The Pennsylvania State University

The Graduate School

Department of Applied Linguistics

LEXICAL BUNDLE USE BY THAI

UNDERGRADUATE LEARNERS OF ENGLISH

A Dissertation in

Applied Linguistics

by

Mathurin Leelasetakul

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2019
The dissertation of Mathurin Leelasetakul was reviewed and approved* by the following:

Xiaofei Lu  
Associate Professor of Applied Linguistics and Asian Studies  
Dissertation Adviser  
Chair of Committee  
Director of Graduate Studies, Department of Applied Linguistics  

Suresh Canagarajah  
Edwin Erle Sparks Professor of Applied Linguistics, English, and Asian Studies  

Ning Yu  
Professor of Applied Linguistics and Asian Studies  

Matthew E. Poehner  
Associate Professor of World Languages Education and Applied Linguistics  

*Signatures are on file in the Graduate School.
This present study explores the use of lexical bundles in the writings of Thai undergraduate learners of English. The study uses a Thai learners’ corpus, TULE2014, compiled from writing assignments and tests of two levels of undergraduate students: first-year students enrolled in a freshman English course (Freshman English group) and third-to fourth-year students in an advanced writing course (English Composition group). The analysis is divided into two parts. The first part explores the use of frequent 4-word lexical bundles by both groups using the frequency-based approach and the structural and functional categories from Biber et al. (2004). The first analysis reveals that the students in the novice Freshman English subcorpus used more lexical bundle types and tokens than the expert students in the Composition subcorpus. The Freshman English students also used more lexical bundles that are not originally-produced than their Composition counterparts. The second part of the analysis uses the Academic Formula List (AFL) from Simpson-Vlach and Ellis (2010) as target bundles and examined the use of those target bundles in the learner corpus. The result shows that the Freshman English group used fewer AFL bundles than the Composition group. The Freshman English group also used more spoken bundles than written ones and the Composition group used more written than spoken bundles in their writing. The analysis provides insights into the changes that can be made to the class materials and class discussion in order to address the issues identified.
# TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................. vi

ACKNOWLEDGEMENTS ...................................................................................................... vii

Chapter 1  Introduction ........................................................................................................ 1

1.1 Background .................................................................................................................. 1
1.2 Focus of the study ....................................................................................................... 3
1.3 Dissertation organization .......................................................................................... 4

Chapter 2  Lexical Bundles ................................................................................................. 6

2.1 Lexical bundles .......................................................................................................... 6
2.2 Cross-register and cross-discipline differences in lexical bundle use .................... 10
2.3 Lexical bundles and the development of second language writing proficiency .......................................................... 12
2.4 Lexical bundles and instruction ............................................................................... 19
2.5 The Academic Formula List .................................................................................... 21
2.6 Alignment in Second Language Acquisition ............................................................ 22
2.7 The present study ..................................................................................................... 24

Chapter 3  Research Methodology .................................................................................... 26

3.1 Research design ........................................................................................................ 26
3.1.1 Data sources ....................................................................................................... 26
3.1.2 The corpus ........................................................................................................ 32
3.1.3 Data analysis ..................................................................................................... 34
3.2 Research questions .................................................................................................. 37

Chapter 4  Lexical Bundles in the Learner Corpus ............................................................ 38

4.1 Overall results .......................................................................................................... 38
4.2 Lexical bundle structures ......................................................................................... 41
4.2.1 Lexical bundles that incorporate verb phrase fragments .................................. 42
4.2.2 Lexical bundles that incorporate dependent clause fragments ....................... 44
LIST OF TABLES

Table 3.1: Writing prompts for English I writing tasks ............................................... 30
Table 3.2: Writing prompts for Composition writing tasks .......................................... 32
Table 3.3: Total word count of the TULE2014 Corpus .............................................. 33
Table 3.4: Freshman English subcorpus word count ................................................. 33
Table 3.5: Composition subcorpus word count ......................................................... 33
Table 4.1: Overall comparison of lexical bundles for Freshman English and Composition ................................................................................................................. 38
Table 4.2: Lexical bundle structures .......................................................................... 41
Table 4.3: Lexical bundle functions ........................................................................... 49
Table 4.4: Prompt-specific lexical bundles ............................................................... 59
Table 4.5: Genre-specific lexical bundles ................................................................... 63
Table 4.6: Originality of lexical bundles .................................................................... 68
Table 5.1: AFL bundles in Freshman English and Composition subcorpora .......... 90
Table 5.2: Referential expressions .......................................................................... 93
Table 5.2: Stance expressions .................................................................................. 107
Table 5.3: Discourse organizing functions .................................................................. 114
I would like to first express my gratitude to Her Majesty Queen Sirikit of Thailand, whose scholarship, the Queen Sirikit Scholarship, has given me this opportunity to further my education in the US.

I would like to thank my dissertation advisor, Dr. Xiaofei Lu, for his guidance and support since the first day I approached him about this dissertation. I am also indebted to my dissertation committee, Dr. Suresh Canagarajah, Dr. Matthew Poehner and Dr. Ning Yu, for their advice on the dissertation. I greatly appreciate the feedback on my research from my committee.

I have received generous support from my colleagues, several of whom are also my mentors and former teachers, at the Faculty of Arts, Chulalongkorn University and for that I am forever thankful.

Thank you to my friends in the APLNG program at Penn State who are my source of motivation and inspiration. You are all amazing people and I am glad I had the chance to know you.

I also found myself a home away from home in the Thai community in State College. I owe a big thank you to my friends here in this college town. Thank you for being welcoming since day one and supportive of me until my last say in State College.

Last but not least, I am eternally grateful to have a great support system back home in Thailand from my family and friends. Thank you, mom and dad, for always believing in me. Thank you, Mace. You are the best sister one could ever ask for and you are my superhero.
CHAPTER 1

Introduction

1.1 Background

It has been established that language in use consists of a substantial proportion of what is called formulaic sequences (Nattinger & DeCarrico, 1992; Schmitt & Carter, 2004). Wray (2002) defined formulaic sequences as “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (p. 465). As it happens, the term formulaic sequences is only one of the terms that have been used by various researches to refer to this aspect of language. Wray (2002) has also compiled an extensive list of terms that have been used to refer to this phenomenon, ranging from “collocations”, “sentence builders”, “lexical or lexicalized phrases”, “multiword units” to “stock utterances”. The use of these prefabricated chunks of language is prevalent both in speech and in writing.

Since formulaic sequences are “store and retrieved whole” (Wray, 2002, p. 465), these language units are processed much quicker in reading tasks than non-formulaic ones in both native speakers and in second language learners as seen from reaction time in experiments (Jiang & Nekrasova, 2007). In terms of production, with the focus on writing, several studies have shown that students benefited from instructions of formulaic sequences (AlHassan & Wood, 2015; Peters & Pauwels, 2015). As such, this feature of language is important in second language acquisition and instruction and it is important for
second language learners to master the use of these formulaic sequences (Nattinger & DeCarrico, 1992; Wray, 2002).

One aspect of formulaic sequences that is of interest in this present study is a multi-word expression that is called lexical bundles. Lexical bundles, usually three or four words in length in a majority of the studies, are “sequences of words that commonly go together in natural discourse” in either conversation or writing (Biber & Conrad, 1999, p. 184) and they are defined and extracted from the corpus using the frequency-based approach.

Corpus-based studies on lexical bundles have shown that lexical bundles are prevalent in conversation and in academic texts (Biber, Conrad & Cortes, 2004). In academic writing, it is found that there are general academic lexical bundles that are generally used across different academic disciplines as seen in the research into the Academic Formula List (Simpson-Vlach & Ellis, 2010). In discipline-specific studies, comparisons are made across academic disciplines, within the same academic discipline but across different groups of participants (native-non-native or expert-novice), or a combination of different disciplines and different levels of participants. The results of the study showed that, in different academic disciplines and in different groups of participants, frequent lexical bundles are not the same or are not used in the same way in terms of their functions (Cortes, 2004; Hyland, 2008a).

Since lexical bundles are prevalent in naturally occurring discourse across different registers and disciplines, they are essentials to language learners since they function as “lexical building blocks” with distinct functions, such as discourse organizers, that helps with production and comprehension of English (Biber & Conrad, 1999, p. 185). Use of lexical bundles also signal group membership, such being a novice or an expert, through
mastery of lexical bundles in a register or a discipline (Hyland, 2008a). Several studies found that novice writers (undergraduate or graduate students) use lexical bundles differently from the experts (published authors).

English language instruction in Thailand has put the main focus on grammatical competence; however, it is evident that multi-word expressions such as lexical bundles are frequent and important features of English.

From previous research, the importance of formulaic language such as lexical bundles is clear. However, this area still requires more research with Thai learners of English, especially in academic writing. This is the focus of the current study, which is discussed in more detail in the following section.

1.2 Focus of the study

This dissertation focuses on lexical bundle use in the writings of Thai undergraduate learners of English. The corpus data used in this study is collected in the first semester of the academic year 2014 at the department of English, faculty of Arts in a university in Thailand. In the undergraduate curriculum, the department of English offers five writing courses: English I, English II, English Composition I, English Composition II and Research Writing. Each course is a pre-requisite for the subsequent courses. The Thai undergraduate students in this study is students who enrolled in the course English I (referred to as the Freshman English group in this dissertation) and English Composition II (referred to as the Composition group). The corpus data is collected from the writing assignments and tests throughout one semester.
The analysis is divided into two parts. The first part explores the use of 4-word lexical bundles by both groups using the frequency-based approach and the structural and functional category from Biber at al. (2004). The second part of the analysis uses the Academic Formula List (AFL) from Simpson-Vlach and Ellis (2010) as target bundles and examined the use of those target bundles in the learner corpus. Both parts of the analysis explore the lexical bundles usage by the students in the Freshman English and Composition subcorpora as well as similarities and differences between the two groups.

1.3 Dissertation organization

There are six chapters in this dissertation. Chapter 1 is this present introduction. Chapter 2 focuses on lexical bundles and previous research in the area. Chapter 3 is the research methodology chapter where I discuss the data collection process and the details of the TULE2014 learner corpus that is used in the analysis, as well as outlining the analytical steps.

The results of the analysis are presented in Chapters 4 and 5. In Chapter 4, I present the results of the analysis of the frequent lexical bundles in Freshman English and the Composition subcorpora, with the following sections: overall results, lexical bundle structures, lexical bundle functions, prompt-specific bundles, genre-specific bundles, originality of lexical bundles, shared-unshared bundles and discussion. In Chapter 5, I compare the lexical bundles in the writing of participants in the Freshman English and Composition subcorpora with the target bundles from the Academic Formulas List (AFL) by presenting overall AFL lexical bundle results, followed by bundles in different functional categories and discussion.
Chapter 6 is the conclusion where I outline summary of findings, pedagogical implications, limitations and directions for future research.
CHAPTER 2
Lexical Bundles

In this chapter, I first discuss how lexical bundles have been defined and operationalized in previous research. I then review previous studies that assessed cross-register or cross-discipline differences in lexical bundle use, examined the relationship of lexical bundles to the development of second language writing proficiency, or designed pedagogical interventions to facilitate learner acquisition of lexical bundles. I conclude with a discussion of the importance of lexical bundles in English language learning and teaching.

2.1 Lexical bundles

Lexical bundles, sometimes also referred to as *n*-grams, chunks or chains, are sequences or strings of lexical items that occur frequently in a particular corpus or a collection of texts in a particular register. Biber and Conrad (1999) stressed that such bundles are “sequences of words that commonly go together in natural discourse” in either conversation or writing (p. 184) and they are considered to be “extended collocations” (p. 183) given the fact that the lexical items in a bundle tend to co-occur together.

In terms of length, lexical bundles may range from 2-word sequences to sequences of 6 or more words. The majority of studies on lexical bundles, however, focus on 4-word sequences, noting that 3-word lexical bundles are often part of longer 4-word lexical bundles and that longer bundles tend to be rare in corpora. For example, the three-word
bundle “I don't think” is part of such following 4-word bundles as “but I don't think”, “well I don't think”, and “I don't think so” (Biber et al., 1999, p. 990).

As part of their research into formulaic language, O'Donnell et al. (2013) explored two ways of identifying and extracting sequences of consecutive lexical items from corpus samples. The first method used frequency of occurrences of the lexical bundles and the resulting bundles from this approach were called “frequency-grams” (p. 96). The other method was association-based and used Mutual Information (MI) to extract what were called “MI-grams” (p. 97). Rather than using the actual frequency of occurrences, this method employed the MI score to extract lexical bundles with “stronger association between the words” (O'Donnell et al., 2013, p. 90). In this current study, the frequency-based approach is used to identify and extract lexical bundles from the corpus.

In previous research, lexical bundles are usually “identified empirically” (Biber & Conrad, 1999, p. 184; Cortes, 2004, p. 400). That is to say, researchers do not embark on a study on lexical bundles with specific sets of lexical bundles in mind. Rather, the lexical bundles that are extracted and analyzed represent the most frequent lexical sequences in the corpus being examined. Lexical bundle studies are thus “frequency-driven” (Biber and Barbieri, 2007, p. 267) in which frequent word combinations are selected for analysis. The frequency cut-off points (normalized frequency per million words) would depend on the researchers and are to a certain degree arbitrary; for example, Hyland (2008a) and Cortes (2004) used the frequency of 20 per million, whereas Biber et al. (2004) used the frequency of 40 per million words. Studies that intend to compare their results against those reported in previous studies, however, would generally adopt a similar cut-off point.
Not all potential lexical bundles that are above the frequency cut-off points are included in the analysis. In most cases, the word sequences that are extracted from a corpus are filtered using an additional set of criteria. Some of the most commonly used criteria are discussed below.

First, as Biber and Conrad (1999) emphasized, word sequences that are considered lexical bundles must occur uninterrupted together as a unit. Therefore, word sequences that cross the sentence boundary are excluded.

Second, word sequences must occur in multiple texts and must not be produced by a single participant. This is to prevent “idiosyncratic uses by individual speakers or authors” (Biber & Barbieri, 2007, p. 268). For example, in the study by Biber and Conrad (1999), a lexical bundle must occur in 5 different texts in the corpus.

Third, overlapping lexical bundles need to be merged. Chen and Baker (2010) identified two types of overlapping bundles: “complete overlap” and “subsumption” (p. 33). In complete overlap, two 4-word bundles are both part of the same five-word bundle. An example from my pilot study based on the analysis of the ICNALE corpus is the 4-word bundles “there is nothing worse” and “is nothing worse than”. Both appear four times in the native speaker sub-corpus and both are part of the 5-word bundle “there is nothing worse than”. The 5-word bundle is thus coded as “there is nothing worse + than”, with a raw frequency of 4. The other type of overlap, “subsumption”, refers to cases where one of the overlapping 4-word bundles is part of the other 4-word bundle. An example from the pilot study is the bundles “it is a good” and “is a good idea”. “It is a good” occurs 14 times in the ICNALE corpus and “is a good idea” appears 6 times. All occurrences of “is a good idea” are part of the longer string “it is a good idea”. Cases of subsumption are marked
In cases of corpora that consist of texts from a specific discipline, lexical bundles that are discipline specific are sometimes excluded from the analysis, such as *second language acquisition research*, or analyzed separately from the more general lexical bundles.

Lexical bundles are usually categorized in terms of structure and function. Structurally, they are usually grouped according to whether they incorporate a certain grammatical structure or not, even though one of the main characteristics of lexical bundles is that they are structurally incomplete and therefore do not conform to traditional grammatical categories such as verb phrase or noun phrase (Biber & Conrad, 1999, p. 183). In their study on lexical bundles in textbooks and lectures, Biber, Conrad and Cortes (2004) proposed a taxonomy that categorized lexical bundles into three broad groups: those that incorporate (1) verb phrase fragments, (2) dependent clause fragments, and (3) noun phrase and prepositional phrase fragments. Chen and Baker (2010), in their comparison of native and non-native speaker bundles, grouped lexical bundles slightly differently into three groups: verb-phrase-based, noun-phrase-based and prepositional-phrase-based. Functionally, lexical bundles have been grouped into three categories: stance expressions, discourse organizers and referential expressions (Biber, Conrad & Cortes, 2004; Chen & Baker, 2010; Ädel & Erman, 2012). Studies that use spoken corpus data, such as the study by Biber, Conrad and Cortes (2004), may include an additional functional category, namely, bundles that fulfill a special conversational function. Hyland (2008b) made use of
three categories for lexical bundles in his analysis of what he calls “academic clusters”: “research-oriented,” “text-oriented” and “participant-oriented” (p. 42).

2.2 Cross-register and cross-discipline differences in lexical bundle use

A large body of research has compared lexical bundle use across different registers or disciplines. With regard to register, Biber et al. (1999) compared lexical bundles in conversation and “academic prose” (p. 993) and found that although both registers yield a large number of lexical bundles both in terms of types and tokens, there are more types of bundles in conversation than in academic prose.

In their comparison of lexical bundles in spoken and written university registers, using corpus data ranging from office hours to written syllabi, Biber and Barbieri (2007) found that “lexical bundles are even more prevalent in non-academic university registers than they are in the core instructional registers” and also that lexical bundles are “very common in written course management (e.g., course syllabi), in contrast to previous research which showed bundles to be much more common in speech than in writing.” (p. 263). From their analysis, they concluded that “the use of lexical bundles is influenced by both mode and communicative purpose” (p. 282).

The second strand of lexical bundle research, genre-based lexical bundle studies, has to do with academic disciplines. Comparisons of lexical bundles are made across academic disciplines, within the same academic discipline but across different levels of participants (native-non-native or expert-novice), or a combination of different disciplines and different levels of participants.

Cortes (2004) compared lexical bundle use between history and biology texts, using
data from published articles and student writings. From the published corpus, it is found that the most frequent lexical bundles from the two academic disciplines are different. Structurally, lexical bundles in history are limited to noun-phrase based and prepositional-phrase based structures, whereas in biology, lexical bundles cover a wider range of structures. In terms of lexical bundle functions, different bundles are used for similar functions; e.g. time reference bundles are used for years of historical events in history and for different stages in development for biology. It is also found that in biology there are several instances of “epistemic–impersonal/probable–possible stance markers” such as “are likely to be” or “it is possible that” (p. 411); however, these bundles are not prevalent in the history corpus. This study has shown that the use of lexical bundles varies depending on the disciplines.

Hyland (2008a) also investigated lexical bundles in different academic disciplines. Using the written data from published authors (research articles) and students writing (doctoral dissertation and master’s theses) from the following academic disciplines: electrical engineering (EE), microbiology (Bio), business studies (BS) and applied linguistics (AL). Hyland’s analysis revealed that overall most bundles are noun- and prepositional-phrase based, similar to what Cortes (2004) discovered in her study. As for similarities and differences across disciplines, the four disciplines differ in terms of ranges and the number of lexical bundles that are found, with engineering at the most and biology at the least. Lexical bundles in the four disciplines are also structurally and functionally different. In short, lexical bundles have been found to be prevalent in naturally occurring discourse across different registers and disciplines. Functionally, they are essential because they serve as “lexical building blocks” for speakers and writers alike (Biber and Conrad,
1999, p. 185) and they serve such functions as “expression of stance, discourse organization, and referential framing” (Biber & Barbieri, 2007, p. 283), which aid both production and comprehension. Biber and Barbieri (2007) even stressed that they are considered pre-fabricated items because of their frequency. Hyland (2008a) also points out another importance of lexical bundles that they signal group membership through mastery of lexical bundles in terms of structures and functions in a register or a discipline. Therefore, if leaners are not fluent or competent in their use of bundles, they would be regarded as “novice or newcomer” or even “outsiders” (Hyland, 2008, p. 5).

2.3 Lexical bundles and the development of second language writing proficiency

Lexical bundles have also been studied in relation to the production of different groups and levels of participants, such as between native and non-native speakers (Römer, 2009; Chen & Baker, 2010; Ädel & Erman, 2012) and between expert or published and novice or student writers (Römer, 2009). Several lexical bundles studies look into production of lexical bundles by participants of different levels of proficiency. The first strand examines lexical bundle production by novice and expert writers in certain academic disciplines. Novice writers in this case are students (either undergraduate, graduate or both) and expert writers are authors of published articles- not necessarily native speakers of English.

Cortes (2004) examined lexical bundle production in published and student writings in the fields of history and biology. The student sub-corpus in her study consists of three levels: two undergraduate levels (lower and upper) and one graduate level. Their production of lexical bundles is compared against lexical bundles produced by expert
authors in the disciplines (these are referred to as target bundles). The target bundles were analyzed first from the published sub-corpus and those target bundles are then identified in the student sub-corpus. For both history and biology students, it is found that “many lexical bundles frequently used by published authors were rarely or never used by students” (p. 412). When students used lexical bundles that are also used by their expert counterparts, they are used for different functions. The example that is given is the lexical bundle “at the same time” in history, which experts used for its literal meaning as opposed to the metaphorical meaning used by the students. Cortes stressed that the mismatch between expert and novice use of lexical bundles, in terms of both frequency and function, does not improve with proficiency (if we were to establish that graduate students are more proficient and more developed). Graduate students are also not “target-like” in their use of lexical bundles (p. 413). From this, Cortes concluded that exposure to reading materials that contain target bundles alone is not enough to improve students’ ability to use those bundles in the same way that the experts do.

