effective argumentation.

4.3.3 Stance-oriented bundles

This functional category is concerned with the two-way interactions between the participants in the text, namely the writer and the reader. By expressing the epistemic function, the evaluative function and the importance of something, stance-oriented bundles help writers convey their attitudes towards claims and establish the appropriate relationship with their readers (Hyland, 2005). There are three functions in this category: centrality-, epistemic- and modality function, which convey the writers' attitudes and evaluations.

The centrality function in (10) refers to the way the writer acknowledges the importance of something, while the epistemic function (11) indicates its certainty or uncertainty. The modality function (12) refers to the presence of the reader to invite and actively engage the reader in the author's arguments the writer is conveying, to involve the reader as a participant in the discourse, and to guide the reader in interpretation (Hyland, 2005). These functions can be seen in the following instances.

(10) Dickinson (1994) found that all cultures use storytelling to persuade and believe that storytelling plays *an important role in* making persuasion successful because storytelling could keep people interested and follow it.

[TH 23-2020]

(11) It *can be seen that* the percentage of the word "Brandy" is 97% which is the highest percentage, while there are just 40% of all students made the word "Serious".

[TH 11-2019]

(12) The data obtained from the research *will be able to* reflect how much code-mixing infiltrated in Thai news.

[TH 30-2020]

4.3.4 Others

Several studies on lexical bundles (Hyland, 2008a; 2008b; Jalilifar, Ghoreishi, & Roodband, 2017; Salazar, 2014) note that they can serve more than one function in different contexts. Such multifunctionality was also found in this study. There are 26 multiword sequences that have multiple functions. For example, the bundle *"affect the use of*" may not only have the function of indicating research procedures, as shown in (13), but also serve as a resultative signal presenting the results of analysis, as shown in (14). Changes in the

functional category are influenced by the position of bundles in the sentence or by how the bundles are used in a contextual environment.

(13) As mentioned above, many requests' strategies study in different cultures showed that cultures might *affect the use of* the strategies as well.

[TH 29-2020]

(14) This research will be beneficial for people *who are interested in* English collocation and they can use this research to be the reference for learning advanced English.

[TH 22-2020]

5. DISCUSSION

The present study set out to identify the frequently used lexical bundles in research reports written by Thai L2 learners of English; to determine the extent to which Thai L2 learners of English and expert writers or native English speakers differ in terms of the lexical bundles they frequently use in writing academic texts; and to identify the pragmatic functions of the lexical bundles in the Thai learner corpus. The identification of lexical bundles was possible through Sketch Engine. To explore possible similarities and differences in the use of bundles, keyword analysis was performed. In classifying the functions of lexical bundles, Durrant's (2017) framework was used as a guide to determine the pragmatic function that each of the lexical bundles serves in the Thai learner corpus. This classification scheme made it possible to organise the lexical bundles based on their typical uses and semantic types as well as determine the extent to which each functional category is used in academic contexts.

The analyses showed that the bundles "*on the other hand", "is one of the", "the meaning of the",* and *"results of this study*" are the most frequent bundles used by Thai L2 English learners when writing research reports. A list of 256 convergent bundles was compiled, which can be a pedagogically useful resource for academic writing, especially research report writing. When comparing the Thai learner corpus with the BAWE and CAE corpora, it was found that *"on the other hand"* is the bundle that occurs the most frequently in all three corpora. The comparative results between the target corpus and the BAWE and CAE corpora also demonstrated that Thai learners overuse certain lexical bundles in academic texts, while the bundles used in the reference corpora are distinctly different and vary greatly in their frequency, indicating the importance of corpus size.

In terms of discourse function, Thai L2 English learners tend to use more research-oriented bundles of description, followed by text-oriented bundles of resultative signals. This finding is consistent with the work of Allen (2009) and Beng and Keong (2015) who indicate that research-oriented lexical bundles presented the highest proportion in their corpus. In academic disciplines, lexical bundles indicating procedure, quantity and resultative signals appear to be predominant. The findings of this study contribute to a better understanding of the features and functions of lexical bundles used in written academic discourse in general, and in written research reports in particular.

Through an in-depth qualitative investigation of the way certain lexical bundles are used in texts, the findings of this study reveal that the choice of linguistic forms or lexical bundles could be applied differently based on different situational contexts. This is in line

with Salazar (2014) who used semantic criteria to procure a refined and pedagogically useful list of bundles for teaching scientific writing in English. Moreover, the lexical bundles can be considered multifunctional when used in a particular context. This supports Biber and Barbieri (2007) and Simpson-Vlach and Ellis (2010) who argue that any specific functional taxonomy for multiword units necessarily suffers from a multiplicity of types and subtypes, which are often domain-specific. This situation is not conducive to distilling the data into a succinct functional model applicable to corpora representing various domains of language use. Therefore, it can be said that the methodology used in this study supports the claim and establishes the importance of qualitative data to complement quantitative analysis in extracting lexical bundles with their different types of functions.

