Lexical Bundles in Native English Speakers' and Thai Writers' Dissertations

Nateethorn Narkprom [a]
nateethorn.nar@dome.tu.ac.th
Language Institute of Thammasat University, Thailand

Supakorn Phoocharoensil [b]
supakorn.p@litu.tu.ac.th
Language Institute of Thammasat University, Thailand

ABSTRACT

The purposes of this corpus-driven study were to compare the use of four-word lexical bundles between native English and Thai dissertation writings. Two language corpora, roughly 1,000,000 words apiece, were gathered from dissertations in the field of English Language Teaching written by both groups of writers. Each corpus was subdivided into three sub-corpora, namely, the Introduction, the Methodology, and the Results and Discussions sub-corpora. Two frameworks employed for the structural and functional analysis of the four-word lexical bundles were Salazar's (2011) adaptations of Biber et al.'s (1999) and Hyland's (2008a). The analysis of lexical bundles was performed using concordance software *AntConc*. The results showed that Thai writers overused lexical bundles in comparison with that of English speaking writers in each part of the dissertations, especially in the Results and Discussion section, which could result from institutional factors such as expectation and practice of Thai universities that expect Thai Ph.D. students to be more critical of the findings and to offer more implication. The structural analysis revealed the overuse of verb-structured lexical bundles throughout the three sub-corpora of Thai writers, which was likely to stem from the non-native speaker's failure to employ noun- and preposition-structured lexical bundles effectively. The proportion of functions of lexical bundles in each section of dissertations written by both groups of writers shared a relatively similar trend, indicating that English speakers and Thai writers conformed to the same convention of dissertation writing.

Keywords: corpus analysis; comparative analysis; lexical bundles; structures; functions

INTRODUCTION

The English language's dominance is visible through a wide distribution of English-written academic publications (Hyland, 2009; Swales, 2004), and a number of non-native post-graduates are required to compose English written dissertations. Despite the fact that the amount of academic prose written by authors who use English as a foreign language has been steadily rising (Hyland, 2006), many non-native scholars often find their dissertations or research writings not living up to standards and expectations set by academically-trained native writers (Salazar, 2011). One of the key revelations made by corpus-driven studies is that native speakers tend to use ready-made or prefabricated groups of words rather than unify isolated words in their writings (Sinclair, 1991) while non-native speakers of English often struggle with using lexical bundles (De Cock, 2003;

[a] *Main author*
[b] *Corresponding author*

eISSN: 2550-2131
ISSN: 1675-8021

Granger, 1998; Howarth, 1998; Nesselhauf, 2005). As a result, subjects involving the use of lexical bundles in academic prose written by native and non-native speakers of English require greater attention because a proper use of lexical bundles can be representative of a language competency level in a particular register (Biber & Barbieri, 2007; Cortes, 2004). This is especially the case for dissertation and academic writings, where it is compulsory for writers to compose clear and concise written content so as to deliver their thoughts and report the empirical findings to academic readers effectively. Academic writing is also determined by a clear convention set by professional writers within the community. For example, influenced by Swales' (1981) notion that language is largely "phraseological" in nature, Davis and Morley (2018) believed that academic and dissertation language is also characterized by its phraseological quality which serves different communicative purposes from everyday English. The study by Biber et al. (1999) demonstrated how lexical bundles are common in academic register, and following studies (e.g. Biber, 2009, Biber, Conrad, & Cortes, 2004; Chen & Baker, 2010; Durrant, 2015; Hyland, 2008b; Pan, Reppen, & Biber, 2016; Perez-Llantada, 2014) highlighted the differences between registers, genres, and disciplines. These researches on lexical bundles have only focused on an individual genre, such as student essays, or other specific fields (e.g., chemistry, history, biology). However, none has looked at lexical bundles in the genre of academic writing in the discipline of English Language Teaching. As a result, the goal of this study is to analyze and compare lexical bundles in academic writing across two groups of writers (English and Thai) in order to provide insight to teachers, non-native postgraduate students, and educational academics. The main findings will respond to three major research questions as follows.

1. What are mutual four-word lexical bundles used in native English and Thai dissertation writings?
2. How are four-word lexical bundles used by English speaking writers structurally different from those used by Thai writers in dissertation writings?
3. How are four-word lexical bundles used by English speaking writers functionally different from those used by Thai writers in dissertation writings?

LITERATURE REVIEW

LEXICAL BUNDLES

Lexical bundles were described as strings of words that statistically exist together at least forty times in one million words across at least five texts in a register (Biber et al., 1999). According to Cortes (2004), one of the defining features of lexical bundles is their fixedness. Lexical bundles are frequently not idiomatic but compositional since the meanings come from the words they consist of. Lexical bundles are structurally incomplete entities, being partial to phrases or sentences with incorporated fragments. Nevertheless, notable scholars have attempted to classify lexical bundles into certain grammatical (Biber et al., 1999) and functional categories (Hyland, 2008a), as shown in Table 1 and Table 2, respectively.

eISSN: 2550-2131
ISSN: 1675-8021

STRUCTURES OF LEXICAL BUNDLES

TABLE 1. Salazar's (2011, p.51) Adaptation of Biber et al.'s (1999) Structural Classification of Lexical Bundles

Noun structures	
NP+of, noun phrase + *of*-phrase fragment	Example: *the purpose of this*
NP+other, noun phrase with other post-modifier fragment	Example: *an important role in*
other NP, other noun phrase	Example**:** *the present research study*
Verb structures	
passive+PP, passive + prepositional-phrase fragment	Example: *are shown in the*
other passive, other passive fragment	Example: *has been described previously*
we+V, verb phrase with personal pronoun *we*	Example: *we found that the*
other V fragment, other verbal fragment	Example**:** *does not require the*
Prepositional structures	
PP+of, prepositional phrase + *of* fragment	Example: *in the presence of*
other PP, other prepositional phrase (fragment)	Example**:** *with respect to the*
Other structures	
V/A+to, verb or adjective *to*-clause fragment	Example: *has shown to be, is likely to be*
V/N+that cl, verb phrase or noun phrase + *that*-clause fragment	Example: *this suggests that the, the possibility that the*
as+V, adverbial-clause fragment	Example: *as seen in the*
be+AP, copula *be* + adjective phrase	Example: *is consistent with the*
other AP, other adjectival phrase	Example: *significantly different from the*
anticipatory it, anticipatory *it* + verb or adjectival phrase	Example: *it is likely that*
others, other expression	Example**:** *in order to be, as well as the*

FUNCTIONS OF LEXICAL BUNDLES

TABLE 2. Salazar's (2011, pp.52-23) Adaptation of Hyland's (2008a) Functional Classification of Lexical Bundles

Research-oriented to build up the writer's activities or real-world experiences	
Location: time or place	Example: *at the beginning of*
Procedure: methodology or study purpose	Example: *the use of the*
Quantification: an amount or number	Example: *a large number of*
Description: qualities or properties of materials	Example: *the appearance of the*
Grouping: groups, categories, parts and order	Example: *a wide range of*
Text-oriented to signal textual organization and to convey its meaning as a message or an argument	
Additive: establishing additive links between elements	Example: *in addition to the*
Comparative: comparing and contrasting different elements	Example: *in contrast to the*
Structuring: markers organizing the text or directing readers	Example: *as shown in figure*
Framing: making arguments by specifying the limitation	Example: *on the basis of*
Citation: citing sources and supporting data	Example: *studies have shown that*
Generalization: signalling accepted facts or statements	Example: *little is known about*
Objective: indicating objective relations between elements	Example: *in order to the, to discuss the*

eISSN: 2550-2131
ISSN: 1675-8021

Participant-oriented to build up the writer's activities or real-world experiences	
Stance: conveying writer's attitudes and evaluations	Example: *may be due to a*
Engagement: addressing readers directly	Example: *it should be noted that*
Acknowledgment: recognizing people that participate in or contributed to the study	Example: *kindly provided by the*