In a similar manner as Cortes, Hyland (2008b) compared the use of lexical bundles (termed “academic clusters” in his study) in published and student writings in four academic disciplines: electrical engineering, microbiology, business studies and applied linguistics. Hyland’s student participants are higher in terms of their academic level than those in Cortes’ study, as the student sub-corpus is collected from doctoral dissertations and master’s theses. Each sub-corpus was analyzed separately for their most frequent lexical bundles. The result of frequency analysis of lexical bundles is in line with what Cortes found: “many of the clusters most frequently used in published academic writing were never, or only rarely, found in the student texts” (p. 50) and vice versa. The lexical
bundles produced by the three groups of participants also differ in terms of structural distribution and functions.

In addition to the expert-novice dichotomy, native and non-native speaker statuses are also factored in for several studies. Römer (2009) looked at 3-word and 4-word lexical bundles (called “n-grams” in her study) in academic writing of three participant groups: non-native speaker students, native speaker students and expert, published authors, who are primarily native speakers. The student writings are from two academic disciplines—linguistics and English literature—whereas the published writings are from the field of linguistics. The sizes of the three corpora are similar, around 200,000 words. The focus of the analysis is on the 20 most frequent 3-word bundles and 4-word bundles. The analysis showed that there is “a considerable overlap” (p. 98) between the lexical bundles that are produce by the native and non-native students. It is also found that the two student groups “differ in similar ways from expert academic writers” (p. 99). Römer (2009) then concluded that novice-expert statuses are the more important factors to consider than native and non-native statuses when it comes to academic writing of advanced learners.

Chen and Baker (2010) investigated lexical bundles in writings by published authors, who are native speakers, and students from various disciplines. The student sub-corpus consists of writings by who are Chinese learners of English and by students who are native speakers of English. The analysis shows comparisons between expert and novice as well as native and non-native speakers. The overall analysis shows that expert writers use more types of lexical bundles than the novice student writers and the Chinese learners used the least types of lexical bundles (p. 30). Chen and Baker concluded that there is a correlation between writing proficiency and the use of lexical bundles. The most proficient
of the three groups—the expert writers—made use of the most types and tokens of lexical bundles whereas the least proficient—the Chinese learners—is at the other end of the spectrum. Chen and Baker also noted that non-native and native student are “similar” when it comes to the lexical bundles they use (p. 44), which is also what Römer (2009) found in her study.

In their study, which was aimed to be comparable to Chen and Baker’s, Ädel and Erman (2012) compared lexical bundles in writings of native speakers and advanced-level Swedish learners of English. Both groups of participants are university-level students and are in the same academic discipline of linguistics. Following similar analytical steps as Chen and Baker, Ädel and Erman found that native speakers used more types of lexical bundles than their non-native counter parts. This is contrary to what Römer (2009) and Chen and Baker (2010) found in their studies where students are found to be similar in their use of lexical bundles, regardless of their native or non-native statuses. However, Ädel and Erman concluded that this might be accounted for by the fact that the non-native students in their study are from the EFL context, not ESL like in the previous studies, or from the different sizes of the native and non-native corpora in their study.

Wei and Lei (2011) compared the production of 4-word lexical bundles by advanced Chinese EFL learners (doctoral dissertations) to published authors in the discipline of applied linguistics. The analysis revealed that the Chinese EFL learners use more types and tokens of lexical bundles than the expert authors, which is similar to what have been found by other researches like Cortes (2004) and Hyland (2008a). Similarities between the EFL learners and expert writers can also be found in the structures of the lexical bundles and certain functions of lexical bundles (research-oriented and text-oriented
bundles). The EFL learners differ from the expert writers in their overuse of passive structures, underuse of anticipatory *it* structures and under use of participant-oriented bundles in comparison to the native speaker production. Wei and Lei concluded that the findings pointed to the possible focus of lexical bundle instruction for Chinese learners.

Like Wei and Lei (2011), Jalali, Eslami Rasekh and Tavangar Rizi (2008) explored lexical bundles used in the field of applied linguistics by comparing published journal articles with master’s theses and doctoral dissertations of Iranian EFL writers. The analysis of types and tokens of lexical bundles revealed that the EFL writers used more types and tokens of lexical bundles in their writings. Although the structures of the lexical bundles produced by the experts and the novice EFL writers are both predominantly phrasal, the proportions of the types of phrasal lexical bundles differ between the two groups. The EFL writers made use of almost twice as many noun phrase bundles. Function-wise, both groups of writers used more research-oriented bundles than the other two types. However, similar to the results on lexical bundle structures, the EFL writers used more types of research-oriented lexical bundles than the experts, which is also what Hyland (2008b) discovered in his study. Jalali et al. cautioned that although the novice writers make use of more lexical bundles than the expert counterparts, this does not mean that the novice writers are proficient in their writing; rather, it indicated that they have yet to master and appropriate the types of lexical bundles that are actually used by experts in the field. Thus, the findings of this study support the need for instruction that focuses on helping learners develop and master the use of lexical bundles that are specific to their academic discipline.

Weinstein (2011) investigated lexical bundle use by two levels of students, novice and upper, in the field of applied linguistics and discourse studies (ALDS). The novice
corpus consisted of papers from an introductory course in ALDS written by first to fourth-year undergraduate students and students in a one-year diploma program. The upper level corpus consists of papers written by graduate students and fourth-year undergraduate students, who received grade A or higher. Both proficiency level groups consist of native and non-native speakers. The analysis of the lexical bundles revealed that the novice group made use of more lexical bundles than the upper-level group and that the novice writers do not share much of the bundles that the upper-level students used. The result supported the notion that patterns of lexical bundle use characterize learners’ novice and expert status in a specific discipline.

Several lexical bundle studies have put the focus on the native-non-native or expert-novice distinctions; however, there has been an attempt to examine lexical bundle use by learners of different proficiency levels. Appel (2011) complied a corpus of non-native speaker writers from the essays section of the Canadian Academic English Language (CAEL) test. Dividing the corpus into high- and low-proficiency levels according to their test scores, Appel made comparisons between the two groups’ use of 4- to 7-word lexical bundles. In general, the low proficiency group produced more types and tokens of lexical bundles than the high proficiency group and they also used more longer lexical bundles. Appel also examined the functions of the lexical bundles produced by the two groups, using the categories proposed by Biber et al. (2004): stance, discourse organizing and referential bundles. The low proficiency group used more types of the three categories than the high proficiency group, and some lexical bundles that were used by the low proficiency group, such as intention/prediction stance bundles, were not used by the high proficiency group at all. Appel also compared her analysis of lexical bundle functions with that of Biber et al.
(2004) and found that the high proficiency students were closer to the native speakers in terms of the proportion of lexical bundle use. Apple claimed that this constitutes evidence for “the potential path of progression” as learners become more proficient and more native-like in their writing (p. 94). However, it should be noted that the essays in the corpus were written on one topic with a specific essay prompt and participants were also exposed to two reading texts in the test. Appel did not eliminate lexical bundles that occurred in the essay prompt or the reading texts; rather, she highlighted this factor in the analysis and found that the high proficiency group used more “originally produced” (p. 105) lexical bundles than the low proficiency group. This showed that lower proficiency learners incorporated more phrases from the reading materials as a way to “to cope with their limited vocabulary and store of formulaic sequences” (p. 111). Given the differences in their use of lexical bundles, lower proficiency level learners could benefit from the kind of instructions that target the use of lexical bundles in academic discourse.

As can be seen from the discussion above, the majority of lexical bundle studies are comparisons between different groups of participants. Longitudinal case studies into the development and acquisition are rare. One study has been carried out by Li and Schmitt (2009) on the acquisition of “lexical phrases” by one Chinese participant over the period of one academic year. The lexical phrases in this study are defined differently from the traditional lexical bundle studies: native speaker intuitions are used to identify the lexical phrases and to judge the appropriateness of lexical phrase production; therefore, lexical phrases in this case can range from 2-word sequences, such as “according to”, to longer sequences like “a summary can be drawn”. Based on native speaker judgment of the lexical phrases and the interview with the subject, Li and Schmitt concluded that the subject, Amy,
demonstrated improvement in her use of lexical phrases and also increase in her confidence. From the interviews, it is found that Amy benefited equally from both implicit and implicit sources of learning with “a slight advantage for incidental learning” (p. 94), although Li and Schmitt noted that Amy benefited more from her writing class than the even more explicit feedback from the native speaker judges.

2.4 Lexical bundles and instruction

Given that lexical bundles are prevalent in natural discourse in different registers and academic disciplines and that there are still significant differences and gaps in lexical bundle use between language learners and native speakers as well as between novice and expert writers, it is worth exploring the effect of instruction on lexical bundle development.

Cortes (2006) designed a pedagogical intervention focusing on lexical bundles for students in the history discipline. Frequent lexical bundles in published articles were identified as target bundles and incorporated into five 20-minute lessons that were given once every two weeks to a class of 8 native speaker students. Students’ written assignments that were produced over the period of 10 weeks were analyzed for the use of the target bundles. The results of the analysis revealed that even though students are reported to be more aware of lexical bundles in published writing, they rarely use the target bundles in their writing and in many cases preferred to use shorter conjunctions in places where published authors would use lexical bundles.

The relationship between lexical bundles, pedagogical intervention and students’ learning has also been investigated with EFL learners. Ranjbar, Pazhakh and Gorjian (2012) designed an experiment that incorporated explicit instruction of lexical bundles in
the forms of role play, discussion, paragraph writing and conversation for a group of EFL undergraduate Iranian students. The control group was exposed to the “conventional method of grammar translation” (p. 248). The purpose of the study is to explore students’ fluency in paragraph writing after the pedagogical intervention. From the analysis of the pre- and post-test scores of both the control and experimental groups, Ranjbar et al. concluded that “striking progress” of the experimental groups of students could be attributed to the explicit instruction of lexical bundles. While there is statistical significance in pre- and post-test scores, the researchers do not provide any details on what lexical bundles are taught in the lessons and whether the students in the experimental group actually used those target bundles in their writings or not. Admittedly, the gains in writing proficiency could be attributed to the instruction of lexical bundles and to the fact that learners incorporated those bundles in their writings; however, without qualitative information on the lexical bundles that were taught and used by the students, it is still difficult to confirm that this was the sole source. The gains in writing proficiency can also be due to other factors.

The pedagogical intervention by Cao (2013) focuses on different task types and the learning of lexical bundles by Chinese EFL students. 70 Chinese EFL second-year undergraduate students were assigned into three task groups: reading comprehension, reading comprehension plus gap-filling, and sentence-writing. Pre-test, immediate post-test and delayed post-test, where learners were asked to give Chinese translations of the target bundles, were administered. Statistical analysis of the two post-test scores led Cao to conclude that the sentence-writing task “produced best initial learning and retention” (p. 960). There is also strong correlation between the “involvement loads” of activities and
students’ test scores (p. 961). While Cao’s study provided insights into the task types that might best benefit learners, the post-tests seems to only test learners’ recognition of lexical bundles meanings but not their ability to use those bundles in actual writing contexts.

From the studies with EFL learners, it is clear that pedagogical intervention might aid lexical bundle learning; however, more studies are needed to investigate learners’ actual production of lexical bundle in their writing, not just from their test scores.

2.5 The Academic Formula List

English language instructors in area of academic writing are most likely familiar with Coxhead’s Academic Word List or the AWL (2000) with 570 word families that are frequent and occur across various academic disciplines. The AWL is a valuable resource on which words to include in second language academic writing instructions.

Coxhead’s AWL has inspired many subsequent studies to comply lists of lexical items, using corpus data, that are specific to certain academic disciplines and registers, such as the EAP Science list (Coxhead & Hirsh, 2007) and the Academic Spoken Word List (Dang, Coxhead & Webb, 2017).

With general academic writing, there are also studies that aim to expand such list to for lexical items longer than one word. For example, Ackermann and Chen (2013) developed the Academic Collocation List (ACL), which includes 2,468 collocations that are frequent in academic writing.

Another example, which is relevant to this dissertation, is the Academic Formula List (AFL) by Simpson-Vlach and Ellis (2010). The list contains the total of 607 formulas, or lexical bundles, that are three to five words in length. These bundles are frequent and
found across different fields of studies, both in speech and in writing. The 607 lexical bundles on the list are divided into 207 core formulas (both spoken and written), 200 spoken formulas and 200 written formulas. They are also categorized based on their functions: referential expressions, stance expressions and discourse organizing functions. With this list, Simpson-Vlach and Ellis (2010) state that it is meant to be used as a “starting point” for subsequent research and for further teaching materials development (p. 502), rather than using the list ‘as is’ in teaching.

**2.6 Alignment in Second Language Acquisition**

This section discusses Atkinson’s sociocognitive approach to second language acquisition (SLA) and one of its three main principles that is relevant to this dissertation. Nishino and Atkinson (2014) outlines the three principles of the socioconitve approach, which are “the separability principle,” “the adaptivity principle” and “the alignment principle” (p. 38). The focus of this section and this dissertation is on the alignment principle.

According to the sociocognitive view, when it comes to second language acquisition, “mind, body, and world function interactively” (Atkinson, 2011, p. 143). That is, this approach maintains that human beings are both “ecological” and “adaptive” since their environment and how they constantly adapt to it are vital to their life and their survival, while considering learning, including second language learning, as “a default state of human affairs” (p. 143).

With the alignment principle, Atkinson, Churchill, Nishino, and Okada (2007) explain that “alignment takes place not just between human beings but also between human
beings and their social and physical environments” (p.171). In second language learning, alignment is when learners interact with “the language being learned” with various “sociocognitive affordances,” which include both human and nonhuman entities such as “literacy tools” (such as text books), “embodied tools” (such as “gestures”), “social tools,” and “individuals and their multiple roles and identities”, among other things (p.172).

In second language writing research, Nishino and Atkinson (2014) examined the interactions between two doctoral students who were working together to write a paper during the 13 video-recorded meeting sessions. The analysis of the interactions revealed that alignment can be observed when the two participants interacted with each other, using conversational fillers, intonations, syntactic structures and “gesture sharing” (p. 50), and with the objects around them, such as a book and the computer screen, as they worked together to write the paper. When summarizing the analysis of the study, Nishino and Atkinson also argued that individual writing session is also as interactive as two people co-writing a paper as in the study since the writer also interacts with their environment, such as writing medium, their location or other reading and writing materials, that is beyond what is going on inside their brain. Nishino and Atkinson also stressed that “sociocognitive alignment is at the heart of all writing” (p. 52).

In this dissertation, the concept of alignment is relevant when looking at how students interacted with the writing prompts given as part of the assignments and how they incorporated the prompts in their own writing. This will be discussed in detail in the chapters on lexical bundle use analysis.
2.7 The present study

English language instruction in Thailand has put the main focus on grammatical competence; however, it is evident that multi-word expressions such as lexical bundles are frequent and important features of English. These areas are still under-researched when it comes to Thai EFL learners’ use of this feature of the English language, especially in general academic writing.

While much lexical bundle research focuses on comparing novice (e.g., undergraduate or graduate student writers) and expert writers (e.g., published authors in the field), those studies treated all undergraduate students as one homogenous, novice group. Additionally, many previous studies focus on lexical bundle use in specific academic disciplines, such as biology, history and applied linguistics. While the focus on discipline-specific bundles is certainly useful and necessary, lexical bundle use in general academic writing needs to be examined as well, given its broader impact on English learners at the undergraduate level in general.

Combining all these needs, this dissertation investigates lexical bundle use among Thai undergraduate students in the faculty of Arts, who in many previous studies, would be all considered novice writers. However, in the current study, the novice students are students in the Freshman English course, which is the English I course with a writing component for first-year students in the first semester. The expert group in this study is the students in the Composition II course, which is a course offered to third- and fourth-year students. The students in both courses are from various majors, such as English, Thai, Spanish, history, philosophy, geography, information studies, and dramatic arts. As such, these courses are not discipline-specific.
In this study, in the first part of the analysis, I adopted frequency-based approach to the study of lexical bundles with the focus on bundles that are four words in length (Biber et al, 1999). In the second part of the analysis, I used the Academic Formula List (AFL) by Simpson-Vlach and Ellis (2010) as target bundles in order to investigate the use of the AFL bundles in the learner corpus. It is hoped that the results of these analyses will add to our understandings of Thai learner’s use of English regarding lexical bundles in writing. The specific research questions addressed in this study are:

1. What are the frequent lexical bundles that are used by the students in both subcorpora?
2. How are the two groups of students similar or different in their use of lexical bundles?
3. Which lexical bundles from the Academic Formula List (AFL) are used frequently by the students in both subcorpora?
4. How are the two groups of students similar or different in their use of lexical bundles from the Academic Formula List (AFL)?

The research design and corpus data collection will be discussed in detail in the following chapter.
CHAPTER 3
Research Methodology

This chapter covers the data collection process, the details of the corpus and the analytical steps to extract the lexical bundles for the analysis.

3.1 Research design
3.1.1 Data sources

The data were collected in the first semester of the academic year 2014-2015 at a large public university in Central Thailand. In the undergraduate curriculum of the Faculty of Arts, five writing courses are offered by the Department of English: English I, English II, English Composition I, English Composition II and Research Writing. Each course is a pre-requisite for the subsequent courses.

English I and English II are required courses for all first-year students in the Faculty of Arts, regardless of their majors. English I is offered in the first semester and English II the second semester of the same academic year. These two courses are general English courses with writing components with the focus on paragraph writing. In English I, students learn how to write three genres of paragraphs: example, reason and description. In English II, the three genres of focus are comparison-contrast, persuasion and narration.

English Composition I, offered in the second academic semester, is a required course for second-year English major students and an elective for students who choose English as their minor. This course is offered in the second semester. The objective of the
course is for students to develop their writing skills from paragraph writings in their first year to argumentative essay writing on different topics.

English Composition II, offered in the first academic semester, is an elective for both English majors and minors. Although students can choose to register for this course when they are in their third year, most students prefer to take this course in their fourth year. In English Composition II, the focus is on writing argumentative essays on different topics. In addition to essay writing that they have learned in English Composition I, students also learn to write an analysis of an essay and a literary analysis.

The most advanced writing course, Research Writing, is an elective course for fourth-year students and is offered in the second semester. In this course, the instruction centers on different stages of writing a research paper, from selecting the research topic to using references in the research. Students are allowed to choose any research topic of their own interest in consultation with the instructor.

These five courses aim at preparing students for higher levels of writing, where students progress from being a beginner, or a novice, in writing in English I to developing more expertise in writing. The writing tasks themselves, in the same way, progress from short, simple paragraph writing to longer argumentative essays to full-fledged research papers. In this sense, the students in the course English I can be considered novice whereas the students in more advanced writing courses, such as English Composition II, would be regarded as relatively expert in the context of the undergraduate curriculum.

The data for this research were collected from the participants who are undergraduate students in the Faculty of Arts, enrolled in either English I or English Composition II. These two courses are the only writing courses offered in the first semester.
The other available courses with an emphasis on writing are English II, English Composition I and Research Writing, offered in the second semester. Altogether, there were 331 participants in this research—295 from English I and 36 from English Composition II.

The two courses and the students who enrolled in them are different in nature. As explained earlier, English I is a compulsory course for all first-year students in the Faculty of Arts. This course consists of 14 sections of about 20-24 students each (around 300 students in total). The writing component of English I aims at training students to write paragraphs of three different genres: example, reason and description. Students are generally assigned to write at least two practice paragraphs for each genre outside of the class hours. They are then assessed in class where they are required to write a paragraph on a given topic within the class hour. At the end of the semester, students are also required to write one paragraph in the final examination (example, reason or description). In total, each student could produce up to 10 different paragraphs in the semester.

The seven main types of paragraphs produced by the students in English I, along with their writing conditions and time constrains, are summarized below:

1. Example paragraph practice (take-home)
2. Example paragraph test (in-class, one hour)
3. Reason paragraph practice (take-home)
4. Reason paragraph test (in-class, one hour)
5. Descriptive paragraph practice (take-home)
6. Descriptive paragraph test (in-class, one hour)
7. Final exam writing test (one hour)
The prompts for paragraph practices were given to the students by their section instructors, with some variations among the sections. The writing prompts of the tests and the final examination were the same for all students. With the exception of the final examination where the writing prompts were part of the English I final examination paper, the three in-class assignment prompts were presented to the students in different manners, depending on the instructors, which include writing the prompts on the whiteboard, making PowerPoint slides for the prompts, giving students paper slips with the prompts written on, and reading the prompts to the students. These different ways in presenting the prompts resulted in minor differences in the actual prompts provided as well. For example, for the reason paragraph test, the students in 13 of the 14 sections were presented with the prompt “Why more Thai people are riding bicycles these days” but those in one section received the prompt “Why more Thai people are riding their bicycles these days” with the possessive pronoun “their” inserted in the prompt. While this did not affect the students’ understanding of the prompt, the slight difference in the prompts has some effect on the lexical bundles that are produced by the students, which will be discussed in subsequent chapters.

