3-7-2017

Lexical Bundles in Applied Linguistics and Literature Writing: A Comparison of Intermediate English Learners and Professionals

Kathryn Marie Johnston
Portland State University

Let us know how access to this document benefits you.

Recommended Citation

Johnston, Kathryn Marie, "Lexical Bundles in Applied Linguistics and Literature Writing: A Comparison of Intermediate English Learners and Professionals" (2017). *Dissertations and Theses.* Paper 3482.

Lexical Bundles in Applied Linguistics and Literature Writing:

A Comparison of Intermediate English Learners and Professionals

by

Kathryn Marie Johnston

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Arts
in
Teaching English to Speakers of Other Languages

Thesis Committee:
Susan Conrad, Chair
Alissa Hartig
Julia MacRae

Portland State University
2017

Abstract

Lexical bundles (fixed sequences of three to four words) have been described as building blocks of discourse, both written and spoken (Biber & Barbieri, 2007), and as a useful mechanical device for creating writing that is suited for its academic field (Hyland, 2008). Having noticed that the academic theses of my students at Longdong University in Qingyang, China seemed very different from professional writing in their fields, I created a thesis project that addressed the question of how professionals in their fields were using bundles and how the learners' use of these bundles in terms of frequency, structure, and function varied from the professionals' use.

In order to answer this question, I compiled four corpora of writing in literature and applied linguistics, representing professional and learner writing in each field. I used concordancing software in order to identify four-word lexical bundles that occurred at least 20 times per 100,000 words and over a range of four texts. I then did a three-part analysis which looked at frequency, structure, and function of these bundles.

The results of the study reveal that professionals in applied linguistics and literature use bundles with different frequency, display different choices of lexical items to fill structural bundles, and use functional bundles differently. These differences seem to reflect the rhetorical needs of each discipline. Further, the learners in each field displayed differences in their use of bundles as compared to the professionals' use. Learners in applied linguistics used more types and tokens of bundles overall, while learners in literature used fewer. Both groups of learners relied more on repetitive use of certain bundles than did the professionals.

Implications of this study are discussed for teaching and curriculum development. The findings can be applied to teaching through creating awareness-raising and guided practice opportunities for the students to see how bundles are used in professional writing and to help them apply this understanding to their own writing.

Table of Contents

List of Tables

List of Figures

Chapter 1 - Introduction

While working as a university English instructor with the Peace Corps in China, I started hearing complaints from the senior students about having to write their graduation theses in English rather than in Chinese as they had in previous years. Many felt like they were not adequately prepared to handle this task and found it to be difficult and frustrating, which made me wonder what their final theses were like and how they compared to writing by more proficient English users. I became interested in creating a corpus comprised of writing samples from the students at my school, which was a low-ranked four-year college located in a remote area in Gansu province, one of the poorest provinces in China. Many of my students scored poorly on the college entrance exam, and they were often not only the first in their family to attend college, but in many cases also the first to complete high school or even middle school. With a comparison of a corpus of my students' writing and the target they were aiming for, I could provide the university with information about the writing of English major students, in terms of both what they were doing and what gaps existed between the writing produced and writing on similar topics prepared by proficient English speakers.

Of course, many features are important for creating effective writing, but I became particularly interested in formulaic language while taking a class in corpus linguistics. I was interested in how prevalent formulaic language is in both spoken and written discourse, and how competent use of it can help a language learner become more proficient in spoken and written registers. As I thought of the Chinese students' papers, I

was curious about what formulaic language they had learned to use and how they used it.

Lexical bundles, which are recurrent sequences of three or more words, are useful in academic writing since academic writing is, in part, characterized by its formal structural requirements. These structural expectations can vary according to academic discipline (Wood, 2015). Proficient use of lexical bundles can help learners' writing become more logical and coherent (Hyland, 2008). Having a description of the variation of bundles between disciplines can then guide teachers in creating targeted lessons that may help their students become more proficient writers in their field.

Previous research on disciplinary variation of lexical bundles has found that they differ between disciplines in terms of overall frequency, structural type, and functional use. For example, Hyland (2008) found that writing in electrical engineering used more distinct types of bundles than did writing in business studies, applied linguistics, or biology. Further, of the specific bundles used by the engineers, many were not found in the writing of other fields. On the other hand, biology writing used fewer bundles than any of the other fields, which Hyland states may be related to how arguments in that field are situated and framed, as compared to engineering. He speculates that engineering writing may be more abstract while biology is more focused on being concrete.