METHOD

DESCRIPTION AND CREATION OF THE CORPORA

Two specialized corpora were compiled by the researchers: the English speaking writer corpus and the Thai writer corpus. The collection of the corpus data was based on three criteria, based on Salazar's (2011) recommendations, including topic, text type, and author profile. First, the texts were doctoral dissertations in the field of English Language Teaching. Secondly, the texts were licensed and retrievable in *PDF* full text through online databases of academic institutions in countries where the English language is officially the first or dominant language such as the United States, the United Kingdom and others in Kachru's (1985) Inner Circle. Thirdly, first names and surnames of dissertation authors must represent English language properties. To further ensure the native speaker status of the text composers, the writers' linguistic and demographic background were double-checked through respondent electronic mails and social media platforms such as their Facebook and LinkedIn personal profile. In certain cases in which the writers' names did not represent Anglophilic properties and personal profile information was not available, the researchers looked up the vita or curriculum vitae, usually at the of the dissertations, to verify the biological backgrounds and upbringings in order to authenticate the author's mother language. The compilation of data for the Thai writer corpus was required to meet the criteria similar to that of the English speaking writer corpus. However, the difference between the two corpora was that all authors must have Thai names and were born in Thailand. Doctoral dissertations written by Thais who graduated from universities in Thailand or abroad were both considered as a part of the Thai writer corpus. Both of the English speaking writer and Thai writer corpora yielded a total of 50 doctoral dissertations and were roughly one-million word large. Each corpus was subdivided into three sub-corpora, namely, the Introduction, the Methodology, and the Results and Discussions sub-corpora, as described in Table 3.

TABLE 3. Description of the Corpora

Names of Sub-corpora	**Word Counts**	
	Eng	**Thai**
Introduction	191,239	154,268
Methodology	215,419	348,614
Results and Discussions	713,041	765,559
Total	1,119,699	1,268,441

DATA PREPARATION

The data was analyzed using corpus software called *AntConc* (Anthony, 2019). The introduction, method, results, and discussions of dissertations that met the predefined criteria were extracted and then converted from the *Portable Document Format* (*.pdf) into the *Plain Text* format (*.txt), using a converter software named *AntFileConverter* (Anthony, 2017). Because of possible misspellings, all samples were proofread and corrected to guarantee errorless and precise data processing. In addition, unnecessary parts such as tables, author names, interview quotes, figures and page numbers were excluded not to mislead the results of the analysis. Eventually, all texts were saved into the *Plain Text* format (*.txt) to be compatible with the *AntConc*.

IDENTIFICATION OF LEXICAL BUNDLES

The lexical bundles were identified using the *AntConc* program with three preconfigured settings: the cut-off frequency, the distribution, and the length of lexical bundles. A list of lexical bundles was created. The criteria for determining lexical bundles are discussed as follows.

The cut-off frequency was to quantify the number of the target lexical bundles. Reviews of literature indicated the value of cut-off frequency depending on the sizes of the target corpus, 20 to 40 times per 1,000,000 words for large corpora (Biber, Conrad, & Cortes, 2004; Cortes, 2004; Hyland, 2008a, 2008b). The current study used the cut-off frequency of 10, equal to 100 occurrences per 1,000,000 words. This cut-off point was higher than the "conservative approach" to lexical bundle analysis of 40 occurrences per a million words, which was adopted by Biber, Conrad, and Cortes (2004).

The distribution of a lexical bundle across the corpus was taken into consideration since using the cut-off points alone could be misleading. A lexical bundle could be used so frequently in a text by a writer that it caused inflation, as a result of the writer's preference to a specific bundle. To avoid such misinterpretation, a lexical bundle should exist in at least 3 to 5 texts (Biber & Barbieri, 2007; Cortes, 2004). Therefore, this study used the distribution of 5 texts across the corpus for the identification of lexical bundles.

Lastly, the length of lexical bundles was defined at 4. According to Hyland (2008a), four-word lexical bundles were more frequent than five-word ones and represented more obvious structures and functions and were the most frequently studied sequences (Chen & Baker, 2010).

ANALYSIS OF LEXICAL BUNDLES

The structural classification of the lexical bundles was adopted from Salazar's (2011) adaptation of Biber et al.'s (1999) because of its practicality and comprehensibility. Regarding practicality, Biber et al.'s taxonomy has been widely cited and modified by a number of previous researchers who empirically investigated lexical bundles. Concerning comprehensibility, Biber et al.'s framework provided researchers with more thorough types of structure of lexical bundles potentially discovered in the corpus. This framework structurally divided the lexical bundles into four broad grammatical categories: noun-structured lexical bundles, verb-structured lexical bundle, preposition-structured lexical bundles, and other-structured lexical bundles. More specifically, Salazar's (2011) adaptation of Biber et al.'s (1999) framework also divided these four broad grammatical categories into 16 subdivisions. Concerning the functions of lexical bundles, the adapted version of Hyland's (2008a) functional taxonomy of lexical bundles by Salazar (2011)

was used in this study. This taxonomy divided the lexical bundles into three broad functional types, namely, research-oriented, text-oriented, and participant-oriented lexical bundles. The adaptation of Hyland's (2008a) taxonomy by Salazar (2011) further subdivided the three broad types of lexical bundle's functions into 15 subdivisions.

RESULTS AND DISCUSSION

LEXICAL BUNDLES IN THE INTRODUCTION SUB-CORPORA OF ENGLISH SPEAKING WRITERS AND THAI WRITERS

Table 4 lists the mutual lexical bundles statistically underused and overused by Thai academics, as compared with the native speakers ($p<0.01$, critical value 6.63), in the Introduction sub-corpora.

TABLE 4. Comparison of Mutual Lexical Bundles in the Introduction Sub-corpora of English Speaking Writers and Thai Writers

LBs	Frequency		LogL	LBs	Frequency		LogL
	Eng	Thai			Eng	Thai	
as a result of	36	10	-10.54	on the other hand	10	25	+10.27
in the field of	15	29	+8.05	the findings of the	10	31	+16.27
is one of the	10	35	+20.60	the results of the	10	23	+8.44

According to Table 5, the structural analysis showed that, in the Introduction sub-corpora, the most frequently used bundles that both English speaking writers and Thai authors utilized were noun-based (39.20%; 35.46%), other-structured (26.40%; 30.00%), preposition-based (25.60%; 23.64%), and verb-based (8.80%; 10.92%), respectively. The analysis revealed that there were 125 types of four-word lexical bundles in the Introduction sub-corpus of English speaking writers. The most frequent lexical bundles used by native speakers of English included *of English language learners*, *the purpose of this*, *meet the needs of*, *as a result of*, and *to meet the needs*. In the Introduction sub-corpus of Thai writers, the analysis showed 110 types of four-word lexical bundles. The most common lexical bundles used by Thai authors were *is one of the*, *the findings of the*, *the effects of the*, *investigate the effects of*, and *in the field of*. The results from both the Introduction sub-corpora of English speaking writers and Thai academics collaborate with earlier findings which revealed that non-English speaking writers utilize a smaller number of lexical bundles (Erman, 2009; Howarth, 1998). These findings are supported by other studies (Adel & Erman, 2012; Biber et al., 1999; Biber, Conrad, & Cortes, 2004) which indicated academic writing composed by English speaking writers were dominated by phrasal bundles (noun- and preposition-based structures). However, the Introduction sub-corpus of Thai writers represented a statistically significant overuse of verb-based bundles in comparison with that of English speaking writers, based on the log-likelihood value of 12.69. These inflated uses of verb-based construction in the Introduction sub-corpus of Thai authors were consistent with the findings of Gungor and Uysal (2016) and Pan, Reppen, and Biber (2016) which discovered that non-native scholars (Turkish and Chinese writers of English language) overused verb-based bundles in their research articles. This excessive use probably stems from the non-native English writers' failure to employ noun phrase and prepositional phrase structures by Gungor and Uysal (2016).