The number of writing prompts presented to the students in the 2014-2015 academic year are summarized in Table 3.1 below:
Table 3.1: Writing prompts for English I writing tasks

<table>
<thead>
<tr>
<th>Writing</th>
<th>Number of Writing Prompts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example paragraph practice</td>
<td>38</td>
</tr>
<tr>
<td>Example paragraph test</td>
<td>6</td>
</tr>
<tr>
<td>Reason paragraph practice</td>
<td>27</td>
</tr>
<tr>
<td>Reason paragraph test</td>
<td>7</td>
</tr>
<tr>
<td>Descriptive paragraph practice</td>
<td>32</td>
</tr>
<tr>
<td>Descriptive paragraph test</td>
<td>14</td>
</tr>
<tr>
<td>Final exam writing test</td>
<td>3</td>
</tr>
</tbody>
</table>

English Composition II is a compulsory course for third-year and fourth-year students who have passed the three pre-requisite courses: English I, English II, and Introduction to Composition. This course generally consists of 4 sections of up to 15 students in each (around 60 students in total). In this class, students engaged in three different genres of writing: argumentative essays, analysis of essays, and literary analysis. As a part of their assessments, they were required to write three argumentative essays on different topics, based on the reading materials covered in class. Two essays were written in response to essay prompts and the students were required to submit two drafts (the first written in class and the second outside of class after they have received feedbacks on the first draft). The essay prompts for these two essays were the same for all students in the four sections. For the third essay, the topics were of the students’ own choosing, with the approval of the instructor. The second writing assessment component was analyses of essays. The students were required to write two analyses of two essays that were part of the reading material in class. The last writing assessment was literary analysis in the final examination on a short story given to the students one week prior to the final examination.
In total, each student produced eight different writings in this course: two drafts, three complete essays, two analyses, and one literary analysis.

The eight main types of written production by the students in the English Composition II course, along with their writing conditions and time constrains, are as follows:

1. Draft 1 of Essay 1 (in-class, one and a half hour)
2. Draft 2 of Essay 1 (take-home)
3. Draft 1 of Essay 2 (in-class, two sessions of one and a half hour each)
4. Draft 2 of Essay 2 (take-home)
5. Research paper (take-home)
6. Analysis of an essay #1 (in-class, one and a half hour)
7. Analysis of an essay #2 (in-class, one and a half hour)
8. Final Exam: literary analysis essay (three hour)

For the 2014-2015 academic year, the writing prompts for the essays were in the format of one to two short paragraphs, highlighting the issues for the students to write about. The prompt for the essay analyses was the complete essays to be analyzed. The two writing prompts for the literary analysis each consisted of only one sentence. Students were given paper slips that contained the writing prompts for all the written assignments.

The number of writing prompts presented to the students are summarized in Table 3.2 below:
Table 3.2: Writing prompts for Composition writing tasks

<table>
<thead>
<tr>
<th>Writing</th>
<th>Number of Writing Prompts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essay 1</td>
<td>3</td>
</tr>
<tr>
<td>Essay 2</td>
<td>4</td>
</tr>
<tr>
<td>Analysis 1</td>
<td>2 (essays)</td>
</tr>
<tr>
<td>Analysis 1</td>
<td>2 (essays)</td>
</tr>
<tr>
<td>Literary Analysis</td>
<td>2</td>
</tr>
</tbody>
</table>

The assignments that were submitted electronically via email were converted to plain text files. Assignments that were handwritten and submitted in hard copy were digitalized and saved as plain text files. The file names were coded to include information of the course, the assignment (e.g. example paragraph test or literary analysis essay) and the students’ assigned identification number.

3.1.2 The corpus

The writing samples collected in this research were compiled into the Corpus of Thai Undergraduate Learners of English 2014 (TULE2014 Corpus). The TULE2014 Corpus consists of two subcorpora: the Freshman English subcorpus, which contains the samples collected from English I, and the Composition subcorpus, which contains the samples collected from English Composition II. The details of the TULE2014 Corpus are summarized below:
Table 3.3: Total word count of the TULE2014 Corpus

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Word Count</th>
<th>Number of Files</th>
<th>Number of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>617,572</td>
<td>2,047</td>
<td>331</td>
</tr>
<tr>
<td>Freshman English</td>
<td>358,091</td>
<td>1,805</td>
<td>295</td>
</tr>
<tr>
<td>Composition</td>
<td>259,481</td>
<td>242</td>
<td>36</td>
</tr>
</tbody>
</table>

Table 3.4: Freshman English subcorpus word count

<table>
<thead>
<tr>
<th>Category</th>
<th>Word Count</th>
<th>Number of Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example Practice</td>
<td>55,692</td>
<td>306</td>
</tr>
<tr>
<td>Example Test</td>
<td>60,410</td>
<td>293</td>
</tr>
<tr>
<td>Reason Practice</td>
<td>34,689</td>
<td>176</td>
</tr>
<tr>
<td>Reason Test</td>
<td>61,615</td>
<td>294</td>
</tr>
<tr>
<td>Description Practice</td>
<td>30,996</td>
<td>156</td>
</tr>
<tr>
<td>A. Person</td>
<td>15,523</td>
<td>84</td>
</tr>
<tr>
<td>B. Place</td>
<td>15,473</td>
<td>72</td>
</tr>
<tr>
<td>Description Test</td>
<td>62,567</td>
<td>289</td>
</tr>
<tr>
<td>A. Person</td>
<td>24,117</td>
<td>115</td>
</tr>
<tr>
<td>B. Place</td>
<td>38,450</td>
<td>174</td>
</tr>
<tr>
<td>Final Exam</td>
<td>52,122</td>
<td>291</td>
</tr>
<tr>
<td>Example</td>
<td>21,954</td>
<td>125</td>
</tr>
<tr>
<td>Reason</td>
<td>14,521</td>
<td>90</td>
</tr>
<tr>
<td>Description</td>
<td>15,647</td>
<td>76</td>
</tr>
</tbody>
</table>

Table 3.5: Composition subcorpus word count

<table>
<thead>
<tr>
<th>Category</th>
<th>Word Count</th>
<th>Number of Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis 1</td>
<td>20,124</td>
<td>30</td>
</tr>
<tr>
<td>Analysis 2</td>
<td>25,992</td>
<td>28</td>
</tr>
<tr>
<td>Essay 1 Draft 1</td>
<td>16,226</td>
<td>36</td>
</tr>
<tr>
<td>Essay 1 Draft 2</td>
<td>55,170</td>
<td>35</td>
</tr>
<tr>
<td>Essay 2 Draft 1</td>
<td>20,118</td>
<td>28</td>
</tr>
<tr>
<td>Essay 2 Draft 2</td>
<td>37,472</td>
<td>24</td>
</tr>
<tr>
<td>Essay 3</td>
<td>42,772</td>
<td>26</td>
</tr>
<tr>
<td>Final Exam</td>
<td>41,607</td>
<td>35</td>
</tr>
</tbody>
</table>
3.1.3 Data analysis

In this study, the analysis is in two parts. The first part is the analysis of frequent lexical bundles that were produced by all participants in both subcorpora. The second part is the analysis of lexical bundles from the academic formula list (Simpson-Vlach & Ellis, 2010) that are used by the students in both subcorpora. For both analyses, I used the concordancing program AntConc version 3.4.1m for Macintosh OS X 2014 (Anthony, 2014) for lexical bundle extraction.

The first part of lexical bundle analysis focused on the types, tokens, structures and functions. Lexical bundles with the length of four words were extracted and refined using the following criteria: the lexical bundle frequency cut-off point at 40 occurrences per million words and production by at least three individual students.

For the first part of the analysis, the frequency cut-off point was 14 times for the Freshman English subcorpus and 10 times for the Composition subcorpus.

The lexical bundles extracted were then further categorized for their structures and functions, using the categorization proposed by Biber, Conrad and Cortes (2004).

Structurally, lexical bundles can be categorized into the following broad categories:

1. Lexical bundles that incorporate verb phrase fragments
2. Lexical bundles that incorporate dependent clause fragments
3. Lexical bundles that incorporate noun phrase and prepositional phrase fragments

In the present analysis two additional subcategories (“there is/are + verb phrase” and “noun/noun phrase + verb phrase”) are added to the first structural category of verb phrase fragments to include all the lexical bundle structures found in the learner corpus.
Functionally, lexical bundles in writing can be grouped into three main functions with the following sub-categories:

1. Stance expression
2. Discourse organizer
3. Referential expression

In the original taxonomy by Biber et al. (2004), there is a fourth category specific to conversational function with the three subcategories: politeness, simple inquiry and reporting. Since the corpus data in this analysis is in the written register, the conversational function was not taken into consideration.

As the data consisted of writing samples in different genres produced in response to different writing prompts, three additional categories were used in the analysis: prompt-specific bundles, genre-specific bundles and originally-produced versus non-originally-produced bundles.

Prompt-specific lexical bundles are bundles found in essays for a specific prompt only. In the same way, genre-specific lexical bundles are bundles found in essays of a specific genre. Originally-produced and non-originally-produced lexical bundles are identified by whether the lexical bundles are repetition of the words or phrases in the essay prompts or not. In the case of essay analyses for Composition, if the lexical bundles were from the essay being analyzed, from the short story, or from the class materials that were given as examples, they were also labeled as non-originally-produced lexical bundles.

The second part of the analysis in this dissertation uses the lexical bundles from the Academic Formula List or the AFL (Simpson-Vlach & Ellis, 2010) as target bundles. The target bundles from the academic formula list were searched in both learner corpora, using
the concordancing program AntConc, and the occurrences of the bundles were noted down for both types and tokens. All the AFL bundles that are found in the learner corpus are filtered further to include only frequent bundles that occur at 40 times per million words or over, which is 14 times for the Freshman English subcorpus and 10 times for the Composition subcorpus. I also use the register and functional categories from the academic formula list (Simpson-Vlach & Ellis, 2010) to group the target AFL bundles that are found in the Freshman English and Composition subcorpora.

By register, the Simpson-Vlach and Ellis’ AFL bundles are grouped into core AFL (spoken and written), primarily spoken and primarily written. By functions, there are 3 main functions: referential expressions, stance expressions and discourse organizing functions. Within each main functional category, the AFL bundles are divided further into subcategories and grouped according to their register status.

In addition to the register and functional categories, the analysis also covered AFL target bundles that are used by students in both learner subcorpora (shared bundles) and those that are used by only students in one subcorpus (unshared).
3.2 Research questions

The research questions addressed in this dissertation are repeated below:

1. What are the frequent lexical bundles that are used by the students in both subcorpora?

2. How are the two groups of students similar or different in their use of lexical bundles?

3. Which lexical bundles from the Academic Formula List (AFL) are used frequently by the students in both subcorpora?

4. How are the two groups of students similar or different in their use of lexical bundles from the Academic Formula List (AFL)?

The first two questions will be addressed in the first analysis in Chapter 4 and the last two questions will be addressed in the second analysis in Chapter 5.
CHAPTER 4

Lexical Bundles in the Learner Corpus

This chapter will present the results of the analysis of the Freshman English and the Composition subcorpora. The results and discussion will cover the following areas: the number of lexical bundle types and tokens, structures, functions, prompt-specific bundles vs. general bundles, genre-specific bundles vs. general bundles, originally-produced bundles vs. non-originally produced bundles, and shared lexical bundles between the two subcorpora.

4.1 Overall results

The results of the types and tokens of lexical bundles of all lengths extracted from both subcorpora are summarized below in Table 4.1:

Table 4.1: Overall comparison of lexical bundles for Freshman English and Composition

<table>
<thead>
<tr>
<th>Subcorpus</th>
<th>Types</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman English</td>
<td>106</td>
<td>3,279</td>
</tr>
<tr>
<td>Composition</td>
<td>94</td>
<td>1,702</td>
</tr>
</tbody>
</table>

The overall comparison in the table above shows that the participants in the Freshman English subcorpus used slightly more lexical bundle types than the Composition subcorpus, at 106 and 94 types respectively. However, when looking at the number of lexical bundle tokens, we can see that the number in the Freshman English subcorpus (3,279 tokens) almost doubles that in the Composition subcorpus (1,702 tokens). In short,
the Freshman English subcorpus contains both more lexical bundle types and tokens than
the Composition subcorpus.

If we narrow down the results further to look at only high-frequency lexical
bundles, i.e., those that occur more than 100 times per million words or, in the Freshman
English subcorpus, 36 times and over, we can see that there are 22 lexical bundles that fall
into this category. While those bundles account for only about 20% of the total bundle
types, they account for almost half of the total lexical bundle tokens in this subcorpus (with
1,552 tokens).

In Composition, bundles that occur 100 times or more per million words are those
with the raw frequency of 26 and more. There are 15 bundles in this category, accounting
for 16% of the total lexical bundle types but 36% of the overall lexical bundle tokens in
the Composition subcorpus (with 618 tokens).

In the Freshman English subcorpus, one reason that might explain the high number
of lexical bundle tokens is the high frequency of the top three lexical bundles that are part
of the writing prompts. These three bundles are *a top fashion model* (304 tokens), *I used to
collect* (143 tokens), and *Thai people are riding* (124 tokens). These three most frequent
bundles alone make up 17.4% of the total lexical bundle tokens (571 out of 3,279 tokens),
with the most frequent bundle *a top fashion model* at 9.3% of the total tokens in the
subcorpus.

These three bundles are part of the writing prompts that were given to the students:
- *a top fashion model* is from the example prompt in the final exam
  “advantages/disadvantages of being a top fashion model”
- *I used to collect* is from the example test prompt “Things you (or someone you know) used to collect as a child” (only one occurrence out of 143 was found in an essay produced for a different writing prompt)

- *Thai people are riding* is from the reason test prompt “Why more Thai people are riding (their) bicycles these days”

The most frequent bundle in the Composition subcorpus is also part of the essay writing prompt: *the online homework services*. However, compared to the Freshman English subcorpus, whose most frequent bundle accounts for 9.3% of the total bundle tokens, the frequency of this top bundle is relatively low, with only 77 tokens or 4.5% of the total lexical bundle tokens.

The difference in token frequency when it comes to lexical bundles that are part of the writing prompts in both subcorpora can be explained by the fact that, in Freshman English, those three bundles are key words in the topic of the writing. Therefore, it is not unusual to see them used in the students’ writings and it is almost impossible to avoid using phrases that contain the keywords from the prompt. In most cases, the lexical bundles are in the topic sentences where students would incorporate most of the writing prompts as part of the first topic sentence. However, there are several cases where the bundles are used repeatedly throughout the paragraph. This is most clear with *a top fashion model* when many students used this bundle instead of using other nouns or noun phrases. In three cases, the bundle is used seven times in a short paragraph.

In Composition, the most frequent bundle in the subcorpora *the online homework services* is used only 77 times and is not as frequent as the Freshman English bundle *a top fashion model*. The bundle is also part of the essay writing prompt but students also used
other nouns or noun phrases, such as “the services(s)”, “this kind of service(s)”, “online service(s),” or even another bundle identified in the result analysis *the online homework service*, instead of *the online homework services* from the writing prompt.

To explore the lexical bundle use by the two groups further, the subsequent sections will discuss the structures and functions of the bundles found in the subcorpora.

### 4.2 Lexical bundle structures

This section discusses lexical bundle structures, using the categories proposed by Biber et al (2004) with additional subcategories to include all the structures identified in the subcorpus. The table 4.2 below summarizes the number of lexical bundles in each functional category:

<table>
<thead>
<tr>
<th>Subcorpus</th>
<th>Lexical bundles that incorporate verb phrase fragments</th>
<th>Lexical bundles that incorporate dependent clause fragments</th>
<th>Lexical bundles that incorporate noun phrase and prepositional phrase fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman English</td>
<td>55 (51.89%)</td>
<td>11 (10.38%)</td>
<td>40 (37.73%)</td>
</tr>
<tr>
<td>Composition</td>
<td>38 (40.43%)</td>
<td>13 (13.83%)</td>
<td>43 (45.74%)</td>
</tr>
</tbody>
</table>

As the table shows, in both subcorpora, the lexical bundles are phrasal in nature (verb phrases, noun phrases and prepositional phrases), with only a small percentage of clausal lexical bundles (10.38% for Freshman English and 13.83% for Composition). In the Freshman English subcorpus, the largest subcategory is lexical bundles that incorporate
verb phrase fragments (51.89%) while in Composition, the largest is those with noun and prepositional phrase fragments. Each subcategory will be discussed separately in the subsequent sections.

4.2.1 Lexical bundles that incorporate verb phrase fragments

In this category, two additional sub-categories of verb phrase fragments are added to include all the bundle structures found in the two subcopora. These structures are “there is/are fragments” and “noun/noun phrase + VP fragments”

In the Freshman English subcorpus, there are 55 lexical bundles with this structure (51.89% of the total bundle types) and in the Composition subcorpus, there are 38 such bundles (40.43%).

Looking further into the subcategories of verb phrase fragment structures, we can divide the 55 bundles in the Freshman English subcorpus into the following 5 verb phrase structures:

- (Connector +) 1st/2nd person pronoun + VP fragment (12 bundles): I used to collect, you don’t have to, I would like to, you have to be, as you can see, I went to the, you have to do, I think it is, you don’t want to, you will have to, you can go to, we don’t have to

- (Connector +) 3rd person pronoun + VP fragment (2 bundles): they don’t have to, it is a good

- Verb phrase (with non-passive verb) (19 bundles): have just woken up, difficult for many students, be a required course, celebrate a friend’s birthday, is a bad fear, is one of the, have a lot of, has a lot of, want to be a, will be able to, have a chance
to, should study Thai literature, have to go to, have to pay for, is a good thing, is a
good choice, not be able to, be one of the, have enough time to

- There is/are fragments (13 bundles): there are a lot, there are many reasons, there
  are many interesting, there are so many, there are three reasons, there are many
disadvantages, there are several reasons, there are lots of, there are many things,
there are several ways, there are many advantages, these are the reasons, there are
many ways

- Noun/Noun Phrase + VP fragment (9 types): Thai people are riding, the last one
  is, riding a bicycle is, studying Thai literature is, some of them are, the last reason
  is, riding bicycles is a, riding a bicycle can, most of them are

In Composition, the 38 verb phrase bundles are divided into six structures:

- (Connector +) 1st/2nd person pronoun + VP fragment (2 bundles): we can see that,
as we can see

- (Connector +) 3rd person pronoun + VP fragment (7 bundles): it is true that, it is
  clear that, so that they can, it is possible that, it is a good, it is important to, they
do not have

- Verb phrase (with non-passive verb) (15 bundles): is one of the, have enough time
to, is going to be, do homework by themselves, does not want to, not be able to, use
online homework services, is not going to, will be able to, shows that Mrs Spence,
is not the only, have too much homework, do not want to, do not have to, is the one
who

- Verb phrase (with passive verb): is supposed to be, can be seen as, written by
Nathaniel Bellows
- There is/are fragments (1 bundle): *there are a lot*

- Noun/Noun Phrase + VP fragment (10 bundles): *some might say that, Mrs Spence is a, Mrs Spence is the, some may say that, Monthita Rojtinnakorn argues that, some people might say, the author argues that, Mrs Spence is not, his parents are away, some people claim that”*

The two subcorpora are similar in that the structures with the most bundles are verb phrase (with non-passive verb) and noun/noun phrase fragments. However, there are several differences between the two subcorpora. The first difference is that there is no verb phrase with passive verb in Freshman English but there are three bundles of that structure in Composition. The second difference is the number of the bundles with pronouns. In Freshman English, 12 bundles are verb phrases with 1st or 2nd person pronouns while only 2 are bundles with 3rd person pronouns. In contrast, in Composition there are more bundles with 3rd person pronouns (7 bundles) than those with 2nd person pronouns “we” (2 bundles). Another difference is the bundles with “there is/are”: in Freshman English, these bundles are abundant (13 bundles) and in most cases are used as part of a topic sentence or concluding sentences; in Composition, however, there is only one “there are” lexical bundle.

**4.2.2 Lexical bundles that incorporate dependent clause fragments**

This second structural category is the smallest structural category for both subcorpora, with only 11 bundles in Freshman English (10.38%) and 13 bundles in Composition (13.83%).
In Freshman English, the 11 lexical bundles that incorporate dependent clause fragments can be further divided into three sub-categories:

- **WH-clause fragment (5 bundles):** *when I was a, when I was young, why it is important, when I was in, why reading tests are*

- **if-clause fragment (2 bundles):** *if you want to, if you have a*

- **(verb/adjective+) to-clause fragment (4 bundles):** *to study Thai literature, to ride a bicycle, to improve my English, to help the environment*

As for the composition subcorpus, the 13 lexical bundle types are also divided further into three categories:

- **WH-clause fragment (2 bundles):** *what they have learned, when it comes to*

- **(verb/adjective+) to-clause fragment (7 bundles):** *to take care of, to do their homework, to solve the problem, to be able to, to play the piano, to pick him up, to make sure that*

- **That-clause fragment (4 bundles):** *that he does not, that online homework services, that it is not, that Thai language is*

Both subcopora contain bundles that are of the WH-clause and to-clause fragments.

In Freshman English, the WH-clause are “when” and “why” bundles whereas in composition the two bundles are “what” and “when” bundles. The “why” bundles in Freshman English are from reason paragraph writings.

The difference between the subcopora is, in Freshman English, there are two bundles that are if-clause fragments but none in Composition. In Composition, however, there are 4 that-clause bundles but this bundle structure is not in the Freshman English subcoprus.
4.2.3 Lexical bundles that incorporate noun phrase and prepositional phrase fragments

The last structural category is the second largest in Freshman English, with 40 bundles or 37.73% of the bundle tokens in the corpus, and the largest category in Composition, with 43 bundles or 45.74% of the bundle tokens.