The findings of this study have theoretical and practical benefits. Theoretically, the study aims to enrich knowledge about the ubiquity of lexical bundles in research report writing. More specifically, it is hoped to increase knowledge about the forms as well as the pragmatic functions of lexical bundles prevalently used in academic writing, especially in research report writing. From a pedagogical perspective, the findings of this study have important implications for the study of academic writing, which is one of the compulsory subjects at universities. As suggested by Kazemi, Kohandani and Farzaneh (2014), multiword sequences are becoming a key indicator in the evaluation of research articles. Meanwhile, Hyland (2008a) claimed that formulaic language is difficult to learn and that failure to use it correctly implicates learners as outsiders. Therefore, it is suggested that students who wish to write and publish research should pay special attention to multiword strings. The findings of this study may improve the quality of students' academic writing and help them present their ideas more clearly to become more credible researchers. This surely will have an impact on the acceptance of Thai scholars in the academic community and their respective field. Finally, the findings of this study are likely useful to English language teaching and education programs in creating teaching materials and designing course syllabi for academic writing courses. Consequently undergraduates, and perhaps postgraduates, may develop good competence in research report writing.

Although the present study sheds new light on the lexical bundles of Thai L2 English learners in writing research reports, it also has limitations. One major limitation is that the study only examined four-word strings and classified the pragmatic functions of these bundles, which provides an incomplete picture of formulaic language. The conclusions on register features are, therefore, rather limited. Another limitation is the classification of bundles. As Ädel and Erman (2012) have convincingly argued, the analytical framework leaves some room for arbitrary interpretations and applications. The functional analysis of the recurrent phrases was performed qualitatively, which may have resulted in possible discrepancies. However, this problem can be eliminated by using intuition-based judgements of EAP instructors, thereby increasing the reliability of the generated list of lexical strings from the analysis to some extent.

Future research should investigate linguistic patterns and the role they play with lexical bundles of academic writing in disciplinary variation. Such a study can deepen our understanding of writing in disciplinary variation and the construction of knowledge in other academic genres, which is beneficial for students, instructors, and novice writers when

writing research reports or perhaps publishing them. Furthermore, the functional patterns of lexical bundles discovered in this study may prove useful for further research and lead to practical applications of lexical bundles in language instruction.

6. CONCLUSION

This study investigated and compared the distribution of lexical bundles produced in academic prose by Thai L2 English learners and expert writers or native English speakers. The data analysis indicates some variations in the production of lexical bundles in a number of important areas evidenced by the use of lexical bundles, thereby highlighting the distinctive features of academic writing between L1 and L2 writers. While previous research on the use of lexical bundles by L2 English writers has primarily focused on L1/L2 contrasts, the present study conducted a keyword analysis with closely comparable corpora of L1 and expert writers to highlight similarities, discrepancies, and unique writing tendencies through the lens of lexical bundle use. These findings also offer insight into the use of recurrent multiword sequences by Thai L2 English learners in specific genres. In addition, the study presents a quantitative analysis complemented by a qualitative examination of pragmatic functions, which provides a more comprehensive picture of lexical bundle use following a manual categorisation and revision of concordance lines.

ACKNOWLEDGEMENT

This research work was funded by the Bachelor of Arts in English (Special Program) of the Faculty of Liberal Arts and Science at Kasetsart University. The author is grateful to Mr Daniel Lebedies for his valuable comments on the earlier version of this paper.

REFERENCES

Abdollahpour, Z., & Gholami, J. (2018). Building blocks of medical abstracts: Frequency, functions and structures of lexical bundles. *Asian ESP Journal, 14*(1), 82-110. https://doi.org/10.1016/j.sbspro.2014.03.473

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. https://doi.org/10.1016/j.esp.2011.08.004

Allen, D. (2009). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS Learner Corpus. *Komaba Journal of English Education, 1*,105-127.

Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora, 4*(1), 71-83. https://doi.org/ 10.3366/E1749503209000227

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 101– 122). Oxford: Oxford University Press.

Altenberg, B., & Mats, E. (1990). Phraseology in spoken English. In A. Jan & M. Willem. *Theory and Practice in Corpus Linguistics* (pp. 1-26). Amsterdam: Rodopi.