TABLE 5. Structural Comparison of Lexical Bundles in the Introduction Sub-corpora of English Speaking Writers and Thai Writers

Structure	Raw Frequency		Normalized Frequency		Type		%		LogL
	Eng	Thai	Eng	Thai	Eng	Thai	Eng	Thai	
NP-based	581	509	3,038.08	3,299.45	49	39	39.20	35.46	1.84
NP+of	481	431	2,515.18	2,793.84	39	32	31.20	29.09	2.50
NP+other	70	56	366.03	363.00	7	5	5.60	4.55	0.00
other NP	30	22	156.87	142.61	3	2	2.40	1.82	0.12
VP-based	148	178	773.90	1,153.83	11	12	8.80	10.92	+12.69
passive+PP	10	46	52.29	298.18	1	4	0.80	3.64	+33.46
other passive	10	32	52.29	207.43	1	3	0.80	2.73	+17.33
we+V	0	0	0.00	0.00	0	0	0.00	0.00	0.00
other V fragment	128	100	669.32	648.22	9	5	7.20	4.55	0.06
PP-based	464	339	2,426.29	2,197.47	32	26	25.60	23.64	1.93
PP+of	279	199	1,458.91	1,289.96	23	15	18.40	13.64	1.77
other PP	185	140	967.38	907.51	9	11	7.20	10.00	0.33
Other	388	399	2,028.87	2,586.4	33	33	26.40	30.00	+11.58
V/A+to	166	105	868.02	680.63	14	7	11.20	6.36	3.87
V/N+that cl	10	45	52.29	291.70	1	4	0.80	3.64	+32.24
As+V	0	0	0.00	0.00	0	0	0.00	0.00	0.00
be+AP	0	0	0.00	0.00	0	0	0.00	0.00	0.00
anticipatory it	31	72	162.10	466.72	3	7	2.40	6.36	+26.78
other AP	0	24	0.00	155.57	0	2	0.00	1.82	+38.70
others	181	153	946.46	991.78	15	13	12.00	11.82	0.18
Total	1,581	1,425	8,267.14	9,237.17	125	110	100	100	

As seen in Table 6, lexical bundles elicited from the Introduction sub-corpora of English speaking writers and Thai writers were categorized based on their textual functions. In the Introduction sub-corpus of the native scholars, text-oriented bundles were mostly preferred at 52.31%, followed by research- and participant-oriented bundles at 35.38% and 12.31%, respectively. Text-oriented bundles were mainly used for introducing the writer's aims (*n*=26), situating arguments by specifying limiting conditions (*n*=16) and organizing stretches of discourse or directing the reader elsewhere in the text (*n*=11). The present findings of text-oriented bundles being the most dominant function in native writings contradicted with other studies (e.g., Biber, 2009; Biber & Barbieri, 2007; Chen & Baker, 2010; Gungor & Uysal, 2016; Jukneviciene, 2009; Salazar, 2011) but corresponded with some (e.g., Demiray Akbulut, 2020; Pan, Reppen, & Biber, 2016). The dominance of text-oriented bundles in the Introduction sub-corpus of English speaking writers was probably affected by the nature of corpus data, which was a collection of Ph.D. dissertations in the fields of English Language Teaching. Research in this field is viewed more as interpretative—and less as experimental knowledge and generally characterized by text-oriented bundles (Hyland, 2008a). Moreover, the characteristic of writing introductions to research, for which statements of hypotheses and purposes were expected (Jha, 2014; Boyd, Rifai, & Annesley, 2010 as cited in Bavdekar, 2015), might play a role in the recurring demonstration of the writer's aims through the uses of text-oriented lexical bundles. The Introduction sub-corpus of the Thai writers was also predominated by text-oriented bundles at 46.67%, succeeded by research- and participant-oriented bundles at 35.88% and 17.50%, respectively. Text-oriented bundles used by Thai authors in the Introduction sub-corpus were chiefly employed for introducing the writer's aims (*n*=13), marking cause and effect relations between elements (*n*=9) and signaling inferences and conclusions drawn from data (*n*=8). The results of text-oriented bundles being the most used

textual function by Thai authors in the Introduction sub-corpus contradicted with the study on functions of lexical bundles conducted with Thai writers by Panthong and Poonpon (2020). This could be due to the fact that Panthong and Poonpon's corpus was a collection of research publications in the fields of medicine, which was more scientifically and experimentally focused. Hence, the text-oriented bundles were outnumbered by research-oriented bundles. The present findings, however, were consistent with other studies on functions of lexical bundles in academic prose written by non-native speakers (e.g., Demiray Akbulut, 2020; Gungor & Uysal, 2016; Pan, Reppen, & Biber, 2016). In addition to the characteristics of the corpus data and introduction writing convention, the overuse of text-oriented bundles in the Thai author's corpus might be a result of non-native scholars viewing those bundles as a highly sophisticated functional category (Hyland, 2012 as cited in Gungor & Uysal, 2016).

TABLE 6. Functional Comparison of Lexical Bundles in the Introduction Sub-corpora of English Speaking Writers and Thai Writers

Functions	Raw Frequency		Normalized Frequency		Type		%		LogL
	Eng	Thai	Eng	Thai	Eng	Thai	Eng	Thai	
Research-oriented	455	541	2,379.22	3,506.88	46	43	35.38	35.83	+37.38
Location	49	107	256.22	693.60	5	10	3.85	8.33	+36.35
Procedure	134	139	700.69	901.03	14	12	10.77	10.00	4.31
Quantification	72	21	376.49	136.13	7	2	5.38	1.67	-19.69
Description	180	177	941.23	1,147.35	18	14	13.85	11.67	3.49
Grouping	20	97	104.58	628.78	2	5	1.54	4.17	+73.06
Text-oriented	884	736	4,622.49	4,770.92	68	56	52.31	46.67	0.04
Additive	36	43	188.25	278.74	2	3	1.54	2.50	3.03
Comparative	20	52	104.58	337.08	2	4	1.54	3.33	+22.44
Inferential	68	118	355.58	764.90	6	8	4.62	6.67	+26.49
Causative	68	126	355.58	816.76	3	9	2.31	7.50	+32.30
Structuring	107	82	559.51	531.54	11	8	8.46	6.67	0.12
Framing	179	85	936.00	550.99	16	6	12.31	5.00	-17.06
Citation	0	32	0.00	207.43	0	3	0.00	2.50	+51.60
Generalization	20	25	104.58	162.06	2	2	1.54	1.67	2.15
Objective	386	173	2,018.42	1,121.43	26	13	20	10.83	-43.93
Participant-oriented	242	148	1,265.43	959.37	16	21	12.31	17.5	-7.18
Stance	55	94	287.60	609.33	6	14	4.62	11.67	+20.42
Engagement	6	34	31.37	220.40	1	5	0.77	4.17	+28.11
Acknowledgement	181	20	946.46	129.64	9	2	6.92	1.67	-116.13
Total	1,581	1,425	8,267.14	9,237.17	130	120	100	100	

LEXICAL BUNDLES IN THE METHODOLOGY SUB-CORPORA OF ENGLISH SPEAKING WRITERS AND THAI WRITERS

The shared lexical bundles significantly underused and overused by Thai authors, $p<0.01$, critical value 6.63, in comparison with the English speaking authors in the Methodology sub-corpora are demonstrated in Table 7.