In Freshman English, the bundles fall into four structural sub-categories:

- (Connector +) noun phrase with of-phrase fragment (15 bundles): *one of the most*, *a lot of money*, *the right side of*, *the left side of*, *behaviors of music fans*, *the center of the*, *scene of a murder*, *the faculty of arts*, *the back of the*, *the end of the*, *the top of the*, *a lot of things*, *corner of the room*, *a lot of people*, *one of my friends*

- Noun phrase with other post-modifier fragment (4 bundles): *a good way to*, *the best way to*, *people in the past*, *people around the world*

- Other noun-phrase expressions (8 bundles): *a top fashion model*, *keeping up with fashion*, *my high school classroom*, *my English I teacher*, *staying in a dormitory*, *the new year’s holiday*, *last but not least*, *being a top model*

- Prepositional phrase expression (13 bundles): *for all arts students*, *during the final exam*, *during final exam week*, *in front of the*, *at the same time*, *in the middle of*, *for a long time*, *on the other hand*, *in faculty of arts*, *in the same time*, *with a lot of*, *for all these reasons*, *in front of me*

In Composition, the 43 noun and prepositional phrase bundles are further categorized into five subcategories:

- (Connector +) noun phrase with of-phrase fragment (15 bundles): *the Ministry of Education*, *the end of the*, *the amount of homework*, *one of the most*, *this kind of
service, the beginning of the, a number of weaknesses, both sides of the, the majority of the, a considerable number of, a lot of homework, a lot of people, the use of the, one of the main, a waste of time

- Noun phrase with other post-modifier fragment (7 bundles): destructive force in the, the fact that the, lesson with Mr Nichols, footpaths in Siam Square, these online homework services, someone else to do, an important role in

- Other noun-phrase expressions (6 bundles): the online homework services, the online homework service, using online homework services, public safety and security, last but not least, doing homework by themselves

- Prepositional phrase expression (13 bundles): on the other hand, at the same time, in the food industry, of online homework services, by pointing out that, into the world of, in order to be, despite the fact that, in the boy’s life, in the first place, in order to get, between the narrator and, in the world of

- Comparative expression (2 bundles): as well as the, as long as the

Comparing the two subcorpora, we can see that Composition has one structural subcategory that Freshman English does not have: comparative expressions.

In Freshman English, in the prepositional phrase expression structure, there is one instance of grammatically incorrect bundle. The bundle in the same time occurs 16 time in this subcorpus and in these occurrences it is clear that the students intended to use at the same time, a bundle which also occurs at a higher frequency (46 times) than the version with the preposition “in”:

- …. students ride a bicycle to school they are doing exercise in the same time…. 
- …being a fashion model has a disadvantage and advantage in the same time in term of working hard.

- … not only pay good wages but also help improve your language skill in the same time.

Incorrect choice of preposition in this bundle can be attributed to the influence of the first language of the students, Thai. In Thai, the phrase whose meaning is the same as at the same time is used with the Thai preposition that translate to “in” in English. As a result, a confusion in preposition use is possible in several cases. However, a higher occurrence of the correct preposition use in at the same time shows that there are more students who can use the expression correctly.

4.3 Lexical bundle functions

The lexical bundles in the corpus can be grouped into three main functions: stance expression, discourse organizer and referential expression. The results from the learner corpus are summarized in Table 4.3 below. In each cell, the number in the first line indicates the number of lexical bundle types in a category and the parenthesized number indicates the percentage of bundle types that are in that category.
Table 4.3: Lexical bundle functions

<table>
<thead>
<tr>
<th>Subcorpus</th>
<th>Stance expression</th>
<th>Discourse organizer</th>
<th>Referential expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman English</td>
<td>23 (21.7%)</td>
<td>18 (17%)</td>
<td>65 (61.3%)</td>
</tr>
<tr>
<td>Composition</td>
<td>19 (20.21%)</td>
<td>20 (21.28%)</td>
<td>55 (58.51%)</td>
</tr>
</tbody>
</table>

In terms of function, both subcorpora are similar in their large proportions of lexical bundles that serve the referential expression (61.3% for Freshman English and 58.51% for Composition). Freshman English has slightly more stance expressions than discourse organizers while in Composition there is only one more bundle in the discourse organizer category than in stance expressions. Each functional category will be discussed separately in the subsequent sections.

4.3.1 Stance expression

In Freshman English, in the first category of stance expressions, only one of the 23 lexical bundle types is in the epistemic stance subcategory (*I think it is*) and the other 22 bundles are attitudinal/modality stance bundles, which are *if you want to, want to be a, I would like to, you don’t want to, to improve my English, to study Thai literature, you don’t have to, be a required course, they don’t have to, you have to be, should study Thai literature, you have to do, have to go to, have to pay for, you will have to, we don’t have to, will be able to, you can go to, not be able to, to help the environment, have a chance to, and riding a bicycle can.*
Similarly, in the Composition subcorpus, of the 19 stance expressions, one is an epistemic stance bundle (it is clear that) and the rest are attitudinal/modality stance (18 bundles): to take care of, that he does not, is going to be, to be able to, do homework by themselves, does not want to, is supposed to be, in order to be, so that they can, not be able to, is not going to, it is possible that, will be able to, do not want to, it is important to, do not have to, to make sure that and in order to get.

With attitude/modality stance bundles, some bundles express different stances, which are desire, obligation/directive, intention/prediction and ability, depending on the context of use. However, in general, the majority of the stance bundles in Freshman English are obligation/directive bundles, such as you don’t have to, you have to be, should study Thai literature and have to go to. In composition, most of the stance bundles in this category are intention/prediction bundles (such as is going to be, in order to be, is not going to and it is possible that) and ability bundles (such as to be able to, do homework by themselves, so that they can and not be able to). In short, even though both subcorpora are similar in the number of attitudinal/modality stances, the actual functions of the bundles they use are different, with Freshman English using more obligations/directive bundles and Composition using more intention/prediction and ability stance bundles.

4.3.2 Discourse organizer

For discourse organizers, in Freshman English, the 18 bundles of this function are further divided into two categories:

- Topic introduction/focus (14 bundles): there are many reasons, there are many interesting, why it is important, the last one is, there are three reasons, there are
many disadvantages, there are several reasons, there are many things, there are several ways, the last reason is, why reading tests are, there are many advantages, last but not least, there are many ways

- Topic elaboration/clarification (4 bundles): as you can see, on the other hand, these are the reasons, for all these reasons

In Composition, the 20 bundles are divided into the two subcategories below:

- Topic introduction/focus (8 bundles): it is true that, some might say that, may say that, Monthita Rojitinnakorn argues that, the author argues that, some people might say, some people claim that, last but not least

- Topic elaboration/clarification (12 bundles): on the other hand, we can see that, the fact that the, when it comes to, as we can see, by pointing out that, despite the fact that, shows that Mrs Spence, as long as the, as well as the, written by Nathaniel Bellows, can be seen as

The difference between discourse organizing bundles in the two subcorpora lies in the functional subcategories. The majority of the stance bundles in Freshman English are topic introduction/focus bundles. However, in Composition, there are more topic elaboration/clarification lexical bundles.

In Freshman English, most of the topic introduction/focus bundles (9 out of 14) are of the “there are” structures and these bundles are primarily used in the topic sentences in the paragraph, followed by the topic or the prompt of the writing, such as:

- There are many reasons why students should not have to wear uniforms.

- Prompt: University students should be required to wear uniform

- There are three reasons why it is important to study Thai literature.
- Prompt: Why it is/isn’t important to study Thai literature

- *There are many interesting* ways to celebrate a friend’s birthday.

- Prompt: Interesting ways to celebrate a friend’s birthday

The use of “there are” in topic sentences is never explained explicitly in the writing packet of the class materials. However, the majority of the students used this structure in the topic sentences of their paragraphs. One reason is that the “there are” structure is presented to the student in the section about brainstorming supporting examples. These two “there are” sentences are:

- There are many ways to help improve your English.
- There are many interesting things for a tourist to see in my city.

Another instance of “there are” in the topic sentence is in a model example paragraph at the end of the example section in the writing packet. The topic sentence from that model paragraph is “At [University name], there are many useful majors that students can choose from.” In the reason paragraph writing, none of the model paragraph uses this “there are” structure as part of the topic sentence. In other model paragraphs in the writing packet for example and writing, the topic sentences are quite specific to the prompts and, unlike the previous example, cannot be readily “borrowed” to be used in other writing:

- My best friend Susan is a very generous person (example)
- My next-door neighbor Mike is a workaholic (example)
- My brother Tony is a very organized person (example)
- Krabi is a great destination for a four-day holiday (reason)
- Smoking is a very bad habit (reason)
In Composition, most of the topic introduction/focus bundles are those that introduce a counterargument in the essay, both in the introduction paragraph and in supporting paragraphs. The use of counterargument bundles reflects how students are trained to write in this course and from the previous courses, English II and Composition I. In this subcorpora, those bundles are *it is true that, some might say that, some may say that, some people might say* and *some people claim that*. They are used in a pattern with the transitional expressions “while,” “although,” “nevertheless” or “however”:

- **While it is true that** the court’s rulings promotes companies awareness is to prevent the harassment, criticizing men’s appearance is still inconsiderate…

- **Although some people might say** that psychological service or therapy cost a lot of money, there are some other options to reach the services cheaper….

- **Some people claim that** Mrs. Spence is an essentially destructive force in the narrator’s life. **However**, others believe that she is a really good house sitter

- **Some may say that** in some subject, like art, some students cannot do well or think that it is not important. **Nevertheless**, most people may not do anything well at the first time…

The two bundles *Monthita Rojinnakorn argues that* and *the author argues that* are mainly used in the first paragraph of essay analysis writing in order to introduce the main idea or thesis statement of the essays that were being analyzed and written about.

With the second discourse organizer subcategory, which is topic elaboration/clarification, in the Freshman English subcorpus, most of the bundles, with the exception of *on the other hand*, are those that lead into the concluding sentences: *as you can see, these are the reasons* and *for all these reasons*. The topic elaboration/clarification
bundles in Composition are primarily those that are used to introduce supporting points such as *we can see that, the fact that the, as we can see, by pointing out that* and shows that *Mrs Spence.*

### 4.3.3 Referential expression

The last functional category, referential expression, is also where the majority of the lexical bundles are. 65 of the lexical bundles in Freshman English (61.3%) and 55 bundles in Compositional (58.51%) serve this function.

In Freshman English, these 65 bundles are categorized further into three groups:

- **Identification/focus (25 bundles):** *a top fashion model, I used to collect, Thai people are riding, is one of the, one of the most, for all arts students, difficult for many students, celebrate a friend’s birthday, behaviors of music fans, keeping up with fashion, scene of a murder, is a bad fear, riding a bicycle is, studying Thai literature is, to ride a bicycle, I went to the, riding bicycles is a, my English I teacher, if you have a, people in the past, staying in a dormitory, being a top model, be one of the, one of my friends, people around the world*

- **Specification of attributes (17 bundles):** *there are a lot, a lot of money, have a lot of, has a lot of, a good way to, the best way to, there are so many, there are lots of, some of them are, a lot of things, is a good thing, a lot of people, it is a good, is a good choice, with a lot of, have enough time to, most of them are*

- **Time/place/text reference (23 bundles):** *in front of the, when I was a, when I was young, at the same time, have just woken up, during the final exam, the right side of, in the middle of, for a long time, the left side of, the center of the, when I was in,
the faculty of Arts, the back of the, the end of the, the top of the, my high school classroom, faculty of Arts, in the same time, corner of the room, during final exam week, the new year’s holiday, in front of me

In composition, the 55 stance bundles are also divided into three subcategories as presented below:

- Identification/focus (35 bundles): the online homework services, the ministry of education, is one of the, the online homework service, the amount of homework, destructive force in the, one of the most, to do their homework, to solve the problem, Mrs Spence is a, of online homework services, this kind of service, online homework services, what they have learned, that it is not, using online homework services, Mrs Spence is the, footpaths in Siam Square, lesson with Mr Nichols, online homework services, use online homework services, someone else to do, Mrs Spence is not, public safety and security, to play the piano, is not the only, that Thai language is, to pick him up, both sides of the, they do not have, doing homework by themselves, between the narrator and, is the one who, the use of the, one of the main

- Specification of attributes (11 bundles): there are a lot, enough time to, number of weaknesses, too much homework, a considerable number of, a lot of homework, it is a good, the majority of the, a lot of people, a waste of time, an important role in

- Time/place/text reference (9 bundles): the end of the, at the same time, in the food industry, the beginning of the, into the world of, his parents are away, in the boy’s life, in the first place, in the world of

Both subcorpora are similar in the high number of identification/focus lexical bundles (25 bundles in Freshman English and 35 bundles in composition). In both
subcorpora, there are several bundles that are used generally in different writing topics, such as *is one of the, one of the most* and *be one of the* in Freshman English and *that it is not, is not the only and the use of the* in Composition.

However, in most cases, for both subcorpora, the bundles in this category are either lexical bundles that are part of the writing prompt or source materials or lexical bundles that are specific to certain writing prompts but not necessarily part of the original writing prompts. In the former case, in Freshman English, lexical bundles such as *a top fashion model, I used to collect, behaviors of music fans* and *my English I teacher* are phrases that are directly from the writing prompts. In the latter case, some bundles are not in the writing prompts but they are used frequently with one prompt, such as *riding a bicycle is* for the prompt “why more Thai people are riding (their) bicycles these days” and *people in the past* for the prompt “why it is/isn’t important to learn Thai literature.”

In a similar manner, several of the identification/focus bundles in Composition are part of the writing prompts, mainly the keywords of the topic that were being written about such as *the online homework services* and *the online homework service*.

Some bundles are not part of the writing prompts but are specific to one particular writing prompt, such as *this kind of service, the amount of homework, and what they have learned* for the essay prompts on online homework services. Another example is *lesson with Mr Nichols and to play the piano* for the two essay prompts in the final literary analysis essays. These bundles are influenced by the writing topics. For bundles that occurs in the writing for the prompt on online homework services, the frequent use of *the amount of homework* and *what they have learned* are not surprising because the discussion is about hiring someone to do homework for the students and how to solve this problem. The
bundles lesson with Mr Nichols and to play the piano are frequent because in the literary analysis writing, students have to refer back to specific parts of the short story to support their arguments. The bundles refer to important parts in the plot of the story.

In the second sub-category of reference bundles, specification of attributes, most bundles in Freshman English are quantifiers, usually with the phrase “a lot” within, such as there are a lot, a lot of money, have a lot of, has a lot of and there are so many. Composition is also similar in the high use of bundles that are quantifiers, there are a lot, enough time to, number of weaknesses, too much homework and the majority of the. The bundles are also more diverse in the sense that these are not cases of similar variations of “a lot” bundles as with Freshman English.

Freshman English also contains a number of lexical bundles that are time/place/text reference. In this subcorpus, the bundles are primarily used as time or place references. Some examples of time references are when I was a, when I was young, have just woken up, during the final exam, and at the same time. Place reference bundles are either a location relative to the writer or the objects being discussed (in front of the, the left side of, the center of the, in front of me) or an actual place (the faculty of Arts and my high school classroom). Only lexical bundle the end of the also serve the function of text reference in some occurrences in these cases:

- Eventually, at the end of the book, they found out that he was actually a good guy and decided to let go of the prejudice.

- … since I want to know everything in the end of the novel…
Other uses of \textit{the end of the} in Freshman English are as time reference (\textit{the end of the season} or \textit{the end of the day}) and place reference (\textit{the end of the alley} or \textit{the end of the line}).

In composition, the bundles that are time/place/text references are primarily place reference (such as \textit{into the world of} and \textit{in the world of}) with just two cases of time reference (\textit{at the same time} and \textit{his parents are away}).

Similar to in Freshman English, the bundle \textit{the end of the}, and in addition \textit{the beginning of the}, serve double functions; however, they are used almost exclusively as text reference bundles. They are used either to refer to the essays that were being analyzed in the essay analysis writing, such as \textit{“the end of the introduction/last paragraph/essay”} or \textit{“the beginning of the essay/paragraph,”} or to refer to the short story that was being discussed in the essay. In the latter case, most occurrences of this bundle are followed by the noun “story” because the phrase “the end of the story” is in the writing prompt:

- By \textit{the end of the story}, his relationship with the two adults gave him important insights…
- …her care is actually shown since at \textit{the beginning of the story}, …

\section*{4.4 Prompt-specific lexical bundles}

Prompt-specific lexical bundles are lexical bundles that are found within particular writing prompts/in writings for a particular writing prompt only. However, they are not necessarily lexical bundles that are in the writing prompts that were presented to the students.
Table 4.4 shows the number of lexical bundles that are prompt-specific and general in both subcorpora:

Table 4.4: Prompt-specific lexical bundles

<table>
<thead>
<tr>
<th>Subcorpus</th>
<th>Prompt-Specific</th>
<th>General</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman English</td>
<td>27 (25.47%)</td>
<td>79 (74.53%)</td>
</tr>
<tr>
<td>Composition</td>
<td>29 (30.85%)</td>
<td>65 (69.15%)</td>
</tr>
</tbody>
</table>

From the table, we can see that there are slightly more prompt-specific bundles in the Composition subcorpora than in Freshman English.

In the Freshman English subcorpus, 27 out of the total 106 lexical bundles (25.47%) are prompt-specific and are in the writings for 16 prompts. In Composition, prompt-specific bundles make up 30.85% (29 out of the total 94 lexical bundles) of the corpus. These 29 lexical bundles are found in the essays for 6 writing prompts.

Looking at the prompt-specific bundles in Freshman English further, 18 of the 27 prompt-specific bundles are non-original bundles, which means they are phrases in the writing prompts that were incorporated by the writings by the students.

These 18 non-original bundles are found in paragraphs for 16 prompts. In most cases, there is one non-original bundle per one prompt. For example, the bundle have just woken up is in the paragraphs for the prompt “describe yourself when you have just woken up” and the bundle behaviors of music fans for the prompt “extreme behaviors of music/sport/movie fans.” However, there are two writing prompts that generate two non-original bundles. The first one is the prompt “your desk during (the) final exam week” with
the bundles *during the final exam* and *during final exam week*. Both are clearly part of the writing prompt. In some sections, the prompt is given to the students with the definite article “the,” which explains why there are two versions. The second case of multiple bundles for one prompt is the bundles *difficult for many students* and *why reading tests are* for the writing prompt “why reading tests are difficult for many students.” The bundle *difficult for many students* occurs 36 times in the corpus and *why reading tests are* 18 times. There are 9 instances in the corpus where these two bundles are used together as *why reading tests are difficult for many students* as part of their topic sentence.

The other 9 cases of originally-produced but prompt-specific bundles are found in three writing prompts. The first one is the prompt “why more Thai people are riding (their) bicycles these days.” There are 5 prompt-specific lexical bundles, one of which is part of the writing prompt: *Thai people are riding*. The other four, while originally produced, have keyword of the writing prompt, riding and bicycle(s): *riding a bicycle is*, *to ride a bicycle*, *riding bicycles is a* and *riding a bicycle can*.

The second prompt is “why it is/isn’t important to study Thai literature.” One of the four bundles, *to study Thai literature* is from the prompt while the other three are originally-produced by students: *studying Thai literature is*, *should study Thai literature* and *people in the past*. The bundle *people in the past* is the only prompt-specific lexical bundle that does not contain keywords from the writing prompt.

Finally, for the prompt “advantages/disadvantages of being a top fashion model,” there are three prompt-specific lexical bundles: one non-original, “a top fashion model (304)”, and two originally-produced, “there are many disadvantages (24)” and “being a top model (15).”
In Composition, among the 29 prompt-specific bundles, only 8 bundles are non-originally produced. In the Composition subcorpus, bundles are considered to be non-originally produced if they are part of the writing prompts, the model essay analyses that were given as examples, or the essays that were being analyzed in essay analysis.

In 6 writing prompts where there are these prompt-specific bundles, three cases are prompts with only one bundle, which is non-original: into the world of (for the literary analysis prompt “The narrator has gained important insights into the world of adults by the end of the story”), destructive force in the (for the literary analysis prompt “Mrs. Spence is an essentially destructive force in the narrator’s life”) and in the food industry (for the essay analysis writing on an article about meat eating). However, there are two writing prompts that generate two non-original bundles: the online homework services and the online homework service (in essays on the topic), and footpaths in Siam Square and public safety and security (in essay analysis writings on the topics on footpath where these two bundles are used in the original essay and the essay included in the students’ analyses without the quotation marks).

In this subcorpus, the other 21 cases of originally-produced but prompt-specific bundles are found in three writing prompts.

The writing prompt that contains the largest proportion of originally-produced prompt-specific lexical bundles (16 types) is the first essay prompt on the online homework services, with the full writing prompt presented to the students as follows:

“What action, if any, should be taken, and by whom, concerning the “online homework services” that were a recent issue in the local press and were mentioned by Gen. Prayuth in his latest national address?”
The 18 lexical bundles types that are specific to this writing prompt are mostly those that contain the key words related to the main focus of the prompt such as *of online homework services, using online homework services, these online homework services, the amount of homework, to do their homework, have too much homework, and this kind of service*. The only four bundles that do not contain any keyword from the writing prompts are *the Ministry of Education, have enough time to, what they have learned and someone else to do*.

The second prompt that also shows multiple use of originally-produced prompt-specific lexical bundles is the literary analysis prompt: “Mrs. Spence is an essentially destructive force in the narrator’s life.” The bundles are *shows that Mrs. Spence, Mrs. spence is not, in the boy’s life and his parents are away*. The first three lexical bundles contain keywords from the prompt and the last one, while containing no keywords, is also used quite frequently in the students’ writing to refer to a period of time that is important to the plot and to the writing prompt.