Appel, R., & Murrey, L. (2020). L1 differences in L2 English academic writing: A lexical bundles analysis. *Journal of English for Academic Purposes, 46*, 1-15. https://doi.org/10.1016/j.jeap.2020.100873

Appel, R., & Trofimovich, P. (2017). Transitional probability predicts native and non-native use of formulaic sequences. *International Journal of Applied Linguistics, 27*, 24-43. https://doi.org/10.1111/ijal.12100

Appel, R., & Wood, D. (2016). Recurrent word combinations in EAP test-taker writing: Differences between high- and low-proficiency levels. *Language Assessment Quarterly, 13*, 55-71. https://doi.org/10.1080/15434303.2015.1126718

Beng, C., & Keong, Y. (2014). Functional types of lexical bundles in reading texts of Malaysian University English Test: A corpus study. *GEMA Online Journal of Language Studies, 15*(1), 77-90. https://doi.org/ 0.17576/GEMA-2015-1501-05

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics, 14*(3), 275-311. https://doi.org/10.1075/ijcl.14.3.08bib

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgård, & S. Oksefjell (Eds.), *Out of corpora* (pp. 181-190). Amsterdam: Rodopi.

Biber, D., & Gray, B. (2013). Nominalizing the Verb Phrase in Academic Science Writing. In B. Aarts, J. Close, G. Leech & S. Wallis (Eds.). *The Verb Phrase in English: Investigating Recent Language Change with Corpora* (pp. 99-132). Cambridge: Cambridge University Press.

Biber, D., & Tracy-Ventura, N. (2007). Dimensions of register variation in Spanish. In G. Parodi (Ed.), *Working with Spanish corpora* (pp. 54-89). London: Continuum.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405. https://doi.org/10.1093/applin/25.3.371

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

Breeze, R. (*2013*). Lexical *bundles* in four legal genres. *International Journal of Corpus Linguistics, 18*(2), 229–253. https://doi.org/10.1075/ijcl.18.2.03bre

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL, 5*, 31-64.

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

Chen, Y. H., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics, 37*(6), 849-880. https://doi.org/10.1093/applin/amu065

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423. https://doi.org/10.1016/j.esp.2003.12.001

Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora, 3*, 43-57. https://doi.org/10.3366/E1749503208000063

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes, 12*(1), 33-43. https://doi.org/10.1016/j.jeap.2012.11.002

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. https://doi.org/10.1016/j.jeap.2016.11.005

Dastjerdi, H. V., & Rafiee, M. (2011). Corpus study of lexical bundles: Journalistic discourse in focus. *Asian ESP Journal, 7*(3), 59-78.

Durrant**, P.** (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics, 38*(2), 165-193. https://doi.org/10.1093/applin/amv011

Durrant, P., & Mathews-Aydinli, J. (2011). A function-first approach to identifying formulaic language in academic writing. *Journal of English for Specific Purposes, 30*(1), 58-72. https://doi.org/10.1016/j.esp.2010.05.002

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. https://doi.org/10.1016/j.jeap.2017.09.002

Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes, 38*, 23-33. https://doi.org/10.1016/j.esp.2014.10.004

Granger, S., & Meunier, F. (2008). *Phraseology: An Interdisciplinary Perspective*. Benjamins: Amsterdam & Philadelphia.

Hasselgård, H. (2019). Phraseological teddy bears: Frequent lexical bundles in academic writing by Norwegian learners and native speakers of English. In V. Wiegand & M. Mahlberg (Eds.), *Corpus Linguistics, Context and Culture* (pp. 339-362). Berlin: De Gruyter.

Hong, J. (2018). Use of lexical bundles by Korean EAP students at two different academic levels. *Foreign Languages Education, 25*(4), 23-52.

Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System, 53*, 13-23. https://doi.org/10.1016/j.system.2015.06.011

Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies, 7*, 173–192. https://doi.org/10.1177/1461445605050365

Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21. https://doi.org/10.1016/j.esp.2007.06.001

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62. https://doi.org/ 10.1111/j.1473-4192.2008.00178.x

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics, 32*, 150-169. https://doi.org/ 10.1017/S0267190512000037

Hyland, K., & Jiang, K. (2018). "We believe that…": Changes in an academic stance marker. *Australian Journal of Linguistics, 38*(2), 139-161. https://doi.org/10.1080/07268602.2018.1400498

Jalali, Z. S., Moini, M. R., & Arani, M. A. (2015). Structural and functional analysis of lexical bundles in medical research articles: a corpus- based study. *International Journal of Information Science and Management, 13*(1), 51-69.