TABLE 7. Comparison of Mutual Lexical Bundles in the Methodology Sub-corpora of English Speaking Writers and Thai Writers

LBs	Frequency		LogL	LBs	Frequency		LogL
	Eng	Thai			Eng	Thai	
allowed the researcher to	27	11	-8.14	the purpose of this	51	22	-14.00
at the beginning of	13	37	+10.37	the reliability of the	10	62	+38.19
at the end of	30	119	+50.75	the results from the	10	28	+7.65
in the field of	10	95	+76.31	the results of the	29	57	+7.40
in the form of	12	44	+17.24	the total number of	10	34	+12.18
participants were asked to	12	53	+25.10	the validity of the	11	38	+13.88
the beginning of the	10	43	+19.86	to participate in the	67	21	-28.66
the effectiveness of the	10	39	+16.34	used in the study	10	32	+10.61
the end of the	17	117	+76.88	used in this study	11	34	+10.75
the participants in the	10	40	+17.21				

Based on Table 8, in the Methodology sub-corpus, English speaking writers employed bundles that were noun-based (40.87%), other-structured (27.88%), preposition-based (18.27%), and verb-based (12.98%), whereas, in that of Thai academics, the bundles were noun-based (30.90%), other-structured (27.24%), verb-based (24.58%), and preposition-based (17.27%), respectively. English speaking writers employed 208 types of four-word lexical bundles. The most frequent lexical bundles used in native writing consisted of *the purpose of this*, *participate in the study*, *the purpose of the*, *the results of the*, and *at the end of*. On the other hand, the Methodology sub-corpus of Thai authors produced 301 lexical bundles. The most used lexical bundles in the Thai writer's corpus were *in the main study*, *at the end of*, *in the field of*, *in the pilot study*, and *the reliability of the*. The findings disagreed with prior studies which claimed non-native English speakers employ fewer (Erman, 2009; Howarth, 1998) and less diverse (Granger, 1998) lexical bundles. However, the results were aligned with other studies (Gungor & Uysal 2016, Hyland, 2008b; Ozturk, 2004; Perez-Llantada, 2014; Pan, Reppen, & Biber, 2016) that found non-native speakers utilize a wide variety of lexical bundles. For instance, the study conducted by Ozturk (2004) revealed that there were twice as many lexical bundles in non-native postgraduates' writing compared to research papers, Master theses, and Doctoral dissertations written by native postgraduate students. This trend of inflated use of lexical bundles in the Methodology sub-corpus of Thai writers was also similarly observed in Gungor and Uysal's (2016) corpus of research articles written by non-native speakers. As a result, the large number of lexical bundles in the Methodology sub-corpus of Thai academics could be representative of recurring use of formulaicity and fixedness in the genre of academic writing (Perez-Llantada, 2014) and reaffirms Greaves and Warren's (2010) and Biber et al.'s (1999) notion that academic prose was dominated by lexical bundles. The Methodology sub-corpus of Thai writers also represented a statistically significant overuse of verb-, preposition-, and noun-based bundles in comparison with that of English speaking writers in descending order of log-likelihood values 227.30, 162.99, 28.61 ($p<0.01$), respectively. These inflated uses of verb-based construction in the Methodology sub-corpus of Thai authors were consistent with the findings of Gungor and Uysal (2016) and Pan, Reppen, and Biber (2016) which discovered that non-native scholars (Turkish and Chinese writers of English language) overused verb-based bundles in the research articles.

The dominance of clausal or verb-phrase structures in the Methodology sub-corpora of both English speaking writers and Thai authors were also supported by the findings of Salazar (2011), which suggested that verb-based constructions, especially passive structures with past-tense verbs, be dominant in the Experimental, Materials and Methods of the research articles.

According to Salazar (2011), these past-tense passive structures feature different action verbs, denoting methodological procedures to shift the emphasis from the researcher to the actions. This is to highlight that the methodological steps were taken cautiously and that the results of those steps would be consistent. The findings also supported generalization about research authors utilizing the passive voice constructions to adhere to standardized research protocols (Tarone et al., 1998 as cited in Salazar, 2011).

TABLE 8. Structural Comparison of Lexical Bundles in the Methodology Sub-corpora of English Speaking Writers and Thai Writers

Structure	Raw Frequency		Normalized Frequency		Type		%		LogL
	Eng	Thai	Eng	Thai	Eng	Thai	Eng	Thai	
NP-based	1,085	2,139	5,036.70	6,135.73	85	93	40.87	30.90	+28.61
NP+of	759	1,581	3,523.37	4,535.10	57	68	27.4	22.59	+33.58
NP+other	235	484	1,090.90	1,388.36	21	22	10.10	7.31	+9.44
other NP	91	74	422.43	212.27	7	3	3.37	1.00	-19.41
VP-based	332	1,272	1,541.19	3,648.74	27	74	12.98	24.58	+227.30
passive+PP	102	624	473.5	1,789.95	10	37	4.81	12.29	+207.52
other passive	132	566	612.76	1,623.57	13	31	6.25	10.3	+121.80
we+V	0	0	0	0	0	0	0.00	0.00	0.00
other V fragment	98	82	454.93	235.22	4	6	1.92	1.99	-19.45
PP-based	461	1,430	2,140.02	4,101.96	38	52	18.27	17.27	+162.99
PP+of	306	584	1,420.49	1,675.21	24	19	11.54	6.31	5.55
other PP	155	846	719.53	2,426.75	14	33	6.73	10.96	+249.58
Other	744	1,349	3,314.47	3,869.60	58	82	27.88	27.24	6.62
V/A+to	312	510	1,448.34	1,462.94	20	32	9.62	10.63	0.02
V/N+that cl	30	57	1,39.26	163.5	3	3	1.44	1.00	0.51
As+V	0	57	0	163.5	0	3	0.00	1.00	+54.85
be+AP	20	0	92.84	0	2	0	0.96	0.00	-38.50
anticipatory it	47	58	218.18	166.37	4	2	1.92	0.66	1.88
other AP	10	20	46.42	57.37	1	1	0.48	0.33	0.31
others	325	647	1,508.69	1,855.92	28	41	13.46	13.62	+9.48
Total	2,622	6,190	12,171.63	17,756.03	208	301	100	100	