The third prompt with original bundles is an essay assignment on the topic of Thai language, with the full prompt below:

“In a radio address on July 29, 2014, Thai National Language Day, NCPO chief General Prayuth Chan-ocha expressed concern about the wrong usage of the Thai language and vocabulary that is now prevalent among the Thai public, and among youth in particular. Similarly, in a Bangkok Poll, 84.7% of respondents believe that the quality of Thai language has degraded to a serious level, perhaps because of a lack of
interest in reading, the advent of social media, and simple laziness. What do you think?”

The lexical bundle that is specific to this prompt is that Thai language is and in most cases this lexical bundle occurs in the context of the verb degrade, which is also in the writing prompt, such as “that Thai language is now degrading” or “that Thai language is being corrupted or degraded.”

Lastly, in essay analyses, in the writings for the essay on footpaths in Siam Square, the lexical bundle Monthita Rojinnakorn argues that is specific to this prompt since it contains the name of the author of the essay.

4.5 Genre-specific lexical bundles

Since prompt-specific bundles discussed in the previous section are genre-specific as the prompts appear in one particular genre (example, reason or description for Freshman English and essay, essay analysis or literary analysis for Composition), this section will focus on bundles that are genre-specific but not prompt-specific.

Table 4.5: Genre-specific lexical bundles

<table>
<thead>
<tr>
<th>Subcorpus</th>
<th>Genre-specific</th>
<th>General</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman English</td>
<td>37 (34.9%)</td>
<td>69 (65.1%)</td>
</tr>
<tr>
<td>Composition</td>
<td>47 (50%)</td>
<td>47 (50%)</td>
</tr>
</tbody>
</table>
Table 4.5 shows that there are more genre-specific bundles in Composition than in Freshman English.

In Freshman English, there are 37 lexical bundles that are specific to a particular genre and 69 bundles are general. Out of the 37 genre-specific bundles, 27 are prompt-specific and were discussed in the previous section and 10 are not prompt-specific.

In Composition, half of the lexical bundles are genre-specific and half are general. Out of the 47 genre-specific bundles, 29 are prompt-specific and previously-discussed. This leaves 18 genres-specific bundles that are not prompt-specific, or in other words, that appear in writings for more than one prompt.

Focusing on the 10 bundles that are genre-specific but not prompt specific, it is found that 2 bundles are in the genre of example writing, 5 bundles in reason writing and 3 in descriptive writing.

For the example paragraphs, the two bundles specific to this genre are *I used to collect* and *there are several ways*. The bundle *I used to collect* is categorized as genre-specific because in the 142 out of 143 instances in the corpus, this lexical bundle appears in the example test for the prompt on things you (or someone you know) used to collect as a child. The other one instance of *I used to collect* occurs in a paragraph with the prompt “favorite childhood toys,” which is an example practice prompt. Therefore, this is not a case of using the bundle from one writing prompt in another subsequent writing since the practice prompt is done before the test.

The other bundle in example writings, *there are several ways*, is more general and used in paragraphs for five different writing prompts in the topic sentence to introduce the
topic of writing, such as “there are several ways to stay healthy” or “there are several ways to improve my English skills.”

The second type of paragraph is reason, there are five lexical bundles that are specific to this genre: why it is important, there are three reason, the last reason is, these are the reasons and for all these reasons.

Similar to the bundle I used to collect in the previous genre, the lexical bundle why it is important is almost prompt-specific because in the 27 instances that they appear in the corpus, 26 are in the reason paragraph in the final exam for the prompt “why it is/isn’t important to study Thai literature.” The other one instance of why it is important is from a reason paragraph practice on wearing school uniforms, which is also written before the final exam prompt on Thai literature.

The other four lexical bundles in reason paragraphs (there are three reason, the last reason is, these are the reasons and for all these reasons) are used as discourse organizers for reason writing.

There are three reason are used primarily to introduce the topic of the writing whereas these are the reasons are mostly found in the concluding part of the paragraph. The last reason is is used to introduce the last supporting reason. Finally, “for all these reasons” is used to precede the conclusion.

In the uses of these bundles in reason writings, it is found that with the bundle the last reason, two cases are used incorrectly by one student. The student used a comma right after the bundle as they would with other transitional expressions such as last of all.

- The last reason is, every tourist can find to have Som Tum all around the country.
- And the last reason is, you will not have time to relax.
The same pattern is also used when the student introduced the second reason with the phrase “the second reason is.”

In the last genre of descriptive paragraphs, the three lexical bundles that are specific to this genre are related to the location or position of the object that are being described: *the right side of, the center of the* and *the back of the.*

For Freshman English, most instances of these lexical bundles reflect the fact that the students are trained to focus on organization and using transitional expressions in their writings to signal the topic sentence, the supporting points and the conclusion in example and reason writing, and to signal the space order, such as left to right or front to back in their description. These are the sections that are explicitly covered in the writing packet in the course materials. In graded assignments, some instructors also deduct points if students fail to use such transitional expressions.

In Composition, 47 genre-specific bundles, 29 of which have already been discussed in the previous section on prompt-specific bundles. This means that there are 18 genre-specific bundles that are found in writings of more than one prompt: 8 bundles are specific to essay writing, 2 are essay analysis and 8 are literary analysis.

The eight bundles that are specific to essays are *one of the* most, *is not going to, as long as the, it is important to, to make sure that, they do not have, a waste of time* and *an important role in.* These bundles are used across several writing prompts.

For essay analysis, the two lexical bundles that are specific to this are *the author argues that* and *a number of weaknesses.* The bundle *the author argues that* is mainly used in the first paragraph of essay analysis writing in order to introduce the main idea or thesis statement of the essays that were being analyzed and written about. The other bundle *a*
number of weaknesses is a phrase in the model essay analysis writing that was given to the students as a guideline on how to write an analysis.

Finally, in the last writing genre, literary analysis, there are eight lexical bundles that are specific to this category: Mrs. Spence is a, Mrs. Spence is the, lesson with Mr Nichols, does not want to, to play the piano, to pick him up, written by Nathaniel Bellows and between the narrator and. The high use of lexical bundles with keywords from the two writing prompts and the short story itself is not surprising because the only literary analysis essays in this subcorpus is on two writing prompts in the final examination and based on only one short story.

Unlike in Freshman English, where several of the genre-specific bundles are discourse organizers particular to the genres, such bundles in Composition are not genre-specific. This is because in all three assignment types, essay, essay analysis and literary essay, students are trained to write in the form of argumentative essays, which include using counterarguments and rebuttal. As such, discourse organizer lexical bundles that are characteristic of argumentative essays, such as it is true that and some might say that, are used across all three writing genres in Composition.

4.6 Originality of lexical bundles

In this study, for the Freshman English corpus analysis lexical bundles are considered to be non-original if they are part of the writing prompts that were presented to the students. Therefore, lexical bundles that are not part of the writing prompts or contain only some words from the writing prompts are considered to be original as in originally-produced by the students. There are also cases where certain lexical bundles can be both
non-original and original. In the Composition subcorpus, lexical bundles are non-original if they are part of the writing prompts that were presented to the students. In the case of essay analysis, lexical bundles are non-original if they are phrases from the original essay that is the focus of the analysis or phrases from the essay analysis models that were given to the students as writing samples. As with the Freshman English subcorpus, there are a few cases here where the lexical bundles can be both original and non-original, which will be discussed subsequently with the examples in this section.

Table 4.6: Originality of lexical bundles

<table>
<thead>
<tr>
<th>Subcorpus</th>
<th>Original</th>
<th>Non-original</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman English</td>
<td>82</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>(77.36%)</td>
<td>(16.98%)</td>
<td>(5.66%)</td>
</tr>
<tr>
<td>Composition</td>
<td>83</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(88.3%)</td>
<td>(8.51%)</td>
<td>(3.2%)</td>
</tr>
</tbody>
</table>

Table 4.6 shows that the majority of the lexical bundles in both learner subcorpora are originally produced by the students. From the percentages of the lexical bundles, there are twice as many non-original bundles in Freshman English than in Composition. Similarly, the number of lexical bundles that are both original and non-original in Freshman English also almost doubles those in Composition.

In Freshman English, the 18 non-original lexical bundles are from 16 writing prompts. Out of these 16 prompts, there are two prompts that have two corresponding lexical bundles. The first one is the prompt “why reading tests are difficult for many students,” for which there are two lexical bundles difficult for many students.
occurrences) and why reading tests are (18 occurrences). However, there are only 9 instances in the corpus where students used the two bundles combined as in the writing prompt, why reading tests are difficult for many students, in their writing. Here are some examples of how the students incorporated parts of the writing prompt in their own writing, primarily as topic sentences and concluding sentences:

- There are many reasons why reading tests are hard for a lot of English One students
- This is why reading tests are big problems for many students.

These are some examples of how students used a part of the writing prompt in their writing and chose to use synonyms for the other part (“hard” and “big problems” instead of “difficult”) to avoid copying the whole prompt. There are also cases where students changed the prompt from an independent clause to a dependent one by removing “why” and some added phrases to signal that they are writing a reason paragraph:

- Reading tests are difficult for many students.
- Reading tests are difficult for many students for many reasons.

The second prompt with two non-original lexical bundles is “your desk during (the) final exam week.” The two lexical bundles are during the final exam and during final exam week. Both are categorized as non-original because there are different versions of the prompt presented to the students in different section, with and without the article “the.” The frequent use of these bundles from the writing prompt is not surprising since it is the main idea of the writing:

- My desk during the final exam may be one of the most disgusting things in the world.
- *During the final exam* week my desk turns into World War II.
- My desk suddenly becomes alive *during the final exam week*.
- My desk is a disaster *during final exam week*.
- My study area *during final exam week* was horrible tangled.

The other 14 non-original lexical bundles are from 14 writing prompts, such as the lexical bundle *have just woken up* from the prompt “myself when I’ve just woken up” (a description prompt) or *celebrate a friend’s birthday* from the prompt “ways to celebrate a friend’s birthday” (an example prompt). While these bundles are primarily used once or twice in the paragraphs in the topic sentences or the concluding sentences, there are also cases where the bundles are used repeatedly throughout a short paragraph. The following example is from the prompt “Why keeping up with fashion is important for so many teenagers” and its non-original bundle *keeping up with fashion*. In one writing, the bundle is used six times:

Many teenagers these days are trying so hard to *keeping up with fashion* but *keeping up with fashion* too much is not a good idea because every fashion is come and go only style is eternal. Yves Saint Laurent said that. First of all, teenagers try to *keeping up with fashion* because they want to follow the trend they may think follow the new trend can make them poppular or maybe they *keeping up with fashion* for make themselves to be accept from other. Moreover, they try to *keeping up with fashion* for show off especially in social media many teenagers have post their photo dressing like they’re going 30 years old tomorrow. So, these several reasons make
me think that we don’t have to *keeping up with fashion* all time to make ourselves look good but we have to know our style and dress in our way, we don’t have to follow every new trend to be good looking.

The usage of the non-original bundles is also ungrammatical: in 5 of the 6 occurrences, “keeping” in the bundle should be in the form of “keep” such as “*trying so hard to keep up with fashion*” rather than “*trying so hard to keeping up with fashion*” in the first one.

Such cases as this shows that sometimes students use the bundles “as is” from the prompt without changing the parts of speech to be grammatical in their own writing. Part of the reason why students use the bundles this way may stem from the fact that they see the materials from the instructors as a safe choice to use.

As for the lexical bundles that are categorized as both original and non-original bundles, there are six bundles that fall into this category.

In four cases, the occurrences of the lexical bundles are almost all non-original bundles that are part of the writing prompt except for one instance where they are found in writings for other writing prompts.

For example, as previously discussed, the lexical bundle *I used to collect* is used 142 times in the example test prompt “things I used to collect as a child” and appeared only once in an example practice on childhood toys, which was written before the example test. The second lexical bundle is *for all arts students*, which is non-original in its 41 occurrences in the writings for the reason test prompt “why English One should/should not be a required course for all Arts students.” The one instance where this bundle is considered originally-produced is in the final exam writing prompt “why it is/isn’t important to study
Thai literature.” The third case is why it is important, which is part of the writing prompt “why it is/isn’t important to study Thai literature” and occurs 26 times in the writings for this prompt. The only other time where it is non-original bundle is in a reason practice prompt on wearing a uniform. Last, the lexical bundle to help the environment is non-original in the 13 occurrences for the writing prompt “ways to help the environment” and originally-produced in the one time it occurs in the reason test prompt on “why more Thai people are riding bicycles these days.”

The other two cases of lexical bundles that are both original and non-original are where the occurrences are primarily non-original with other original occurrence in several writing prompts. The first one is the proper noun the Faculty of Arts: in 16 out of 22 occurrences, this lexical bundle is part of the reason practice writing prompt on “why I chose to study at the Faculty of Arts” and the 6 other occurrences are for 4 different writing prompts. The other case is to improve my English. This lexical bundle is used non-originally 13 times in writings for the example practice prompt “ways to improve my English skills” whereas the other 4 occurrences are in three different prompts.

In Composition, this subcorpus shows a relatively small number of lexical bundles that are non-original (8 bundles) and three bundles that are both original and non-original.

The 8 non-original bundles are from one essay prompt, two literary analyses and essay analyses.

For the essay prompt, the two lexical bundles, the online homework services and the online homework service, are keywords from the writing prompt on the online homework services. Similarly, the two non-original bundles destructive force in the and into the world of are phrases that are part of the literary writing prompts that were
incorporated in the students’ writings: “Mrs. Spence is an essentially destructive force in the narrator’s life” and “the narrator has gained important insights into the world of adults by the end of the story.”

The other four non-original bundles are from essay analyses. As explained earlier, the bundles are considered non-original if they are either phrases from the original essays that are the focus of the analysis or are phrases in the analysis model writings that were given as guidelines. Three non-original bundles are phrases from the original essay: *in the food industry* is from the essay on eating meat and the other two bundles, *footpaths in Siam Square* and *public safety and security* are from the essay on footpaths in Siam Square. Finally, one non-original bundle, *a number of weaknesses*, is from the analysis model writing and in the subcorpus it is used in 13 analyses of three different essays. In the use of *a number of weaknesses* in some writings, the students also used the same sentence structure frame as the model analyses that were given as samples, changing only a few words in the sentence:

Model writing 1: Unfortunately, there are *a number of weaknesses* in Duangkamol’s argument in favor of a single title for both married and unmarried Thai women. (model writing 1)

- Student’s writing: Unfortunately, there are *a number of weaknesses* in Monthitas defense of the removal of stalls…
- Student’s writing: Unfortunately, there are *a number of weaknesses* lying in her ethos.
Model writing 2: However, despite these considerable strengths, the essay does have *a number of weaknesses* that undermine the overall effectiveness of the Angkana’s argument.

- Student’s writing: However, despite these considerable strengths, the essay does have *a number of weaknesses* that undermine the overall effectiveness of Rojtinnakorn’s argument.

- Student’s writing: However, despite these considerable strengths, the article does have *a number of weaknesses* that undermines the overall effectiveness of the author’s argument.

This issue of students copying the model writing occurs primarily with essay analysis writing. It can be explained by the fact that this writing assignment is relatively new to the students. Even though in class we analyze articles in the course materials in terms of the strengths and weaknesses as the main classroom discussion activity, it is only with this assignment when they are asked to write down the analysis, which is graded on. Students were also given only two examples to model on. While this does not excuse the behavior where some chose to copy the sentences and writing structures, it sheds light on why many students cling on to the model paragraph as their writing template.

Only three lexical bundles are categorized as both original and non-original bundles: *by pointing out that, a considerable number of* and *the end of the*.

For the bundle *by pointing out that*, 12 of its 14 occurrences in the corpus are non-original use where this bundle is a phrase from the essay analysis model that was incorporated in the students’ writing of essay analysis. Like in the essay analysis model, this bundle is used to describe the authors’ strategy in writing (e.g. arguing, rebutting or
refuting a claim, *by pointing out that*...). The other two originally-produced occurrences are in two different essays: an essay on the online homework services and an essay on tourism.

Similarly, the other lexical bundle *a considerable number of* is not originally produced in 10 out of 11 occurrences because it is also part of the essay analysis model that is used in essay analysis writing and they are used to discuss strengths or weaknesses of an essay. The only one originally-produced occurrence is in an essay, the third essay writing task, on public display of affection in Thailand.

It could be argued that these two lexical bundles might be non-original in all their occurrences because the students might have used these bundles in the essays after seeing them in the essay analysis model. However, without interviewing the students who used these bundles in their essay writing about their usage of the bundles in essays, there is no definite answer. As such, these two bundles are considered both original when occurring in essays and non-original when occurring in essay analyses.

The last lexical bundles that are both original and non-original depending on the context of use is *the end of the*. This bundle is relatively high in frequency: with 52 occurrences, this is the third most frequent bundle in the Composition subcorpus. For this bundle, 37 out of 52 instances are considered to be non-original since the bundles are used in literary analysis essays in the final examination. The bundle *the end of the* is in one of the two writing prompts in the final exam: The narrator has gained important insights into the world of adults by the end of the story. This bundle is used in 15 essays for this prompt and in 22 essays written for the other writing prompt. Since students saw both prompts in
the final exam, the use of *the end of the* in the essays for both prompts are considered non-original.

The other 15 uses of *the end of the* that are originally-produced are for many different prompts in different writing tasks.

### 4.7 Shared and unshared lexical bundles

In the Freshman English and Composition subcorpora, there are 12 bundles that are shared between the two. The shared bundles account for 11.3% of the total bundles in the Freshman English subcorpus and 12.8% of the Composition subcorpus. The shared bundles are:

- a lot of people
- it is a good
- one of the most
- at the same time
- last but not least
- the end of the
- have enough time to
- not be able to
- there are a lot
- is one of the
- on the other hand
- will be able to

In most cases, there is no particular pattern of use in either subcorpora that differentiates the use of the bundles. For example, the bundles *a lot of people, it is a good, is one of the, on the other hand, one of the most* are used across different writing tasks and with several different words.

Cases of similar pattern of use in certain bundles are with the use of the bundles *last but not least, not be able to* and *there are a lot*. *Last but not least* is used similarly in both subcorpora to precede the last supporting points in the paragraphs or essays. The bundle *not be able to* is primarily used after the modal verbs *might, may* and *will* in both Freshman English and Composition. The last bundle *there are a lot* is used similarly in
both subcorpora in that in almost all occurrences they are followed by the preposition of. Only two cases in Freshman English and one in Composition where they are used with the adjective more as in “there are a lot more to see” and “there are a lot more car users” in Freshman in English and “there are a lot more negative effects” in Composition.

The shared bundles that are used differently or show different patterns of use are at the same time, have enough time to, the end of the and will be able to.

The bundle at the same time is used in both subcorpora to talk about concurrent events or things but with slightly different patterns of use in each subcorpus. In Freshman English, this bundle is primarily at the end of the clauses or sentences:

- …it’s a great way to exercise and make new friends at the same time.
- … you can travel and work at the same time.
- … have a hobby that is beneficial and enjoyable at the same time.

In Composition, at the same time is at the clause- or sentence-final positions as in the Freshman English subcorpus. However, there are also several occurrences where the bundle is used at the beginning of a phrase, a clause or a sentence:

- ….will help the society, and at the same time asking for the equal treatment towards the homosexual.
- … but at the same time it harms students.
- Some people view the homework service as another abomination that adds to the student’s educational obstructions. At the same time, some people point out that the real threat is what creates the demand for such service in the first place.

In most occurrences of have enough time to in Freshman English, the bundle follows negatives such as do not/don’t and will not/won’t. There is only one case where
the bundle is not used with negative: “they will have enough time to go shopping.” In Composition, about half of the occurrences (9 out of 19) is preceded by negatives such as do/did/does not and will not. The other half is affirmative, such as “will also have enough time to,” “can relax and have enough time to” and “students should have enough time to improve”

The next bundle the end of the is used predominantly in Freshman English with physical locations such as the end of the “line,” “alley,” “leg” or “right side of my lips” or time such as the end of the “day” or “season” whereas in Composition, the bundle is mostly used as text reference to refer to the essay that is being analyzed “the end of the introduction/last paragraph/essay” or with the noun “story” (e.g. by/at the end of the story) because the phrase “the end of the story” is in the writing prompt.

The last bundle will be able to is used with different pronouns within the two subcorpora. In Freshman English, about half of the occurrences of the bundle is with 1st or 2nd person pronoun, such as I/you/we will be able to. The other half are either 3rd person pronoun or other pronouns, such as they/she/students will be able to. In composition, the majority of the bundle will be able to is used after 3rd person pronoun “they” or other pronouns, such as teacher/students/tourists/people will be able to, and only one 2nd person pronoun (we will be able to).

In short, the shared lexical bundles in the two subcorpora are mainly used similarly in that they present no particular pattern in their usage or that they share a certain pattern of use. However, there are also a number of bundles that are used differently by the two group of participants when it comes to what words are used with those bundles and where the bundles are used.
4.8 Discussion

From the results presented in the previous sections, the two research questions that are addressed in this section are:

Research questions 1: What are the frequent lexical bundles that are used by the students in both subcorpora?

Research questions 2: How are the two groups of students similar or different in their use of lexical bundles?

The first research question on what lexical bundles are frequently used by both groups is discussed in detail in the previous sections. There are 106 lexical bundles that are frequently used by the Freshman English students and 94 lexical bundles by the Composition students.