Jalilifar, A., Ghoreishi, S. M., & Emam Roodband, S. A. (2017). Developing an inventory of core lexical bundles in English research articles: A cross-disciplinary corpus-based study. *Journal of World Languages, 3*(3), 184-203. https://doi.org/10.1080/21698252.2017.1301279

Kashiha, H. (2015). Recurrent formulas and moves in writing research article conclusions among native and non-native writers. *Pusat Pengajian Bahasa dan Linguistik, 21*(1), 47–59.

Khamkhien, A., & Wharton, S. (2020). *Constructing subject-specific lists of multiword combinations for EAP: A case study. Yearbook of Phraseology, 11*(1), 9-34. https://doi.org/10.1515/phras-2020-0003

Kilgarriff, A., Pavel, R., Pavel, S. & David, T. (2004). The Sketch Engine. In W. Geoffrey & S.V. Williams (Eds.), *Proceedings of the Eleventh EURALEX International Congress* (pp. 105-116). Université de Bretagne-Sud, Lorient: EURALEX.

Le, T. N. P., & Harrington, M. (2015). Phraseology used to comment on results in the discussion section of applied linguistics quantitative research articles. *English for Specific Purposes, 39*, 45-61. https://doi.org/10.1016/j.esp.2015.03.003

Lee, C. (2013). Using lexical bundle analysis as discovery tool for corpus-based translation research. *Perspectives, 21*(3), 378-395. https://doi.org/10.1080/0907676X.2012.657655

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing, 18*, 85-102. https://doi.org/10.1016/j.jslw.2009.02.001

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. https://doi.org/10.1016/j.esp.2011.07.002

Malá, M. (2020). Phraseology in learner academic English: Corpus-driven approaches. *Discourse and Interaction, 1*, 75-88. https://doi.org/10.5817/DI2020-2-75

Mbodj-Diop, N. (2016). *Lexical bundles in medical research articles: Structures and functions*. Unpublished master's thesis. Michigan State University.

McCarthy, M., & Carter, R. (2006). This that and the other: Multi-word clusters in spoken English as visible patterns of interaction. In M. McCarthy (Ed.), *Explorations in Corpus Linguistics* (pp. 7-26). New York: Cambridge University Press.

Meunier, F., & Granger, S. (2008). *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins Publishing.

Moreno, A. I., & Swales, J. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes, 50*, 40-63. https://doi.org/10.1016/j.esp.2017.11.006

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signaling in academic lecturers. *International Journal of Corpus Linguistics, 11*(3), 283-304. https://doi.org/10.1075/bct.17.03nes

Omidian, T., Shahriari, H., & Siyanova-Chanturia, A. (2018). A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. *Journal of English for Academic Purposes, 36*, 1-14. https://doi.org/10.1016/j.jeap.2018.08.002

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. https://doi.org/10.1016/j.jeap.2015.11.003

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes, 14*, 84-94. https://doi.org/10.1016/j.jeap.2014.01.002

Pojanapunya, P., & Todd, R. W. (2018). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory, 14*(1), 133-167. https://doi.org/10.1515/cllt-2015-0030

Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction, 3*(1), 95-119. https://doi.org/10.1075/etc.3.1.06rom

Römer, U. (2009). English in academia: Does nativeness matter. *Anglistik: International Journal of English Studies, 20*(2), 89-100.

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. Amsterdam: John Benjamins Publishing.

Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins Publishing.

Simpson-Vlach, R. & Ellis, N. (2010). An academic formulas list: new methods in phraseology research. *Applied Linguistics, 31*(4), 487–512. https://doi.org/10.1093/applin/amp058

Silva, L. G. (2017). Compilation of a Brazilian written English corpus. *E-scrita, 8*(2), 32-47.

Shin, Y. (2019). Do native writer always have a head start over non-native writers? The use of lexical bundles in college students' essays. *Journal of English for Academic Purposes, 40*, 1-14. https://doi.org/ 10.1016/j.jeap.2019.04.004

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes, 12*(3), 214-225. https://doi.org/10.1016/j.jeap.2013.05.002

Valipoor, L. (2010). *A corpus-based study of words and bundles in chemistry research articles*. Unpublished master's thesis, University of Kashan.

Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC journal, 42*(2), 155-166. https://doi.org/10.1177/0033688211407295

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. https://doi.org/10.1016/j.jeap.2014.03.002

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.

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

Yoon, C., & Choi, J. (2015). Lexical bundles in Korean university students' EFL compositions: A comparative study of register and use. *Modern English Education, 16*(3), 47-69. http://doi.org/10.18095/meeso.2015.16.3.03