Based on Table 9, in the Methodology sub-corpus of English speaking writers, research-oriented bundles outnumbered text- and participant-oriented bundles (60.56%, 30.52%, and 8.92%, respectively). The native scholars utilized research-oriented bundles to indicate events, actions, and methods (*n*=76), demonstrate quality, degree and existence (*n*=30), and suggest place, extremity and direction (*n*=11). The results of research-oriented bundles being the most prominent function in non-native writings were aligned with other studies (e.g., Biber, 2009; Biber & Barbieri, 2007; Chen & Baker, 2010; Gungor & Uysal, 2016; Jukneviciene, 2009; Salazar, 2011). For example, in Salazar's (2011) study, research-oriented bundles were the most frequently used bundles in native speaker's corpus at 51.3%, followed by text- and participant-oriented bundles at 42.4% and 6.3%, respectively, which shared the same descending order of functional category of lexical bundles with this study. The inflation of research-oriented bundles in the Methodology sub-corpus of English speaking writers was presumably a result of the characteristic of the given sub-corpus, which included only the methodology section and excluded the other parts of the dissertations. Thus, the majority of lexical bundles in this sub-corpus were research-oriented and mainly indicated methodological procedures. This was consistent with Hyland's (2008a)

observation and Salazar's (2011) findings that the recurring use of research-oriented bundles were representative of the research authors highlighting research by providing accurate descriptions of experimental methods, procedures and equipment to ensure attestable and replicable interpretation of the findings. Likewise, the Methodology sub-corpus of Thai writers produced the highest proportion of research-oriented bundles at 59.4%, followed by text- and participant-oriented bundles at 35.85% and 4.75%, respectively. Thai authors mainly employed research-oriented bundles in the Methodology sub-corpus to express methodological procedures (*n*=117), provide description (*n*=31), suggest place, extremity and direction (*n*=22), and illustrate quantification (*n*=16). The results of research-oriented bundles being the most frequent functional category used by Thai academics in the Methodology sub-corpus disagreed with certain studies on functions of lexical bundles used by non-native scholars (e,g., Chen & Baker, 2010; Gungor & Uysal, 2016; Pan, Reppen, & Biber, 2016; Salazar, 2011). However, the findings were similar to other studies that discovered research-oriented were preferred in academic prose by non-native scholars (e.g., Ädel & Erman, 2012; Biber, 2009; Biber & Barbieri, 2007; Jukneviciene, 2009) and in particular by Thai authors (Panthong & Poonpon, 2020). In Panthong and Poonpon's study, the largest proportion of lexical bundles produced by Thai medical researchers were research-oriented. This could be attributed by the mutual nature of this study's Methodology sub-corpus of Thai academics and Panthong and Poonpon's corpus of Thai medical journals. This study's sub-corpus dissected only the methodological part of dissertations for functional analysis while Panthong and Poonpon's data were gathered from scientific journals in which methodological procedures were heavily concentrated. As a result, research-oriented bundles were mostly employed in both studies.

TABLE 9. Functional Comparison of Lexical Bundles in the Methodology Sub-corpora of English Speaking Writers and Thai Writers

Functions	Raw Frequency		Normalized Frequency		Type		%		LogL
	Eng	Thai	Eng	Thai	Eng	Thai	Eng	Thai	
Research-oriented	1,555	3677	7,218.49	10,547.48	129	192	60.56	59.40	+164.67
Location	183	637	849.51	1,827.24	14	22	6.57	10.29	+94.60
Procedure	969	1,987	4,498.21	5,699.71	76	117	35.68	32.10	+37.44
Quantification	21	326	97.48	935.13	3	16	1.41	5.27	+195.63
Description	321	655	1,490.12	1,878.87	30	31	14.08	10.58	+11.86
Grouping	61	72	283.17	206.53	6	6	2.82	1.16	3.25
Text-oriented	886	2,219	4,112.91	6,365.21	65	109	30.52	35.85	+127.77
Additive	54	99	250.67	283.98	3	5	1.41	1.60	0.55
Comparative	16	66	74.27	189.32	2	4	0.94	1.07	+13.37
Inferential	105	253	487.42	725.73	8	11	3.76	4.09	+12.36
Causative	79	77	366.73	220.87	5	2	2.35	1.24	-9.94
Structuring	176	845	817.01	2,423.88	17	40	7.98	13.65	+213.36
Framing	82	372	380.65	1,067.08	8	17	3.76	6.01	+86.96
Citation	0	31	0.00	88.92	0	3	0.00	0.50	+29.83
Generalization	0	0	0.00	0.00	0	0	0.00	0.00	0.00
Objective	374	476	1,736.15	1,365.41	22	27	10.33	7.69	-11.94
Participant-oriented	181	294	840.22	843.34	19	15	8.92	4.75	0.00
Stance	42	0	194.97	0.00	5	0	2.35	0.00	-80.85
Engagement	19	35	88.20	100.40	2	3	0.94	0.57	0.21
Acknowledgement	120	259	557.05	742.94	12	12	5.63	4.18	+7.02
Total	2,622	6,190	12,171.63	17,756.03	213	316	100	100	

LEXICAL BUNDLES IN THE RESULTS AND DISCUSSIONS SUB-CORPORA OF ENGLISH SPEAKING WRITERS AND THAI WRITERS

Table 10 presents the statistical underuse and overuse of lexical bundles shared by Thai academics and the English native authors ($p<0.01$, critical value 6.63) in the Results and Discussion sub-corpora.

TABLE 10. Comparison of Mutual Lexical Bundles in the Results and Discussions and Sub-corpora of English Speaking Writers and Thai Writers

LBs	Frequency		LogL	LBs	Frequency		LogL
	Eng	Thai			Eng	Thai	
a result of the	76	22	-63.10	on the findings of	34	19	-14.57
a summary of the	34	18	-15.69	on the other hand	13	177	+100.53
a wide range of	32	15	-17.17	participants in this study	41	34	-8.33
are presented in table	12	46	+8.23	recommendations for future research	82	13	-94.52
as one of the	10	43	+9.29	results of this study	42	18	-24.87
as well as the	24	103	+22.19	should be noted that	42	35	-8.43
as well as their	60	26	-35.12	the data from the	11	72	+25.53
as well as to	87	25	-72.61	the development of the	17	59	+8.71
at the beginning of	13	51	+9.50	the effectiveness of the	11	90	+38.25
at the end of	20	122	+40.46	the end of the	17	121	+46.15
at the same time	11	62	+18.95	the fact that the	10	45	+10.40
findings of this study	62	32	-29.58	the findings from the	10	107	+53.90
from the analysis of	66	11	-74.48	the findings in this	29	17	-11.58
in a way that	30	14	-16.18	the findings of the	10	144	+83.80
in addition to the	12	44	+7.25	the implementation of the	14	67	+16.85
in response to the	29	11	-19.36	the majority of the	13	80	+26.76
in terms of the	12	118	+56.61	the results from the	26	126	+32.20
in the process of	38	27	-10.90	the results of the	10	205	+135.63
in the use of	21	194	+89.54	the use of the	11	84	+33.90
is one of the	13	56	+12.14	there was a significant	10	46	+10.96
it is important to	49	26	-22.54	to be able to	11	57	+15.83
it should be noted	10	38	+6.69				

The structural analysis in Table 11 reveals that, in the Results and Discussions sub-corpus, English speaking writers used bundles that were noun-based (45.33%), preposition-based (27.11%), others (22.22%), and verb-based (5.33%), while in that of Thai writers, the bundles were others (35.65%), noun-based (33.26%), preposition-based (16.63%), and verb-based (14.45%), respectively. From the Results and Discussions sub-corpus of English speaking writers, 225 types of lexical bundles were elicited. The most common lexical bundles employed by native speakers were comprised of *as a result of*, *the results of the*, *the purpose of this*, *the results of this*, and *in the area of*. These results are aligned with others (Adel & Erman, 2012; Biber et al., 1999; Biber, Conrad, & Cortes, 2004) which found English speaking writers preferred using noun- and preposition-based bundles in academic prose. In addition, the topmost occurrence of noun-phrase structure in the Results and Discussions sub-corpus of English speaking writers was consistent with earlier findings (Biber et al., 1999; Byrd & Coxhead, 2010; Hyland, 2008b, Salazar, 2011) which discovered that native speakers of English predominantly employed nominal structures in academic writing and strengthened Swales' (2008) concept of academic register being noun-oriented.