To address the second research question on how the two groups are similar or different in their use of lexical bundles, looking at the total number of lexical bundles, we can see that the participants in the Freshman English subcorpus, the beginner or novice in academic writing, uses both more lexical bundle types and tokens than the Composition subcorpus, the expert in academic writing at the undergraduate level at this university. The number of lexical bundle types is slightly higher in Freshman English than in Composition (106 vs. 94). However, in Freshman English the tokens of lexical bundles (3,279), or the actual occurrences of the lexical bundles, are almost twice the number in Composition (1,702). This result is in line with previous studies where the novice groups are found to use more lexical bundle types and tokens than the expert groups (Jalali et al., 2008; Wei & Lei, 2011; Weinstein, 2011). However, it should be noted that the novice group in previous studies are at a much higher level than the novice group in this study. In the study in Jalali
et al. (2008), the novice group data is from graduate-level theses and dissertations. Similarly, in the research into Chinese EFL learners, Wei and Lei (2011) used doctoral dissertations to represent the novice group. In Weinstein (2011), the novice corpus is from writings by all levels of undergraduate students. In this dissertation, the novice group is first-year undergraduate students who had just started this first English course with a writing component at university level.

The difference in the lexical bundle tokens between both subcorpora, 3,279 in Freshman English and 1,702 in Composition, can be partly attributed to the high frequency of non-original bundles, or bundles that are part of the writing prompts and materials. In the Freshman English subcorpus, there are 24 types of such bundles, which is almost a quarter of the total types and those bundles are about one-third of the total lexical bundle tokens (1,139 out of 3,279 bundles). In Composition, there are only 11 bundles that are non-original and their occurrences are only 16% of the total lexical bundle tokens (286 out of 1,702).

It could be argued that since these bundles are not originally produced by the students and are part of the writing prompts or course materials that were from the teachers, investigating how these bundles are used might not be useful. However, it is important to look at how students interact with the writing prompts and how the bundles are used in context. In a study by Appel (2011) where the corpus data is from essays of the same writing prompt, the study also explored how different levels of learners (low and high proficiency) make use of the prompts that are available to them from the writing prompt and reading materials. Although the focus is different from this dissertation where learners are seen as novice-experts rather than low-high proficiency, lower proficiency learners in
Appel’s study used more phrases from the prompt and the reading materials than the higher group. Appel concluded that this is one strategy to make up for their lack of proficiency.

In the present study, as discussed earlier, the high use of phrases from the writing prompt as non-original bundles in Freshman English is mainly because how the prompt is presented to the students. Writing prompts are concise and in the form of a noun phrase or an independent clause, such as in the final exam paper:

- Example: advantages/disadvantages of being a top fashion model
- Reason: Why it is/isn’t important to study Thai literature

The prompts in Freshman English are quite specific in the sense that the students are not required to narrow the writing topic down any further. In most cases, writing prompts can be readily used as a topic sentence with minor changes:

- **Being a top fashion model** is not good as any other people’s imagination.
- There are many **advantages of being a top fashion model**.

In addition to using the prompts as part of the writing, in several cases, phrases with keywords from the prompts are also used repeatedly throughout the writing:

There are many advantages of being **a top fashion model**. First of all, when someone being **a top fashion model**, they must be a good figure. They must always take care themself and make them have healthy body. Secondly, being **a top fashion model** make them to be good mental health and confidentially person. They must have been training for a model, so they must concentrate with something in front of them all the time when they walk on cat walk. They also can solve the problem when they have any
accident. Last of all, being a top fashion model can make them wealthy. If they are a top fashion model of the world, they will be paid by a large of money. They can use money to do something what they want. So, there are examples of advantages of being a top fashion model and I think that some people want to be a top fashion model for sure.

The example above is not an isolated case; many paragraphs contain lexical bundles from the writing prompts that are used repeatedly in this manner. Such frequent use of phrases from the writing prompt can be explained by the alignment principle from the sociocognitive approach where human adjust themselves to their environment, both “human” and “non-human environments” in social interactions (Nishino & Atkinson, 2015, p. 39). In academic writing, the most immediate environment for the students are the writing prompts from the instructor and in some cases the course materials. The student picked up a key phrase, generally with the main keywords, from the prompt they chose and used them rather than switching to pronouns or synonyms. This is because materials from instructors are seen as a safe choice, especially when it comes to graded assignments, which is a high stake to the students. It is better to use phrases that are already available to them and are grammatically correct rather than trying to come up with alternative phrases in paraphrasing. Students might know that in order to “survive” the writing assignment, so to speak, it might be safer to adhere to what is given from the instructors than to experiment with new words or phrases. As seen in the sample of one student’s paragraph, using a top fashion model repeatedly as a unit without changing anything in that unit, such as changing to the plural “top fashion models” to agree with the plural sentence subject, is one clear
example of adhering and aligning to what they might perceive as teacher-approved lexical items and are safe to use.

In Composition, the use of non-original bundles is less frequent than the Freshman English and also slightly different. There are non-original bundles that are keywords from the writing prompts or the essay being analyzed and also non-original prompts that are from the model writing of essay analysis. The use of non-original prompts that are from the model writings is usually in the form of a larger structure than just a four-word bundle. From the example in the previous section, examining the use of the non-original bundle *a number of weaknesses* in context reveals that in most of their usage, the students also used a similar sentence frame as the model writing:

Model writing 1: Unfortunately, there are *a number of weaknesses* in Duangkamol’s argument in favor of a single title for both married and unmarried Thai women. (model writing 1)

Student’s writing: Unfortunately, there are *a number of weaknesses* in Monthitas defense of the removal of stalls…

The function of the longer “copied” phrases in students’ writings is also similar to the model writing. In the model writing, this sentence “ Unfortunately,…” serves as a discourse organizer when the writer wanted to move from discussing the strengths in the essays to the weakness, which are the main issues in the analysis. Students also used this sentence in a similar manner. Such usage can also be explained by the alignment principle. In this new task, analysis writing, with only a few examples on how to write, some students might think that it is safer, if not easier, to use the approved sentence parts from the model writing given by the instructors. In fact, students were only given two similarly written
model analyses for this new writing task. As such, it might seem safer for students to use such structure in order to show the instructors that they know how to write a proper analysis in the same way as the instructors do in the model writing.

In short, the Freshman English and the Composition groups are different in the number of the lexical bundle types and tokens they used. The difference is partly attributed to the number of non-original bundles that are used by each group and the use of these non-original bundles is important to the analysis since it reveals some issues about the students’ writing strategies.

Structurally, the two group are similar in their high use of lexical bundles that are phrasal (verb phrases, noun phrases and prepositional phrases) and low use of dependent clause structure. The similarities and differences are further highlighted in the subcategories of each main structure.

There are several sub-structures that are frequent in both subcorpora, such as non-passive verb phrase, noun/noun phrase + verb phrase, WH-clause, to-clause, noun phrase with of-phrase, noun phrase with other post modifiers, other noun phrase and prepositional phrases. However, some structures are frequent in one subcorpus but are few in the other such as 1st and 2nd person pronouns verb phrase in Freshman English, 3rd person pronoun verb phrase in Composition, and “there is/are” phrases in Freshman English. There are also certain substructures that are found in one learner subcorpus but not in the other: passive verb phrases in Composition, if-clause in Freshman English, that-clause in Composition and comparative expressions in Composition.

The difference in the use of some substructures can be explained by the nature of the topics of writing and how the students write. For example, in Freshman English, several
of the writing prompts relate to the personal experience of the students. The selection of such writing prompt is intentional on the instructors’ part because the students are considered beginners in the writing course and the instructors wanted to ensure that the topics they were given to write about will be relatively easy to brainstorm ideas for within the one-hour time slot for the graded assignments. Therefore, one frequent substructure for Freshman English is 1st and 2nd person pronouns verb phrase. Some students also prefer to draw on personal experiences as supporting points for topics that are not personal experience, resulting in the use of those 1st and 2nd person pronoun bundles. The composition topics, on the other hands, are much less personal but some are still relevant to the students. Thus, there are fewer use of 1st and 2nd person pronoun bundles in Composition.

In terms of function, both groups are similar in that the majority of the lexical bundles are referential expressions. Further similarities and also differences can be observed in the functional subcategories. With stance expressions, the bundles in both subcorpora are almost all attitude/modality stance expressions; however, in the Freshman English subcorpus, the bundles are primarily obligations/directive bundles and in Composition there are more intention/prediction and ability bundles than other stance expressions.

With discourse organizing bundles, the two group are also different in the use of the bundles in the subcategories: more topic introduction/focus bundles in Freshman English and more topic elaboration/clarification in Composition. The reason behind the difference lies in the difference between the writings in the two courses. Most writing in Freshman English are short paragraphs, with the emphasis put on having one topic
sentence, three supporting points and one concluding sentence. Most of the bundles found are therefore those that are obvious topic introduction/focus bundles such as “there are” + topic of paragraph.

In the last functional category, referential expressions, the difference is in the use of time/place/text reference bundles. In the Freshman English subcorpus, most of the bundles are time or place references whereas in composition most of the bundles are place references, with two bundles that are almost exclusively text references, referring to the essays that were being analyzed or the short story.

With lexical bundle structures and functions, looking at the main categories alone, both subcorpora appear to be similar; however, a closer look into the subcategories and lexical bundle use in context can reveal the differences between the two group.

Examining lexical bundle use in context also reveals ungrammatical use of lexical bundles in Freshman English that is not present in the Composition subcorpus. The first example has to do with the lexical item in the bundle. The bundle *in the same time* is used where the correct form for the meaning intended should be *at the same time*, a bundle which is also frequently used in the corpus. The incorrect choice of preposition is most likely from Thai, the students’ first language, in which the preposition “at” is used for the same expression.

Another case of ungrammatical use of lexical bundles in Freshman English is with structure. There is a case where the student used the lexical bundles that are from the writing prompt as a whole without changing the part of speech at all. With the bundle *keeping up with fashion*, some instances of the bundle use in context is incorrectly following “to” as in “trying so hard to *keeping up with fashion.*”
These ungrammatical uses of lexical bundles are found in the Freshman English subcorpus, the novice group.
CHAPTER 5

Use of Lexical Bundles on the Academic Formulas List (AFL) in the Learner Subcorpora

This chapter is the second part of the analysis. In this chapter, the discussion is on comparing the lexical bundles in the writing of participants in the Freshman English and Composition subcorpora with the target bundles from the academic formulas list (AFL).

The results and discussion will follow the categories presented in the AFL. Overall AFL lexical bundle results will be presented first, followed by bundles in different functional categories.

The AFL contains a total of 607 academic formulas (or lexical bundles) categorized according to their main functions: referential function, stance expression and discourse organizing function. Each functional category contains subcategories and the bundles are then grouped according to their register as core AFL (written and spoken), primarily spoken and primarily written.

5.1 Overall results

Table 5.1 is divided into two sections: the first is overall comparison and the second comparison in the three functional categories. There are five result columns in both sections. The first result column, labeled AFL, is the number of the lexical bundle types in the Academic Formulas List, which is provided for reference. The second column is the result from the Freshman English subcorpus (FE) that includes all the bundles, and the third is also the result from the Freshman English subcorpus but for bundles whose frequency is
over 40 words per million (FE 40). Similarly, the fourth column is for the result from the Composition subcorpus (Compo) and the last column is for Composition bundles with frequency of 40 per million words and over (Compo 40).
Table 5.1: AFL bundles in Freshman English and Composition subcorpora

Section 1. Overall Comparison

<table>
<thead>
<tr>
<th></th>
<th>AFL</th>
<th>FE</th>
<th>FE 40</th>
<th>Compo</th>
<th>Compo 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core AFL written &amp; spoken</td>
<td>207</td>
<td>125</td>
<td>28</td>
<td>176</td>
<td>69</td>
</tr>
<tr>
<td>Primarily spoken</td>
<td>200</td>
<td>98</td>
<td>32</td>
<td>76</td>
<td>24</td>
</tr>
<tr>
<td>Primarily written</td>
<td>200</td>
<td>78</td>
<td>10</td>
<td>130</td>
<td>33</td>
</tr>
<tr>
<td>Total</td>
<td>607</td>
<td>301</td>
<td>70</td>
<td>382</td>
<td>126</td>
</tr>
</tbody>
</table>

Section 2. Comparison by Functional Categories

**Category A. Referential Function**

<table>
<thead>
<tr>
<th></th>
<th>AFL</th>
<th>FE</th>
<th>FE 40</th>
<th>Compo</th>
<th>Compo 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core AFL written &amp; spoken</td>
<td>165</td>
<td>94</td>
<td>17</td>
<td>139</td>
<td>48</td>
</tr>
<tr>
<td>Primarily spoken</td>
<td>39</td>
<td>20</td>
<td>8</td>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td>Primarily written</td>
<td>87</td>
<td>34</td>
<td>6</td>
<td>56</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>291</td>
<td>148</td>
<td>31</td>
<td>218</td>
<td>71</td>
</tr>
</tbody>
</table>

**Category B. Stance Expression**

<table>
<thead>
<tr>
<th></th>
<th>AFL</th>
<th>FE</th>
<th>FE 40</th>
<th>Compo</th>
<th>Compo 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core AFL written &amp; spoken</td>
<td>17</td>
<td>14</td>
<td>1</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>Primarily spoken</td>
<td>79</td>
<td>44</td>
<td>14</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>Primarily written</td>
<td>81</td>
<td>36</td>
<td>4</td>
<td>62</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>177</td>
<td>94</td>
<td>19</td>
<td>106</td>
<td>37</td>
</tr>
</tbody>
</table>

**Category C. Discourse Organizing Functions**

<table>
<thead>
<tr>
<th></th>
<th>AFL</th>
<th>FE</th>
<th>FE 40</th>
<th>Compo</th>
<th>Compo 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core AFL written &amp; spoken</td>
<td>25</td>
<td>17</td>
<td>10</td>
<td>23</td>
<td>12</td>
</tr>
<tr>
<td>Primarily spoken</td>
<td>82</td>
<td>34</td>
<td>10</td>
<td>23</td>
<td>5</td>
</tr>
<tr>
<td>Primarily written</td>
<td>32</td>
<td>8</td>
<td>1</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>139</td>
<td>59</td>
<td>20</td>
<td>58</td>
<td>18</td>
</tr>
</tbody>
</table>
When taking all occurrences of the target bundles into account, we can see that there are more target AFL bundles in the Composition subcorpus (382 types or 62.93% of the AFL bundles) than in the Freshman English subcorpus (301 types or 49.6%).

The difference is more pronounced when filtering the results further to include only frequent target bundles. When looking at only target AFL bundles that occur at 40 times and more per million words, which translates into a raw frequency cut-off point of 14 bundles for the Freshman English subcorpus and of 10 for Composition, there are almost twice as many target bundles in the Composition (126 types or 20.8%) than in the Freshman English subcorpus (70 types or 11.5%). These numbers show that the more expert students in the Composition subcorpus are progressing more towards using the academic bundles in the academic formulas list than the beginner students in the Freshman English year.

The first section of Table 5.1 also summarizes the number of lexical types in three different categories of core AFL (both written and spoken bundles), primarily spoken and primarily written.

With all instances of the AFL target bundles in account, both the Freshman English and Composition subcorpora contain more core AFL bundles than spoken and written ones. In the Freshman English subcorpus, 125 bundles are core AFL (60.4% of the 207 core AFL target bundles) while in Composition, 176 bundles are core AFL (85% of the core AFL target bundles).

As for spoken and written bundles, the Freshman English subcorpus has more spoken target bundles than written ones (98 spoken vs. 78 written bundles). In contrast, Composition has considerably more written than spoken bundles (130 written vs. 76
spoken bundles). Freshman English also uses more spoken bundles in the AFL list than Composition (98 bundles vs. 76 bundles).

When taking frequency of use into consideration and only including lexical bundles that occur 40 times or more per million words in each subcorpus, the proportions of core, spoken and written target bundles in the subcorpora change slightly.

Out of the 70 frequent AFL bundles in the Freshman English subcorpus, spoken bundles is the largest category (with 32 bundles or 45.7%), followed by core bundles at 28 bundles (40%). There are only 10 bundles that are in the written, which is a small percentage of the frequent bundles of the subcorpus (14.3%).

In the Composition subcorpus, out of the 126 frequent target bundles, over half are core bundles (69 bundles or 54.8%) and there are more written bundles (33 bundles or 26.1%) than spoken ones (24 bundles or 19%).

Overall, we can say that the Freshman English group uses more core and spoken target bundles than written ones while the Composition group uses primarily core bundles, with more occurrences of written than spoken target bundles.

It can be concluded from this section of the table that the participants in the Composition subcorpus are already progressing towards the usage of written academic bundles in their writing, which sets them apart from the novice/beginner writers in the Freshman English group, who demonstrate more usage of spoken academic bundles.
5.2 Functional categories

This section focuses on the results presented in the second section of Table 5.1, which pertain to the three functional categories, i.e., referential functions, stance expressions and discourse organizing functions. In the AFL, the three categories are also divided further into subcategories as well as into core, spoken and written bundles.

The three functional categories make up different proportions in the AFL with referential function as the largest category (47.94% or 291 bundles), followed by stance expression (29.16% or 177 bundles), and discourse organizing functions (22.9% or 139 bundles).

The discussion in this section will be organized by the three categories, with the corresponding sections in Table 5.2 reproduced in each section.

5.2.1 Referential expressions (category A)

Table 5.2: Referential expressions

<table>
<thead>
<tr>
<th></th>
<th>AFL</th>
<th>FE</th>
<th>FE 40</th>
<th>Compo</th>
<th>Compo 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core AFL written &amp; spoken</td>
<td>165</td>
<td>94</td>
<td>17</td>
<td>139</td>
<td>48</td>
</tr>
<tr>
<td>Primarily spoken</td>
<td>39</td>
<td>20</td>
<td>8</td>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td>Primarily written</td>
<td>87</td>
<td>34</td>
<td>6</td>
<td>56</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>291</td>
<td>148</td>
<td>31</td>
<td>218</td>
<td>71</td>
</tr>
</tbody>
</table>

In this first functional category, the proportions of core, spoken and written bundles in the AFL are different, with core being the largest (165 bundles or 56.7%) and spoken the smallest (only 39 bundles out of the total 291 referential bundles, or 12.3%). The number of written bundles is also over twice as larger as the spoken ones. It is not surprising
to see that the target bundle occurrences in the learner subcorpora also follow the same proportion structure.

In the case of all occurrences of the target AFL bundles, in both the Freshman English and Composition subcorpora, the proportions are similar to those in the AFL, with core referential bundles as the largest category (about 63% for both) and with more written than spoken bundles.

Narrowing down the target bundles in the learner corpus to only high frequency bundles, core bundles still make up the largest category for both subcorpora (17 bundles or 54.8% and 48 bundles or 67.6%, respectively), with almost equal numbers of spoken and written AFL bundles. In the Freshman English subcorpus, out of the 31 referential bundles, there are 8 spoken bundles (25.8%) and 6 written ones (19.4%). In Composition, there is only one more written bundle than spoken (12 bundles or 16.9% and 11 bundles or 15.5%, respectively).

5.2.1.1 Core referential expression bundles

In terms of frequent core referential bundles, in the Freshman English subcorpus, only 2 out of the 31 bundles are unique to this group. The rest of the bundles are shared with the composition subcorpus. These two bundles are *there are three* (quantity specification) and *have the same* (contrast and comparison).

The frequent occurrence of *there are three* that is unique in the Freshman English subcorpus is expected, since this is how many students write their topic sentences for example and reason paragraphs. This bundle occurs 64 times in the corpus and for the majority of the occurrences (59 times), *there are three* appears at the beginning of a topic
sentence and of a paragraph, mostly followed by phrases from the writing prompts or paraphrases of the writing prompts:

- **There are three** things I used to collect as a child.
  (Prompt: Things you or someone I know used to collect as a child)

- **There are three** extreme behaviors of Korean music fans.
  (Prompt: Extreme behaviors of music/sport/movie fans)

- In school and university, **there are three** major sources making students feel stressed. (Prompt: Sources of stress in students)

- **There are three** main reasons why atychiphobia (fear of failure) is a bad fear to have. (Prompt: Why atychiphobia (fear of failure) is a good/bad fear to have)

- **There are three** main reasons for you to study Thai literature.
  (Prompt: Why it is/isn’t important to study Thai literature)

This bundle *there are three* appears to have come from a similar sequence that occurs in a topic sentence from a model paragraph given in the writing packet, which is part of the official course materials for students. In the packet, one of the model paragraphs contains the following sentence as the topic sentence: At [university name] *there are many* useful majors that students can choose from. Since students are encouraged to have a proper topic sentence for each paragraph and are in some class sections with certain instructors strictly graded on that, many students choose to incorporate phrases from sample paragraphs in the course materials in their writing. However, the writing packet in the course materials does not cover how to write a topic sentence at all.
As for the other core referential bundles in the Freshman English, have the same, this bundle is used across different writing prompts and different paragraph types.