Meanwhile, 505 types of four-word lexical bundles were identified in the Thai writers' Results and Discussions sub-corpus. The most frequently utilized lexical bundles in the Thai authors' corpus were made up of *it was found that*, *the results of the*, *in the use of*, *on the other hand*, and *of the present study*. The results from both the Results and Discussions sub-corpora contradicted with previous studies (Gungor & Uysal 2016, Hyland, 2008b; Ozturk, 2004; Perez-Llantada, 2014; Pan, Reppen, & Biber, 2016) that discovered English as a foreign language writers used a fewer types of lexical bundles. Significantly, the inflated number of types of lexical bundles in the Thai writers' Results and Discussions sub-corpus was approximately two times greater than in the number of types of lexical bundles used in the native speakers' corpus. This phenomenon was, however, consistent with the studies conducted by Gungor and Uysal (2016) and Ozturk (2004) that mutually revealed that the number of lexical bundles in non-native writing doubled that of academic prose written by native speakers. According to previous research (Cortes, 2004; Granger, 1998; Kaszubski, 2000; Nesselhauf, 2005, as cited in Salazar, 2011), this circumstance was not unprecedented and did not necessarily reflect the higher linguistic proficiency on the Thai authors' part, in comparison with English speaking writers. In contrast, the extremely frequent occurrence of four-word lexical bundle used by non-native speakers of English, or in this specific case, Thai writers, could be translated as the act of overusing certain bundles by opting for the easiest lexical choices or "lexical teddy bears", as originally coined by Hasselgren (1994).

One observation is made by the researcher in terms of Thai authors having a tendency to overuse particular lexical bundles in their writings, particularly in the results and discussions section, which contributed to extensive lexical repetition, rather than an actual expertise in the use of lexical bundle. An obvious example is the presence of *it was found that*, the first-ranked lexical bundle in the Results and Discussions sub-corpus of Thai authors, which occurred too repetitively only in a few texts throughout the entire sub-corpus. In other words, *it was found that* was narrowly but overly present in a couple of texts and was not equally distributed across the whole corpus. This similar trend of overuse also happened to be applied with other top-ranked four-word lexical bundles in the Discussions sub-corpus of Thai writers. Moreover, the much higher occurrences of four-word lexical bundles in the Thai author's results and discussions sub-corpus could be interpreted as institutional factors such as expectation and practice of Thai universities that expect Ph.D. students to more be critical of the findings and to offer more implication.

In addition to the researcher's personal observation, Salazar (2011) demonstrated a set of lexical bundles that was significantly overused by non-native writers, in comparison with those used by native speakers of English. Among those overused lexical bundles by non-English speaking writers was *on the other hand*, *in the present*, *the present study*, and *as well as*. Similarly to the findings of the current study, these four-word and partial-to-four-word lexical bundles also made its ways to the upper positions of the list of the most frequent lexical bundles identified in the Results and Discussions sub-corpus of Thai academics. The inflated occurrences of four-word lexical bundles in the Results and Discussions sub-corpus of Thai authors evident in the current study are also consistent with the findings of Gungor and Uysal (2016) who performed a comparative analysis on the four-word lexical bundles written by both native English and non-native English scholars. In particular, they identify that four-word lexical bundles *on the other hand*, *the results of the*, *it was found that*, and *at the end of* are the most frequent bundles in the non-native corpus. In a similar vein, all of these most frequent four-word lexical bundles found in Gungor and Uysal (2016)'s corpus of non-native writers were ranked in the top ten of the current study's Results and Discussions sub-corpus of Thai writers. The Results and Discussions sub-corpus, likewise, showed an overuse of the verb-based structures. This excessive use of verbal

construction in the Results and Discussions sub-corpus of Thai academics coincided with the findings of Gungor and Uysal (2016) and Pan, Reppen, and Biber (2016), which found nominal and prepositional structures being outnumbered by verb-based constructions in non-native academic writings. This inflated occurrence of verb-based bundles was viewed as the non-native writer's inability to use noun and prepositional phrase structures (Gungor & Uysal, 2016).

Regarding researches on lexical bundles in Thai context, the inflation of verb-based construction, especially passive structure, in the present study's sub-corpora of Thai writers supported the findings of Panthong and Poonpon (2020) but contradicted with the findings of Leelasetakul (2014), who found a small occurrence of passive verb lexical bundles in Thai author's writings. Leelasetakul discovered that the number of passive verb lexical bundles was limited in the writings of first-year university students, however, the number increased in the composition written by third- and fourth-year university students, enrolled in advanced writing courses. The plausible explanation for the higher number of passive verb lexical bundles in the present study's corpora of Thai writers seemed to be associated with the different levels of proficiency of the target writers under study. Assuming the first-year and third- and fourth-year students enrolled in an advanced writing course in Leelasetakul's study, to be respectively, low- and middle-proficiency level students, the latter group of students, as a result of their proficiency, employed more passive verbs in their writing, in comparison with the former. In the present study, the data were gathered from the writings of Thai authors who are Ph.D. students and it was safe to consider these writers as highly proficient writers. Based on the proficiency levels, high-proficiency authors used a higher number of passive constructions, in comparison with low- and middle-level proficiency writers in Leelasetakul's study. The findings of Panthong and Poonpon (2020) also supported the present study's assumption on the higher occurrence of verb-based structures being indicative of higher level of proficiency. Quite similar to the present study's findings, Panthong and Poonpon (2020) found the passive verb lexical bundles as the most frequently used verbal fragments. Moreover, their data were collected from Thai authors' published medical journals and a higher level of the writing proficiency was expected. In addition, since the corpus data of this study and Panthong and Poonpon (2020) were highly academic, it was not uncommon to see a high occurrence of passive verb lexical bundles, which was also discovered in other previous studies (Biber, Conrad, & Cortes, 2004; Cortes, 2004; Hyland 2008a, 2008b; Kwary et al., 2017).

TABLE 11. Structural Comparison of Lexical Bundles in the Results and Discussions Sub-corpora of English Speaking Writers and Thai Learners

Structures	Raw Frequency		Normalized Frequency		Type		%		LogL
	Eng	Thai	Eng	Thai	Eng	Thai	Eng	Thai	
NP-based	2,239	4,634	3,140.07	6,053.09	102	168	45.33	33.26	+690.81
NP+of	1,673	3,040	2,346.29	3,970.95	71	109	31.56	21.58	+311.05
NP+other	524	1,371	734.88	1,790.85	28	51	12.44	10.1	+334.51
other NP	42	223	58.9	291.29	3	8	1.33	1.58	+123.15
VP-based	150	1,466	210.37	1,914.94	12	73	5.33	14.45	+1,150.02
passive+PP	12	871	16.83	1,137.73	1	41	0.44	8.12	+1,037.17
other passive	35	258	49.09	337.01	3	13	1.33	2.57	+176.33
we+V	0	0	0.00	0.00	0	0	0.00	0.00	0.00
other V fragment	103	337	144.45	440.2	8	19	3.56	3.76	+115.02
PP-based	1,223	2,649	1,715.19	3,460.22	61	84	27.11	16.63	+441.29
PP+of	911	1341	1,277.63	1,751.66	42	46	18.67	9.11	+54.90

other PP	312	1,308	437.56	1,708.56	19	38	8.44	7.52	+589.59
Other	937	4,250	1,314.10	5,551.51	50	180	22.22	35.65	+2,061.46
V/A+to	179	612	251.04	799.42	10	30	4.44	5.94	+220.79
V/N+that cl	211	1,355	295.92	1,769.95	14	59	6.22	11.68	+853.55
As+V	0	282	0.00	368.36	0	7	0.00	1.39	+371.25
be+AP	0	138	0.00	180.26	0	9	0.00	1.78	+181.68
anticipatory it	189	920	265.06	1,201.74	7	27	3.11	5.35	+474.21
other AP	48	237	67.32	309.58	2	11	0.89	2.18	+123.60
others	310	706	434.76	922.2	17	37	7.56	7.33	+131.65
Total	4,549	12,999	6,379.72	16,979.75	225	505	100	100	

As seen in Table 12, the functional analysis of lexical bundles in the Results and Discussions sub-corpus revealed native writers predominantly opted for text-oriented bundles, rather than research- and participants-oriented (47.46%, 41.95%, and 10.59%, respectively). In the Results and Discussions sub-corpus of English speaking writers, text-oriented bundles were employed to signal inferences and conclusions drawn from data (*n*=26), situate arguments by specifying limiting conditions (*n*=24), and mark cause and effect relations between elements (*n*=7). The present results of text-oriented bundles outnumbering other functional categories contradicted with some previous studies (e.g., Biber, 2009; Biber & Barbieri, 2007; Chen & Baker, 2010; Gungor & Uysal, 2016; Jukneviciene, 2009; Salazar, 2011) but were aligned with the findings from the Introduction sub-corpus and those of Demiray Akbulut (2020) and Pan, Reppen, and Biber (2016). The escalation of text-oriented bundles in the Results and Discussions sub-corpus of English speaking writers potentially resulted from the source of corpus data. By separating the result and discussion parts from the other parts of the dissertation, the corpus was heavily themed upon reporting the results and presenting its interpretations. Hence, the indication of inferences and conclusions made from data was frequently found through the use of text-oriented bundles. Furthermore, native scholars largely used text-oriented bundles to discuss the findings because the function was essential in organizing and delivering the arguments in research articles (Gungor & Uysal, 2016). Similar to those of native speakers, lexical bundles in the Results and Discussions sub-corpus of the Thai authors were text-oriented bundles, succeeded by research- and participant-oriented bundles (59.56%, 30.42%, and 10.02%, respectively). Text-oriented bundles were employed by the Thai writers in the Results and Discussions sub-corpus to signal inferences and conclusions drawn from the data (*n*=110), compare and contrast different elements (*n*=56) and introduce text-reflexive markers that organize stretches of discourse or direct the reader elsewhere in the text (*n*=55). That text-oriented bundles were preferred by Thai authors to other functional categories in the Results and Discussions sub-corpus was inconsistent with the earlier research on functions of Thai authors' uses of lexical bundles (Panthong & Poonpon, 2020). This could be influenced by the characteristics of Panthong and Poonpon's corpus of medical journals, which was more experimental and less interpretative. Hence, the text-oriented bundles were less preferable than research-oriented bundles. However, the corpus of the present study was a compilation of a collection of Ph.D. dissertations in the field of English Language Teaching, which, according to Hyland (2008a) was more interpretative, less experimental, and typically distinguished by text-oriented bundles (Hyland, 2008a). The findings, nevertheless, were aligned with other studies on functions of lexical bundles in non-native speaker's academic writing (e.g., Demiray Akbulut, 2020; Gungor & Uysal, 2016; Pan, Reppen, & Biber, 2016). Furthermore, Thai academic's preference of text-oriented lexical bundles could be attributed to the perception of non-native authors that the use of this textual function, rather than others, better represented a higher level of academic expertise. (Hyland, 2012 as cited in Gungor & Uysal, 2016).

TABLE 12. Functional Comparison of Lexical Bundles in the Results and Discussions Sub-corpora of English Speaking Writers and Thai Learners

Functions	Raw Frequency		Normalized Frequency		Type		%		LogL
	Eng	Thai	Eng	Thai	Eng	Thai	Eng	Thai	
Research-oriented	1,806	3,990	2,532.81	5,211.88	99	167	41.95	30.42	+695.74
Location	235	398	329.57	519.88	9	13	3.81	2.37	+31.66
Procedure	397	1,447	556.77	1,890.12	21	55	8.90	10.02	+563.01
Quantification	373	882	523.11	1,152.10	22	38	9.32	6.92	+177.92
Description	516	848	723.66	1,107.69	31	42	13.14	7.65	+59.75
Grouping	285	415	399.70	542.09	16	19	6.78	3.46	+15.93
Text-oriented	2,281	8,086	3,198.97	10,562.22	112	327	47.46	59.56	+3,046.69
Additive	145	242	203.35	316.11	6	6	2.54	1.09	+18.17
Comparative	188	1,326	263.66	1,732.07	11	56	4.66	10.2	+883.90
Inferential	644	3,536	903.17	4,618.85	26	110	11.02	20.04	+2,002.17
Causative	284	393	398.29	513.35	7	17	2.97	3.10	+10.73
Structuring	239	1,168	335.18	1,525.68	19	55	8.05	10.02	+604.01
Framing	444	639	622.69	834.68	24	25	10.17	4.55	+22.81
Citation	33	297	46.28	387.95	2	27	0.85	4.92	+224.58
Generalization	0	27	0.00	35.27	0	3	0.00	0.55	+35.55
Objective	304	458	426.34	598.26	17	28	7.20	5.10	+21.36
Participant-oriented	462	923	647.93	1,205.65	25	55	10.59	10.02	+125.40
Stance	123	380	172.50	496.37	11	29	4.66	5.28	+120.09
Engagement	194	207	272.07	270.39	9	16	3.81	2.91	0.00
Acknowledgement	145	336	203.35	438.89	5	10	2.12	1.82	+65.01
Total	4,549	12,999	6,379.72	16,979.75	236	549	100	100	

CONCLUSION AND PADAGOGICAL IMPLICATIONS

The findings revealed that Thai authors overused lexical bundles in comparison with English speaking writers in each part of their respective dissertations, especially in the results and discussion section, which could result from institutional factors such as expectation and practice of Thai universities that expect Thai Ph.D. students to more be critical of the findings and to offer more implications. The structural analysis revealed the overuse of verb-structured lexical bundles throughout the three sub-corpora of Thai academics, which plausibly stemmed from the non-native speaker's failure to employ noun- and preposition-structured lexical bundles effectively. The proportion of functions of lexical bundles in each section of the dissertations written by both groups of writers shared a relatively similar trend, indicating that English speakers and Thai writers conformed to the same convention of dissertation writing.