In the Composition subcorpus, core referential bundles make up the largest category, and 33 out of the 48 bundles in this category are unique to this subcorpus (i.e., not shared with the Freshmen English subcorpus). These bundles are summarized below:

- intangible framing attributes (15 bundles): *the problem of, in terms of, the process of, focus on the, is based on, fact that the, the fact that the, the idea that, point of view, the issue of, the nature of, the concept of, the case of, in the case, the use of*
- tangible framing attributes (4 bundles): *the amount of, parts of the, the value of, the level of*
- quantity specification (2 bundles): *large number of, the number of*
- identification and focus (9 bundles): *that it is, an example of, the example of, is not the, is that the, this is not, it does not, that there is, this is a*
- contrast and comparison (2 bundles): *related to the, different from the*
- deictics and locatives (1 bundle): *the real world*

As can be seen from the list, the bundles are mainly specification of attributes bundles (intangible framing attributes, tangible framing attributes, and quantity specification), and identification and focus bundles. Most core referential bundles in Composition are used across different topics and writing assignments. There are two bundles, however, that seem to be more specific to a particular writing prompt than others: *the problem of and the amount of*. The first bundle *the problem of* appears 23 times in this subcorpus, among which 16 are used in essays for the writing prompt on online homework services:
- Therefore, the Ministry of Education is the most appropriate to solve *the problem* of online homework services…
- …it will also gradually help solve *the problem of* homework overload and grade competition
- Thus, in order to solve *the problem of* online services at its core, …
- The Online Homework Service problem is like the mirror that reflects *the problem* of Thai Education System…
- In conclusion, *the problem of* students hiring online homework services should be solved at its causes

The other bundle, *the amount of*, occurs 73 times in the subcorpus, 39 of which are followed by the noun “homework” in the online homework services essay prompt:
- … have not enough time comparing *the amount of* homework they have to finish.
- … teachers have to reduce *the amount of* homework given.
- … a lot of students together with parents also feel that *the amount of* homework is overwhelming…
- When *the amount of* homework is decreased, …

Such patterns of use for these two bundles can be explained by the fact that in the first essay writing task, a large number of students (63 out of 242) chose to write on the prompt on online homework services. In addition, the articles that were used as part of the prompt also frame the issue of online homework services as a major problem. Even though the words “*problem*” and “*amount*” did not appear in the articles and the writing prompt, it is not difficult to see why these students used the words and the bundles *the problem of* and *the amount of* to discuss their opinions on the issue.
The two subcorpora also share 15 core referential bundles:

- intangible framing attributes (2 bundles): the fact that, the meaning of

- tangible framing attributes (1 bundle): part of the

- quantity specification (1 bundle): a number of

- identification and focus (10 bundles): it is not, is not a, is that it, there is a, there is no, such as the, it can be, that there are, which is not, which is the

contrast and comparison (1 bundle): in the same

From the list it is clear that most shared bundles are identification and focus ones. In most cases, the two groups use these shared bundles across different writing topics and writing tasks. However, there are certain bundles that are used differently or show different patterns of use by the two group. The first case is there is a. In Freshman English, this is used primarily with concrete things. In most cases, this bundle is used in descriptive writing:

- …there is a beautiful woman standing there…

- …there is a dog who seems to be…

- On my desk, there is a big pile of paper…

The bundle is also used in example and reason paragraph writing as well:

- …there is a fresh game where you can play sports…

- There is a heavy traffic in the city.

- Finally, there is a well-known social networking application called Facebook.

The use of there is a in the Composition group is mostly with concepts, such as:

- For example, there is a generalization in paragraph 3 and 4.

- Secondly, there is a belief that fanfictions misrepresent characters…
- there is a chance that students will really learn the subject…
- However, there is a thin line between language degradation and language change.

Another example of different patterns of use of a shared bundle between the two subcorpora is with the bundle it can be. In, Freshman English, the bundle is followed by a past participle in the passive structure (9 occurrences), a noun (6 occurrences), and an adjective or adjective phrase (5 occurrences):

- … and it can be adopted to use in any other subject or in real life. (passive)
- … it can be kept as long as the owners want to. (passive)
- … it can be dangerous… (adjective)
- … it can be somewhat difficult… (adjective phrase)
- … it can be anything… (noun)
- … It can be dry food… (noun phrase)

In the passive structure with it can be, 3 out of 9 cases are where the past participle is not in the correct form: it can be do (done), it can be rode (ridden), and it can be hold entirely in one hand (held).

With the composition subcorpus, almost all occurrences of it can be, except for two, is in a passive structure with past participle:

- It can be concluded that service charges that restaurants charges
- It can be inferred that people might not be aware of their expenditure
- … but for some it can be seen as vulgar…
- … and it can be said that this is what most coffee shops embrace and plan for.
The two non-passive use of *it can be* is followed by an adjective phrase (*it can be* immensely dangerous for passerby...) and a noun clause (*...it can be* that Mrs. Spence has started a slow progress...).

The bundle *a number of* also shows a particular pattern of use in the Composition subcorpus but not in Freshman English. In Freshman English, this bundle is followed by many different nouns and noun phrases such as “people”, “cars”, and “unusual hobbies.” In Composition, however, 13 out of 49 instances of *a number of* are used with the noun “weaknesses” only in the essay analysis tasks:

- … the essay does have *a number of weaknesses* that undermine…
- Unfortunately, there are *a number of weaknesses*…
- Despite *a number of weaknesses*, however, there are also many strengths…

These uses of *a number of* with the noun *weaknesses* are a direct influence from the two model writings of essay analysis that is included in the course materials for the students. In the model writings, the four-word bundle *a number of weaknesses* is used in several sentences, such as:

- Unfortunately, there are *a number of weaknesses* in Duangkamol’s argument in favor of a single title for both married and unmarried Thai women. (model writing 1)
- However, despite these considerable strengths, the essay does have *a number of weaknesses* that undermine the overall effectiveness of the Angkana’s argument. (model writing 2)

In some of the student’s writings, when the bundle *a number of* is used with “weaknesses,” the students also used the same sentence structure frame as the model
writings, changing only a few words in the sentence. From the example above from model writing 1, there are three cases where students used the same sentence beginning:

- Unfortunately, there are a number of weaknesses in Monthita in terms of pathos.
- Unfortunately, there are a number of weaknesses in Monthitas defense of the removal of stalls…
- Unfortunately, there are a number of weaknesses lying in her ethos.

Similarly, from model writing 2, there are four cases of students copying the sentence from the model:

- However, despite these considerable strengths, the essay does have a number of weaknesses that undermine the overall effectiveness of Rojtinnakorn’s argument.
- Despite these considerable strengths, the essay does have a number of weaknesses that undermines the overall effectiveness of Monthita’s argument.
- However, despite these considerable strengths, the essay does have a number of weaknesses that undermine the other effectiveness of the writer’s arguments.
- However, despite these considerable strengths, the article does have a number of weaknesses that undermines the overall effectiveness of the author’s argument.

5.2.1.2 Spoken referential expression bundles

With spoken referential bundles, in Freshman English, only 2 of the 8 bundles are unique to the subcorpus and 6 are shared with composition. The two are identification and focus bundles: there was a and the best way to. The first bundle there was a is predominantly used in description paragraphs to describe concrete things such as “there was a big space under the sinks…” and “there was a big white bay window” and “there
was a half-empty glass.” The other bundle the best way to is used in many different writing prompts and paragraph genres.

In Composition, there are 11 spoken referential bundles and 5 are unique to this group, which are intangible framing attributes (the idea of, the kind of), identification and focus (and this is), and contrast and comparison (to do with, to each other). There 5 bundles are used across writing prompts and writing genres. There is one bundle that demonstrates a particular pattern of use: to do with. This bundle is used primarily in the pattern “has/have nothing to do with + noun” such as:

- homosexual parenting has nothing to do with a child’s psychology
- religious and societal values have nothing to do with an actual value of an occupation

The 6 shared bundles are intangible framing attributes (this kind of), identification and focus (this is the) and deictics and locatives (at the end, at the end of, the end of, the end of the).

The four deictics and locatives bundles are actually overlapping with the shared phrase “the end” and these four are used differently in Freshman English and Composition.

In Freshman English, they are used to reference both time, physical locations and textual locations, such as:

- He then breaths violently with hurt and passed away at the end. (time)
- At the end of the day … (time)
- Just like the end of the battlefield where the injured soldiers fully lying over the ground (physical location)
- The line waiting for load the package is so long that you can’t ever see *the end of* it. (physical location)

- since I want to know everything in *the end of the* novel (textual location)

- For instance, some of *the end of the* Jamsais book always happy (textual location)

In composition, these four bundles are mostly used in essay analysis and literary essay to refer to textual locations. In essay analysis, those bundles are used to refer to the locations in the essay that was being analyzed:

- The writer also uses emotional appeal effectively in *the end of* P.4,…

- Additionally, *at the end of* the essay, her message is directly sent to the audience, …

- …. before reading *the end of the* article…

In literary essays, these bundles are used when students referenced back to the short story as supporting details for their arguments.

- …when the narrator accuses her of being a liar *at the end*.

- *At the end of the* story, the narrator decides to resolve his problem himself.

- Some readers might say that they don’t see such transformation since by *the end of* the story…

The other two shared bundles *this kind of* and *this is the* are used in various writing prompts and genres; however, with *this kind of* in composition, there are some words that are used with this bundle more than others. In the 78 instances of *this kind of*, 7 are followed by “language” and found in the prompt on degradation of Thai language; 7 are followed by “homework” and found in online homework service prompt; 21 are followed by the
noun “service” and 6 by the noun “services”, both of which are also in the online homework service essay prompt.

5.2.1.3 **Written referential expression bundles**

In the Freshman English subcorpus, 4 of the 6 written referential bundles are shared with composition, leaving only 2 bundles that are not shared: one intangible framing attribute bundle (*there are several*) and one identification and focus bundle (*his or her*). The bundle *his or her* is found to be used in different writing prompts and genres. The bundle *there are several*, on the other hand, are primarily used as part of the topic sentence along with the phrases or paraphrases with the key words from the prompt:

- *There are several* amazing places to visit in Bangkok.
- (Prompt: Places for tourists to visit in Thailand)
- *There are several* disadvantages of being a top fashion model.
- (Prompt: Advantages/disadvantages of being a top fashion model)
- *There are several* extreme behaviors of music fans these days.
- (Prompt: Extreme behaviors of music/sport/movie fans)
- In my opinion, *there are several* sources of student stress.
- (Prompt: Sources of stress in students)
- *There are several* things I used to collect as a child.
- (Prompt: Things you or someone I know used to collect as a child)

As discussed in the previous section with the core bundle *there are three*, the use of the phrase “there are” is one of the frequently used method to write a topic sentence because there is an example of such use from the writing paragraph with *there are many* in
the sentence “at [university name] there are many useful majors that students can choose from.”

However, there are also some instances where the bundle there are several is not used as part of the topic sentence. Two instances of there are several are used as part of the conclusion (e.g. in conclusion, I think there are several benefits…) and general descriptions in the supporting points (e.g. inside the box, there are several colorful of pens and pencils…).

In Composition, there are 8 written bundles out of 12 that are only unique to this group:

- intangible framing attributes (1 bundle): to the fact that
- quantity specification (2 bundles): there are no, are a number of
- identification and focus (5 bundles): can be seen, does not have, this means that, that it is not, which can be

Most of the bundles from the list above are used across different topics and writing tasks. However, the use of are a number of seems to be specific to one particular genre and pattern. In 9 out of 10 occurrences, are a number of is used in essay analysis in the pattern “there + are a number of + weaknesses/strengths.” As discussed in the previous section with the core bundle a number of, this is a direct influence from the model analysis that were given to the students as a guideline on how to write an analysis. The sentence in the model analysis writing is “Unfortunately, there are a number of weaknesses in Duangkamol’s argument.”

Three uses of a number of are the same as that in the model writing:

- Unfortunately, there are a number of weaknesses in Monthita in terms of pathos.
- Unfortunately, there are a number of weaknesses in Monthita’s defense of…
- Unfortunately, there are a number of weaknesses lying in her ethos.

Other uses of this bundles are still very similar to the model:
- Throughout the essay, there are a number of strengths to Monthita’s argument.
- However, despite these flaws, there are a number of strengths in Monthita’s defense…
- There are a number of strengths in Montira’s arguments.

As for shared written referential expression bundles, there are four such bundles: Identification and focus (they do not) and contrast and comparison (on the other, on the other hand, the other hand).

All bundles are used in writings for several different writing prompts and genres. However, the three contrast and comparison bundles contain “the other” core phrase.

In Freshman English, except for three instances, the use of on the other is followed by “hand.” The three exceptions are:

- … if you make any little mistake on the runway, on the other day you will be talked about and be embarrassing.
- On the other side also have many warm tone painted houses.
- On the other side are also mountains of paper which I don’t even know which one should I read first.

Similarly, with “the other hand,” most of the occurrence is in the bundle on the other hand with the exception of 5 cases (out of 25) where the other hand is incorrectly used with the preposition “in”, when students meant to use on the other hand, such as:
- … in Bangkok, you rarely find a place for your car. In the other hand, bicycles don’t need too much space for parking…

- …if your bicycle crashes tress, the only thing you have to do is pick it up and continue riding. In the other hand, if you have an accident as you’re driving a car…

This wrong use of preposition can be explained by the influence of the students’ first language. In Thai, the expression whose meaning and usage is similar to “on the other hand” can be directly translated into English as “in the other way.” Therefore, the students might be using the preposition “in” instead of “on” because of this.

In Composition, it is similar to the Freshman English where almost all use of the three-word on the other is part of the four-word bundle on the other hand (33 out of 35), with only two exceptions, i.e., on the other side and on the other way around.

In 33 occurrences of the other hand, all but on case is part of the longer bundle on the other hand. The exception is only one case of incorrect proposition use with “in” instead of “on”

- In the other hand, Mrs. Spence treats the boy with warmness and case.

### 5.2.2 Stance expressions (category B)

Table 5.2: Stance expressions

<table>
<thead>
<tr>
<th></th>
<th>AFL</th>
<th>FE</th>
<th>FE 40</th>
<th>Compo</th>
<th>Compo 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core AFL written &amp; spoken</td>
<td>17</td>
<td>14</td>
<td>1</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>Primarily spoken</td>
<td>79</td>
<td>44</td>
<td>14</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>Primarily written</td>
<td>81</td>
<td>36</td>
<td>4</td>
<td>62</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>177</td>
<td>94</td>
<td>19</td>
<td>106</td>
<td>37</td>
</tr>
</tbody>
</table>
In the stance expression category, the AFL bundles are primarily spoken (44.6%) or written (45.8%) ones, with only 17 core bundles (9.6%). Looking at all occurrences of target bundles in both subcorpora, the proportions in both are also similar to those in the target list with core bundles accounting for the smallest percentage (15% for Freshman English and 13% for Composition). The difference is that there are more spoken bundles than written ones in Freshman English. In contrast, for Composition, there are twice as many written than spoken bundles.

The difference is more marked when narrowing down the results to frequent bundles (40 per million words). In Freshman English, almost all target stance bundles (14 out of 19 bundles or 73.7%) are spoken, with just one bundle that is core and four that are written. The Composition subcorpus, however, contains more written target stance bundles (20 bundles or 54.1%) than core and spoken bundles combined (9 bundles or 24.3% and 8 bundles or 22.2%), with only a difference of one bundle between the last two, making the proportion different from the target bundle list, where spoken and written bundles make up the majority of the stance bundles:

In short, for this functional category, the majority of the target AFL bundles used are spoken bundles in the Freshman English subcorpus but written bundles in the Composition subcorpus.

5.2.2.1 Core stance expression bundles

In Freshman English, there is only one core stance bundle (obligation and directive), make sure that, and it is shared with Composition. In other words, there is no unique core stance bundles in this subcorpus.
In Composition, however, 8 out of the 9 core stance bundles are unique to the group. Those bundles are hedges (may not be, likely to be, it may be), epistemic stance (according to the, to show that, out that the), expressions of ability and possibility (to use the) and evaluation (the importance of). Most of these bundles are used across different prompts and genres, with no particular pattern of use, with the exception of out that the and to use the. The former is used with two verbs “point” and “turn,” as shown in the following examples:

- …the writer points out that the argument to stop eating…
- At the same time, some people point out that the real threat is…
- …by pointing out that the true problem is homosexual discrimination…
- …but it turns out that the author has delivered the message…
- If the result turns out that the school teachers give too much homework…

As for the ability and possibility stance bundle, to use the, 19 out of 31 instances are for the essay prompt on online homework services, most commonly followed by variations of reference to the online homework services:

…it is against morality to use the homework services…

… there might be a lot of factors leading students to use the online homework services.

People claim that the students have to use the online homework services because…

… why some students decide to use the services rather than trying to do homework…

… the decision to use the service corrupts student’s personality
The sole shared bundle, *make sure that*, is used in several writing prompts in both subcorpora, primarily in example and reason writing in Freshman English and only in essays in Composition. In Freshman English, almost half of the occurrences (8 out of 18) of *make sure that* are followed by second person singular “you/your” and two are with “they,” as illustrated in the examples below:

- Also, *make sure that you* do not eat anything…
- You have to *make sure that your* habit is healthy…
- … and *make sure that you* wear it correctly
- They just want to *make sure that they* will see…

In contrast, in Composition the majority of occurrences of this bundle is followed by “they/their”, with none followed by “you/your”:

- … it will help them *make sure that they* gain a deeper understanding…
- … Thai parents needed to *make sure that their* future son-in-law has enough money…
- …spare much money to *make sure that their* children can take the exam…

**5.2.2.2 Spoken stance expression bundles**

In Freshman English, the majority of the spoken stance bundles are only unique to this group (10 out of 14 bundles), which are:

- hedges (1 bundle): *it looks like*
- epistemic stance (2 bundles): *and you can, as you can see*
- obligation and directive (2 bundles): *you want to, you need to*
expressions of ability and possibility (3 bundles): *you can see, that you can, so you can*

- intention/volition, prediction (2 bundles): *if you want, if you want to*

Most bundles in this category are used across various topics and genres but there are still some that have specific patterns of use. For example, the bundle *it looks like* is primarily used in description writing. Of the 15 instances, only one is used in example writing. In description writing, this bundle is used to describe physical appearance of things:

- *It looks like* my head just hang on my neck but can’t be straight by itself.
- …hair doesn’t look like silk, but *it looks like* a brown nest.
- Now *it looks like* pieces of broken mirror lay messily on my desk…

Another example is the epistemic stance bundle, *as you can see*, which is primarily used to lead into conclusion:

- Now, *as you can see*, reading test is hard and challenging for many students by many reasons.
- *As you can see*, I used to collect unuseful things but I’m really proud of my collections!
- *As you can see*, being a top fashion model is fabulous.
- *As you can see*, too much bad things will come to your life….

In composition, half of the spoken stance bundles (4 out of 8) are unique: hedges (*it might be, it could be*), and obligation and directive (*we need to, to make sure*). The four bundles are used in various prompts and writing tasks, with no distinctive pattern of use.
Similarly, the 4 bundles that are shared between the Freshman English and Composition subcorpora are also used in various topics and writing tasks. Those bundles are hedges (a kind of), obligation and directive (we have to, make sure that), and expressions of ability and possibility (be able to).

5.2.2.3 Written stance expression bundles

In the category of stance bundles, there are only 4 such bundles in the Freshman English subcorpus, which are all shared with the Composition group.

In Composition, there are 16 unique bundles, a stark contrast to the Freshman English group. These unique bundles also span across several stance expression subcategories (5 out of 6) except for intention/volition and prediction function and they are also used across different writing prompts and tasks:

- hedges (1 bundle): are likely to
- epistemic stance (3 bundles): be seen as, be considered as, can be considered
- obligation and directive (2 bundles): should not be, needs to be
- expressions of ability and possibility (5 bundles): it is possible, it is possible that, are able to, can be found, can also be
- evaluation (5 bundles): important role in, it is necessary, it is clear, it is clear that, it is impossible

The bundles be considered as (24 instances) and can be considered (19 instances) are sometimes also combined into a larger bundle can be considered as (8 instances). Notably, for the bundle be considered as, 10 out of its 24 uses are preceded by “should not” or “should”: 
- Grades **should not** *be considered as* a criterion for determining a student’s success…
- …the non-violent drug abusers **should not** *be considered as* a criminal…
- This kind of behavior **should** *be considered as* cheating because…

The 4 shared bundles between the two subcorpora are from two stance subcategories epistemic stance (*if they are*) and evaluation (*it is important, it is important to, the most important*).

In Freshman English, 37 out of 49 uses of *it is important* and out of 43 uses of *it is important to* are as part of a larger bundle *it is important to study Thai literature*, which is in turn a large part of the writing prompt for a reason paragraph in the final exam. The writing prompt given to the student is “why it is/isn’t important to study Thai literature” and the examples below are how the students worked the prompt into their writing, primarily as or as part of the topic sentence:

- *It is important to study Thai literature.*
- There are many reasons why *it is important to study Thai literature.*
- There are three reasons why *it is important to study Thai literature.*
- Many people have several reasons to tell that *it is important to study Thai literature.*

In many cases, students either shorten the prompt to make it a complete sentence or add phrases to signal the topic sentence and the paragraph genres to the prompt, as in the examples above from the subcorpus.
5.2.3 Discourse organizing functions (category C)

Table 5.3: Discourse organizing functions

<table>
<thead>
<tr>
<th></th>
<th>AFL</th>
<th>FE</th>
<th>FE 40</th>
<th>Compo</th>
<th>Compo 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core AFL written &amp; spoken</td>
<td>25</td>
<td>17</td>
<td>10</td>
<td>23</td>
<td>12</td>
</tr>
<tr>
<td>Primarily spoken</td>
<td>82</td>
<td>34</td>
<td>10</td>
<td>23</td>
<td>5</td>
</tr>
<tr>
<td>Primarily written</td>
<td>32</td>
<td>8</td>
<td>-</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>139</td>
<td>59</td>
<td>20</td>
<td>58</td>
<td>18</td>
</tr>
</tbody>
</table>

This last functional category has the least number of bundles when compared to the other two functions. Spoken bundles have the largest proportion within this category (82 bundles out of 139 or 59%), and there are slightly more written bundles (23%) than core bundles (18%).