Byrd and Coxhead (2010) support the utilization of a list of lexical bundles as a foundation to instructional design and curriculum development. However, they assert that instructors and authors should be adequately informed of the way the list is created and compiled. Based on this view, the list offered by the current study may fulfil all the requirements for choosing lexical bundles for instructional practices. This is because all of the prerequisite information, ranging from the methodological steps into creating the target corpora, the textual description from which the list is generated, its representation of the language essential to students, and the method of elicitation being used, is presented. In addition, grouping lexical bundles will simplify the task of instructors when considering for which lexical bundles are needed. Certain lexical bundles should

be included in teaching materials, textbooks, or learner dictionaries in the form of manageable lists of expressions coordinated by a single function.

LIMITATIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

The current study poses some limitations that could be addressed in future research. First, the present study focuses only on four-word lexical bundles since they are more common (Biber et al., 1999) and representative of a wide range of functions and structures (Hyland, 2008b), as previously shown in the previous chapter. However, the three- or five-word length lexical bundles might also be worth investigating, particularly for smaller corpora where time constraint is not a pressing issue, because those lexical bundles are also structurally and functionally classifiable. Second, the current study follows Biber et al.'s (1999) structural and Hyland's (2008a) functional taxonomies, adapted by Salazar (2011). Despite their overall thoroughness, both taxonomies are yet to fully cover several minor aspects of lexical bundles emerging from the data, particularly in terms of structures. More refined and newly revised structural and functional frameworks might be needed for future researchers in order to investigate the lexical bundles more thoroughly. Third, the corpus size of 1,000,000 words may not be fully generalizable for how four-word lexical bundles are actually employed by English speaking writers and Thai academics in ELT dissertation writings. As a result, larger corpora might be more representative of this particular written register between the English speaking writers and Thai authors. In addition, the creation of the corpora understudy was based on its availability and accessibility online and gathered from dissertations published by top-tier universities in English speaking countries and Thailand. Inclusions of ELT dissertations from lower- or middle- tier universities, both domestic and from abroad, might allow researchers to come to a better generalization on how four-word lexical bundles are used among the two groups of writers.

REFERENCES

Adel, 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

Anthony, L. (2017). AntFileConverter (Version 1.2.1) [Computer Software]. Waseda University, Tokyo, Japan.

Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Waseda University, Tokyo, Japan.

Bavdekar, S. B. (2015). Writing Introduction: Laying the Foundations of a Research Paper. *Journal of the Association of Physicians of India*, *63*, 44-46.

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

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

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., & Quirk, R. (1999). Longman grammar of spoken and written English. Harlow: Longman.

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*(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.

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

Davis, M., & Morley, J. (2018). Facilitating learning about academic phraseology: Teaching activities for student writers. *Journal of learning development in higher education*. Special Edition, 1–17. https://doi.org/10.47408/jldhe.v0i0.468

De Cock, S. (2003). Recurrent sequences of words in native speaker and advanced learner spoken and written English. Unpublished Ph.D thesis, Catholic University of Louvain, Belgium.

Demiray Akbulut, F. (2020). A bibliometric analysis of lexical bundles usage in native and non-native academic writing. *Journal of Language and Linguistic Studies*, *16*(3), 1146-1166. https://doi.org/10.17263/jlls.803583

Durrant, P. (2015). 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

Erman, B. (2009). Formulaic language from a learner perspective: What the learner needs to know. In B. Corrigan, H. Quali, E. Moravcsik, & K. Wheatley (Eds.). *Formulaic language* (pp. 27-50). Amsterdam: John Benjamins.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.). *Phraseology: Theory, analysis, and applications* (pp. 145-160). Oxford: Oxford University Press.

Greaves, C. & Warren, M. (2010). What can a corpus tell us about multi-word units? In A. O'Keeffe & M. J. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 212–226). London: Routledge.

Gungor, F., & Uysal, H. H. (2016). A comparative analysis of lexical bundles used by native and non-native Scholars. *English Language Teaching*, *9*(6), 176-188. https://dx.doi.org/10.5539/elt.v9n6p176

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, *4*(2), 237–258.

Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.). *Phraseology: Theory, analysis, and applications* (pp. 161-186). Oxford: Oxford University Press.

Hyland, K. (2006). The 'other' English: Thoughts on EAP and academic writing. *The European English Messenger*, *15*(2), 34-38.

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.

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

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

Jha K. N. (2014). How to Write Articles that Get Published. *Journal of clinical and diagnostic research: JCDR*, *8*(9), XG01-XG03. https://doi.org/10.7860/JCDR/2014/8107.4855

Jukneviciene, R. (2009). Lexical bundles in learner language: Lithuanian learners vs. native Fspeakers. *KaLBOTYRa*, *61*, 61-72. https://doi.org/10.15388/Klbt.2009.7638

Kachru, B. (1985). Standards, codification and sociolinguistic realism: the English language in the outer circle. In R. Quirk & H.G. Widdowson (Eds.). *English in the world: Teaching and learning the language and literatures* (pp. 11-30). Cambridge: Cambridge University Press.

Kaszubski, P. (2000). Selected aspects of lexicon, phraseology and style in the writing of Polish advanced learners of English: A contrastive, corpus-based approach. Unpublished Ph.D thesis, Adam Mickiewicz University, Poznan, Poland.

Kwary, D., Ratri, D., & Artha, A. (2017). Lexical Bundles in Journal Articles across Academic Disciplines. *Indonesian Journal of Applied Linguistics*, *7*(1), 131-140. https://doi.org/10.17509/ijal.v7i1.6866

Leelasetakul, M. (2014). Exploring lexical bundles in Thai EFL learners' writings. *Thoughts*, 1, 101-125.

Nesselhauf, N. (2005). Collocations in a learner corpus. Amsterdam: John Benjamins.

Ozturk, Y. (2014). Lexical bundle use of Turkish and native English writers: A corpus-based study. Unpublished M.A. thesis. Anadolu University, Eskişehir, Turkey.

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.

Panthong, P., & Poonpon, K. (2020). Lexical Bundles in Thai Medical Research Articles. *Journal of Studies in the English Language*, *15*(1), 59-106.

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

Salazar, D. (2011). Lexical bundles in scientific English: A corpus-based study of native and non-native writing. Unpublished Ph.D thesis, University of Barcelona, Spain.

Sinclair, J. (1991). Corpus, concordance, collocation. Oxford: Oxford University Press.

Swales, J. (1981). Aspects of article introductions. Birmingham, England: University of Aston.

Swales, J. (2004). Research genres: Exploration and applications. Cambridge: Cambridge University Press.

Swales, J. (2008). Foreword. In D. Belcher & A. Hirvela (Eds.). *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions* (p. 13). Ann Arbor: University of Michigan Press.

Tarone, E., Dwyer, S., Gillette, S., & Icke, V. (1998). On the use of the passive and active voice in astrophysics journal papers: With extensions to other languages and other fields. *English for Specific Purposes*, *17*(1), 113-132.

ABOUT THE AUTHORS

Nateethorn Narkprom is a Ph.D. student in the English Language Teaching program at Language Institute of Thammasat University. His research interest is Corpus Linguistics.

Supakorn Phoocharoensil is an Associate Professor of English at Language Institute of Thammasat University, Bangkok, Thailand. He is also the Editor-in-Chief of LEARN Journal: Language Education and Acquisition Research Network, indexed in SCOPUS and ERIC. His areas of research interest include Second Language Acquisition, Corpus Linguistics, and English Collocations and Formulaic Language.

eISSN: 2550-2131
ISSN: 1675-8021