For all occurrences of the target bundles in the learner subcorpora, the Freshman English is similar to the AFL in that it has a larger proportion of spoken bundles (34 out of 59 bundles or 57.6%); however, it differs from the AFL in that its proportion of core bundles (28.8%) more than doubles that of written bundles (13.6%). In Composition, about 80% of the discourse organizing bundles (46 out of 58 bundles) are equally divided between core and spoken bundles (23 bundles or 39.6% for each), with only 12 bundles that are written (20.8%).

Similar to the previous functional category, the difference in proportion is also markedly different when narrowing the results to only frequent bundles. In the Freshman English subcorpus, there are no written discourse organizing bundles. Frequent discourse organizers in the Freshman English subcorpus are half core bundles and half written bundles (10 bundles in each).
With Composition, core bundles (the smallest category in AFL) is the largest and the majority of the bundles (12 bundles or 66.7%) with only 5 spoken and one written discourse organizing bundles. Both subcorpora are similar in the lack of written discourse organizers, but Composition has over twice as many core bundles than spoken ones.

5.2.3.1 Core discourse organizing bundles

In Freshman English, there is only one unique bundle among the 10 core discourse organizing bundles, i.e., the topic introduction and focus bundle for example if. This bundle is used in both reason and example paragraphs. Such sage is expected since in the writing packet given to the students, students are encouraged to use the transitional expression “for example” to signal a supporting example in example paragraph writing and to support their reasons with examples in their reason paragraph writing.

- I have interesting ways to celebrate a friend’s birthday. For example, if my friend’s birthday near come, …(example)

- Firstly, students can read Thai literature by themselves. For example, if they want to know about literature… (reason)

Among the 12 core discourse organizing bundles in the Composition subcorpus, there are three bundles that are unique to this group: cause and effect topic elaborations (the result of, so that the) and discourse marker (in other words). They are used in various prompts and tasks.

When it comes to shared bundles, there are more shared than unique ones in this functional category: 9 core discourse organizing bundles are shared:

- topic introduction and focus (2 bundles): for example the, for example in
- topic elaboration (cause and effect) (4 bundles): *due to the, as a result, in order to, because it is*

- discourse markers (3 bundles): *as well as, at the same, at the same time*

Most of these bundles are used in several different writing prompts and genres in both corpora, with a few cases where they are specific to some genres. As previously discussed in Freshman English where the unique bundle *for example if* is used in both reason and example paragraph writings, the two shared bundles, *for example the* and *for example in*, are also used only in these two paragraph types but not in the descriptive paragraph type.

In composition, 45 of the 60 occurrences of the bundle *for example in* are followed by paragraph numbers and only with essay analysis writings:

- *For example, in* paragraph 3, the author starts her rebuttal by…
- *For example in* par 3. she refuse the opinion…
- *For example, in* paragraph 1,3 and 5 where she asks herself if
- *For example, in* the first paragraph, she questions…

### 5.2.3.2 Spoken discourse organizing bundles

In the subcategory of spoken discourse organizing bundles, the Freshman English composition has more unique bundles than the composition and the shared one. 7 out of 10 spoken stance bundles are only used by the Freshman English group, with only 3 shared with the Composition. These unique bundles are

- topic introduction and focus (4 bundles): *if you have, if you have a, to look at, you have a*
- topic elaboration (cause and effect) (1 bundle): so if you
- discourse markers (2 bundles): and if you, but if you

There are certain overlaps among the bundles. The overlaps between the first two (if you have and if you have a) are clear: in 15 out of 38 occurrences if you have is part of a longer bundle if you have a. With the bundle you have a, in about half of its occurrences (15 out of 29) it is also part of the four-word bundle if you have a. Only in 3 instances (out of 28) so if you is part of the longer bundle so if you have and only in 1 instance it is part of a five-word bundle so if you have a. The bundle and if you is not used with the verb “have” at all, while for but if you, it is part of but if you have in 3 out of 32 cases, and in 2 of those it is part of a longer bundle but if you have a.

These spoken discourse organizing bundles are used across different topics and in different writing genres.

Similarly, the only 2 bundles (out of 5) in the Composition subcorpus that are unique to this group are also used in different prompts and genres. Both bundles are topic elaboration bundles: non-causal (come up with) and cause and effect topic (in order to get).

There are only three shared bundles in this category: introduction and focus bundles (first of all, look at the), and cause-and-effect topic elaboration bundles (the reason why).

In Freshman English, the bundle first of all is very has a much higher frequency (467 occurrences in 467 different paragraphs, out of the total 1,805 writings) than in Composition (52 occurrences in 49 writings, out of the total 242 texts). Such high frequency in Freshman English is expected because this bundle is highlighted in the writing packet on the use of transitional expressions in paragraph writing.
In terms of pattern of use, the bundle *look at the* is primarily used in description paragraphs (16 out of 19 uses) in Freshman English and it is mainly used to introduce concrete or physical things as part of the description:

- And when you *look at the* middle of the desk,…
- If you *look at the* entrance, there a people coming in…

In Composition, *look at the* is used in different writing topics and genres, and can be used with both physical objects or the topic to be discussed:

- When we *look at the* photos and read the captions,… (object)
- First, let’s *look at the* definition of plagiarism. (topic)

Another shared bundle *the reason why* also has different patterns of use between the two groups. With Freshman English, the bundle is primarily used in reason paragraph writing. However, several of the usages are mostly incorrect, such as when they are used as part of the conclusion:

- In conclusion, that’s all *the reason why* I choose to major in Thai and I hope I really enjoy it so much :)
- That is all *the reason why* I think English One should be a required course for all Arts students.
- So all above this is *the reason why* Somtum is one dish every tourist must try in Thailand.

In most cases, the correct usage is supposed to contain the plural form “reasons” since the paragraph would list several supporting reasons (usually three reasons) for the writing prompt. This incorrect use of singular and plural could be from the fact that in their first language, Thai, plural forms are not marked in the same way as English does.
There are also cases where the use of *the reason why* is ungrammatical.

- *The reason why* atychiphobia is a good fear to have.
- *The reason why* wearing CU uniform.

These two examples are what the students used as the topic sentences in their paragraph. Structurally, these are noun phrases starting with the bundle *the reason why* and not a complete sentence.

Such ungrammatical use with the bundle *the reason why* is not present in the Composition subcorpus. All the 20 occurrences of *the reason why* are grammatical:

- Thus, *the reason why* PunPun still get criticism is because media still have no limit…
- *The reason why* many students focus so much on grades is because they would like to…
- Mr. Jakkriz also overlooks *the reason why* these students turn to the online homework service

5.2.3.3 Written discourse organizing bundles

There are no written discourse organizing bundles in the Freshman English subcorpus. The only one such bundle in Composition is *even though the* and it is used across different topics and writing tasks.
5.3 Discussion

From the results presented in the previous sections, the two research questions that are addressed in this section are:

Research questions 3: Which lexical bundles from the Academic Formula List (AFL) are used frequently by the students in both subcorpora?

Research questions 4: How are the two groups of students similar or different in their use of lexical bundles from the Academic Formula List (AFL)?

The third research question of this dissertation is addressed in detail in the earlier sections in this chapter. To summarize, out of the 607 bundles in the AFL, 70 are frequently used by the Freshman English group and 126 are used by the Composition group. Dividing this further, the students in the Freshman English subcorpus use 28 core bundles, 32 spoken bundles and only 10 written bundles. The composition group uses 69 core bundles, 24 spoken bundles and 33 written bundles. The use of the lexical bundles by the two groups are quite distinct.

This brings us to the fourth research question on how the two groups of students are similar or different in their use of lexical bundles from the Academic Formula List (AFL). From the results presented in the previous sections, focusing only on frequent bundles that are at the frequency of 40 per million words and over, we can see that overall the expert Composition group makes use of more bundles from the AFL than the beginner academic writers in the Freshman English group. This shows that the students in the faculty are progressing towards greater use of academic lexical bundles in the AFL list as they become more advanced in terms of their writing.
Dividing the target AFL bundle use in both subcorpora even further into core, spoken and written bundles, the difference between the two groups is that Freshman English group uses spoken bundles the most (54.7%) and written bundles the least (only 14.3%). In contrast, in the composition subcorpus, 54.8% of the target AFL bundles are core, with only 19% that are spoken bundles. High usage of spoken bundles in Freshman English can be explained by the fact that this is the writing class in the university. From my own teaching experience, some students had never written a paragraph prior to taking the course. In order to transition the students into writing comfortably, several of the prompts given to the students draw on their personal experience as supporting points, such as things they collect, their hobbies, people they know, school or university experience, or classroom description. This invites more conversational, spoken language into the writings themselves. In contrast, with Composition students, the essay writing topics are less personal and more on social issues at the time. Several of the prompts are based on prior readings of news articles. Students in this course have also finished the two Freshman English courses with the paragraph writing components and the Composition I course; therefore, students are aware of how to write an argumentative essay and of audiences of their writing.

A close look at target AFL bundles in both learner subcorpora reveals several issues with learners’ writings.

The first issue is the problem of using bundles that are not originally produced. This refers to lexical bundles that are part of the writing prompts or model writings. The use of non-originally-produced bundles are not problematic in itself especially since in some cases this cannot be avoided. An example of this is the bundle *it is important to* whose
occurrence is primarily in paragraphs written for the reason paragraph prompt “why it is/isn’t important to study Thai literature.” In most cases, the bundle *it is important to* is used as a chunk with *study Thai literature*, forming a seven-word bundle. Students used the writing prompt in their own writing because students are not asked to narrow the prompt down further for their writing. The prompts are also mostly readily usable as topic sentences with minimum change (omitting or adding some words), such as in the example given earlier, and at this level or writing, instructors do not put an emphasis on paraphrasing or not copying the prompts. Students are more likely to be penalized and have points taken off of their writing for not having a proper topic sentence than for copying the prompt and using it as one. Adhering to the prompts and using them in their paragraph is a safe choice in writing.

In the case of Composition, the problem of using non-originally-produced bundles is different from in Freshman English. This can be seen from the context of use for *a number of* and *are a number of*. Looking at how these bundles are used in students’ writings reveal that in essay analysis writings, when these bundles are used, in several cases the students also copy the whole sentence from the essay analysis model, not just the bundles alone. The following examples show similarities between the essay analysis model and the students’ analysis writings:

- Essay Analysis Model: Unfortunately, there are *a number of weaknesses* in Duangkamol’s argument in favor of a single title for both married and unmarried Thai women.
- Students’ Analysis: Unfortunately, there are a number of weaknesses in Monthitas defense of the removal of stalls…
- Essay Analysis Model: However, despite these considerable strengths, the essay does have a number of weaknesses that undermine the overall effectiveness of the Angkana’s argument.

- Students’ Analysis: However, despite these considerable strengths, the essay does have a number of weaknesses that undermine the overall effectiveness of Rojtinnakorn’s argument.

Such uses of sentences from the model analysis writing by the students is rather surprising since in the Composition II course and previously in Composition I, the courses covered a topic of plagiarism and paraphrasing. However, as explained earlier, this writing assignment is new to the students and they are only given two similar analysis models, written by only one instructor, as a guideline for writing. Much like in the Freshman English group, adhering to the phrases and sentences given from the instructors are possibly seen as a safer, in addition to easier, choice in a graded assignment. In my own class, I have also noticed my own students lifting phrases and sentences from the analysis models and using them as a whole or with slight modifications in their own writing. However, from the corpus results, this issue is also prevalent in other sections of the same course too. From discussing this with my students in class, the responses from the students who relied heavily on the model analysis in their writing are that they do not know how to write an analysis so they cling on to the models.

What these issues of using non-originally produced has shown us as instructors is the need for extra materials. In the case of Freshman English, an additional section on how to incorporate or paraphrase the prompts into their own topic sentence and conclusion can be beneficial to the students since the current writing packet does not cover writing topic
sentences and conclusions at all, except from a short paragraph that says a paragraph must have a topic sentence and a concluding sentence. With Composition, it is clear that instructors need to revise the materials given to the students since giving them two similar model analysis writings ends up with students using sentences from the models as frames for their writing. There is a tension between writing in the same way as the model analyses are written and not plagiarizing from the model.

The second major issue that is visible from studying the target AFL bundles in the subcorpora is ungrammaticality. This is a problem in Freshman English as seen in the use of the two bundles *the other hand* and *the reason why*.

With the bundle *the other hand*, there are 5 instances out of 25 where the bundle is used with the preposition “in,” rather than with “on” as a longer bundle *on the other hand*, which is the other 20 occurrences. This is an influence of Thai, their first language, where a similar expression with the same meaning and function, is used with the preposition “in.” In Composition, there is only one instance of *in the other hand*.

The other bundle *the reason why* is primarily used in topic sentences and conclusions. However, it is used incorrectly since in those sentences, the “reason” actually refers to multiple reasons that would be discussed (introduction) or have been discussed (conclusion) in the paragraphs. Freshman English students also used *the reason why* as noun phrases starting with the bundle, rather than a sentence, as a topic sentence. In these cases, sometimes students just add “the reason” in front of the writing prompt they chose to write for, such as “*the reason why* atychiphobia is a good fear to have” for the prompt “why atychiphobia is a good/bad fear to have.”
This issue of ungrammatical use of bundles, especially in Freshman English writing, calls for a change in course materials for the students. The ungrammatical uses occur primarily with topic sentence and conclusion, the areas which are not covered in the writing packet, except for, as mentioned earlier, saying that a paragraph needs a topic sentence and a conclusion. It is clear from these instances of use from the corpus that students have problems with writing a topic sentence and a conclusion, from a minor issue of using singular-plural forms of words to a major one of writing a grammatically correct sentence. It can be argued that the instructors for the course can easily bring this to the students’ attention when correcting or grading the writing with these errors on. However, with 14 different sections of students in the course and 14 different instructors, this means 14 different teaching styles, especially in terms of allocating class hour to teaching about writing and commenting on students’ work. Some instructors discuss paragraph writings in class with students while some assign students to read the writing packet on their own and write. Since it is clear in this chapter that writing a topic sentence and a conclusion can be problematic for some students, it is beneficial for future students to have a section where they are explicitly guided on how to write ones.

In conclusion, the analysis of the target AFL bundles in the learner subcorpora shows how different the Freshman English and Composition groups are, as well as highlighting some issues that can be addressed in further class material.
CHAPTER 6

Conclusion

This chapter concludes the dissertation with a summary of the major findings and a discussion of the pedagogical implications of the findings, the limitations of the dissertation research, and future directions of corpus-based research on Thai learners of English.

6.1 Summary of findings

The analysis of the lexical bundle use by Thai learners of English in this dissertation has shown how the novice and expert groups of undergraduate students used lexical bundles in their writings and how they also used the lexical bundles in the Academic Formula List (AFL) that are deemed important in academic writing.

The first part of the analysis shows that the students in the Freshman English subcorpus, the novice group, used more lexical bundle types and tokens than the expert students in the Composition subcorpus. The difference in the lexical bundle use is partly explained by the frequent use of non-original lexical bundles that are part of the writing prompts by the Freshman English students. Structurally, the majority of the bundles that both groups use are phrasal bundles. However, when looking at the subcategories of each main structure, differences between the two group can be observed. Similarly, in terms of function, the majority of the bundles are referential expressions in both groups. The differences are in the functional subcategories, such as the more frequent use of
obligations/directive stance bundles and the Composition in the Freshman English subcorpus and the more frequent use of intention/prediction and ability stance bundles in the Composition subcorpus. The novice Freshman English group also uses almost twice as many non-originally produced lexical bundles as their expert counterpart. Most of the non-original bundles are phrases from the writing prompts.

The second part of the analysis with target bundles from the Academic Formula List (AFL) shows that the novice Freshman English group used fewer AFL bundles than the expert Composition group. The two group are also different in the register of the AFL bundles: the Freshman English group used more spoken bundles than written ones and the Composition group used more written than spoken bundles. The result of the AFL target bundle analysis reveals that the expert group, the Composition students, are already using bundles that are core (both written and bundles) and written academic lexical bundles in their academic writing whereas the novice group are still using spoken bundles in their academic writing.

6.2 Pedagogical implications

The pedagogical implications from this research is primarily on how to adjust the learning materials in both writing courses to address the issues that were identified in the analysis, such as ungrammatical uses of some lexical bundles, excessive use of bundles that are part of the writing prompt, and use of phrases or clauses of the model writing that are bordering on plagiarism.
From the first part of the analysis, with the Freshman English course, certain sections can be added in the writing packet as well as the classroom discussion. Since one of the issues identified in the analysis is overusing lexical bundles that are part of or that are the writing prompts, what can be done is adding sections on how to write a topic sentence from the writing prompts. The current class material on writing only focuses on how to narrow down the writing topic in order to make the topic focused enough as a paragraph. However, the reality is that in assignments and tests students are given writing prompts that are very specific and mostly are usable as a topic sentence. As such, students should be trained on how to paraphrase the writing prompts so that they do not end up repeating the same chunks of words throughout the paragraph. Since this freshman course is divided into 14 sections and taught by 14 different instructors with different teaching methods and styles, making changes to the core class material for writing is one way to ensure that this particular area is addressed in all sections and for all students.

In Composition, the main issue that is identified in the analysis is cases where students incorporated phrases or sentences from the model writing. While the result of the lexical bundle analysis shows that students frequently used some 4-word bundles that are from the writing samples in the class material in their own writing, looking at a broader context of use of those 4-word bundles revealed that the students actually incorporated more than just 4-word bundles. In certain cases, the whole sentences are lifted, with minor changes of some words, from the source materials. Since this is a case of plagiarism, it is a serious issue that needs to be addressed. Although in all courses the department explicitly include a section on plagiarism in the course syllabus and disciplinary actions that entails
cases of plagiarism by students, the findings from this study shows that, even in an advanced writing course where students should be familiar with and careful about not plagiarizing, some students still plagiarized from the class materials in one particular writing task that they were not familiar with. Future composition courses should also include sessions on plagiarism and dependence on model essays and writings. In this particular course, more discussion should be dedicated to the essay analysis task since the cause of the problem where students used a large portion of the model writing in their own work is that the task is relatively new to the students and the students might be unsure about how to write such essay analysis assignments. There need to be a balance between providing samples of how to write and encouraging students to not depend on the writing samples.

In the second part of the analysis on Academic Formula List (AFL) lexical bundles, the results of AFL bundle use in the learner corpus can also contribute to future writing course planning. The analysis shows that the expert students are already using primarily core bundles and more written AFL bundles than spoken ones while the beginner group, the Freshman English, uses more spoke bundles from the list than core and written bundles. The instructors for these courses can focus on integrating more core and written AFL lexical bundles into the writing courses, especially with Freshman English courses. More research is needed on this part to decide on which AFL bundles to be included and how to do so.
In a broader context of teaching, it is also important to highlight the concept of lexical bundles and their uses to students along with other formulaic language such as collocations and idioms.

6.3 Limitations

This study is not without limitations. The first limitation lies in the data collection process. The majority of the data collected, especially tests and final examination writings, are handwritten. It was time-consuming to convert the handwritten paragraphs and essays into digital .txt files and checking to ensure the accuracy of the digitization. While tests and final examinations will most likely remain handwritten, for future data collection and corpus compilation projects, writing practices and some graded assignments can be digital submission.

Another limitation is the opportunity to interview the students about their choices to use certain bundles. While it was expected that students would incorporate keywords from the writing prompts, what was surprising and not expected was the repeated use of bundles that are part of the writing prompts as a whole regarded of the parts of speech of the bundles and the use of bundles from the model writing that extended beyond the four-word boundary to, in several cases, the whole longer sentences. Since those issues were not expected, the research methodology was not planned for the interview component, in addition to the data collection of the students’ writings. This component can also be included in future studies because I plan to expand the learner corpus that would allow the
possibility of tracking students over the period of four years where they take the
department’s writing courses. This will be discussed in the following section.

6.4 Future directions

This dissertation is a starting point in further corpus-based research into Thai
learners of English and English language writing, with several future directions.

The learner corpus can be further expanded to include writings from other courses
that are offered by the department. In this study, the corpus data was collected from two
courses in one semester. The learner corpus can be further expanded to include written
assignments and tests from all the writing courses that the department offers and even from
discipline specific courses such as literature and linguistic courses where students are
usually assigned to write term papers as part of the assessment. In addition, if the corpus
data is collected in such manner from several courses over several academic year, the
corpus can also be expanded in terms of time to include students’ writing from their first
writing course to their last. This would allow for a longitudinal study where researchers
can follow certain students’ progress over time from first to final year. Since the corpus
data are naturally-occurring and collected from the writings that the students would
normally write in their assignments or test, not writings from tasks that are specific to just
a certain study, the learner corpus that is compiled and used for analysis in this dissertation
can also be used in future research on writings of Thai learners of English.
Bibliography


VITA
Mathurin Leelasetakul

EDUCATION

Masters of Arts in English as an International Language (EIL) at Chulalongkorn University, Bangkok, Thailand.

Bachelor of Arts in English (First Class Honors, Gold Medal) at Chulalongkorn University, Bangkok, Thailand.

PUBLICATION