When looking at the structural and functional variation between these fields, Hyland also found that writers in biology and engineering more frequently used bundles with a passive structure, while biology and applied linguistics used higher proportions of bundles including a *prepositional phrase + of*. Similar differences arise in the distribution of functional bundles between all groups. This research is useful for helping provide a

basic understanding of which lexical bundles are discipline-specific and which are shared more broadly.

Other research has focused on using these descriptions of professional writing in different disciplines and comparing that with the frequency, structure, and function of lexical bundles as they occur in learner writing. For example, Cortes (2004) compared the writing of learners and professionals in history and biology. She found that learners often did not use the same bundles that professionals did, and when they did use the same bundles the ways in which learners used them were different from the professionals.

A better understanding how lexical bundles are used by professionals in both applied linguistics and literature writing can help to better inform the writing curriculum at Longdong University. It is also important to understand how the learners are currently using bundles, and if they are using shared bundles or structures, how the learners' use compares to the professionals'. To do this, it is necessary to examine how bundles vary in use according to frequency, structure, and function in each discipline, and compare how the learners are currently using these target bundles.

In order to discover how student writing at Longdong University compared to learner writing in previous research, I developed a thesis project that examines the use of lexical bundles in my students' academic writing compared to the use in professional authors' writing.

Overview of the thesis

The following chapters will introduce the literature relevant to the current study, and at the end of the second chapter, I present my research question. Chapter 3 introduces

the context for the study, reviews the corpus compilation procedures, and describes the

frameworks that were used for analyzing the data. Chapter 4 presents the results and

discussion of my analysis, which answers the research question in three parts: the first

concerns bundle frequency differences between literature and applied linguistics writing

at the professional level, and how the learners in each group differ from the experts; the

second part compares structural types of bundles found in each discipline, and the

variation between learners and professionals within each area; the third part addresses

how professionals in each area use the bundles for functional purposes, and how the

learners in each area vary. Chapter 5 starts with a summary of the results of the analysis,

and then I review the implications these results have for teaching and include a sample

lesson that can be used in the classroom. I also discuss the limitations of the present study

and conclude the thesis with suggestions for future research.

Chapter 2 - Literature Review

2.1 Introduction

In this chapter, I review the literature relevant for my study. The chapter opens by defining *formulaic language* and reviewing studies which explain how formulaic language is an important part of our competence in a language, whether first or additional. I then turn to focusing on the advantages of formulaic language study for second language learners, reviewing the studies which have found that formulaic language offers processing advantages, increases confidence, and contributes to more natural-sounding language. I then shift focus to the particular type of formulaic language studied in my thesis, the lexical bundle. I define its characteristics and review what researchers have generally found about the use of lexical bundles in written registers. I then turn to a review of the studies of lexical bundles in second language academic writing, particularly those in a Chinese context, with the aim of providing the niche for my research. Finally, I present the research questions for the present study.

2.2 Formulaic language

In general, formulaic language can be defined as "a sequence, continuous or discontinuous, or words of 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" (Wray, 2002, p. 9). In other words, formulaic language is a fixed set of words (that may have open slots in it) that is currently theorized to be stored and retrieved as whole phrases, or chunks, rather than being analyzed or comprised of the separate parts.   Formulaic language is likely an important

building block of first language acquisition, as a review of the research in this area suggests that formulaic language is a central part of the acquisition process (Bannard & Lieven, 2012).

Formulaic language has also been found to be important for second language learners. The set phrases or chunks allow learners to have an early grasp on pragmatically appropriate language even while they may struggle with other aspects of language learning (Wood, 2015). For example, Schmidt (1983) followed the language learning of an adult learner who, while he struggled with grammatical aspects of the language, was able to rely on formulaic sequences to effectively communicate. In later research, Ellis (1996) claims that these formulaic sequences serve as a basis for more advanced language learning, where these sequences serve as the building blocks for later grammar acquisition.

There has also been a growing body of research looking at the importance of acquiring formulaic sequences for writing. While writing contains less formulaic language than speech (Biber, et al., 2002), researchers have noted that having mastery over formulaic sequences common to writing is necessary for learners to produce writing that is acceptable to native speakers (Cowie, 1992). Additionally, with more students entering into academic language programs, and these programs' focus on mastery of academic writing, a handle on these written forms is essential for success.

While formulaic language has been shown to be important for language learners, both for ease of learning and accelerated pragmatic competence, the study of formulaic language can sometimes be hard to follow because many different types of formulas are

studied: fixed expressions (Jesperson, 1924), extended collocations (Firth, 1951), lexical

phrases (Nattinger & DeCarrico, 1992), to name just a few (Wray, 2000; Wood, 2015).

Some of these terms are clearly defined, and others are not. The present study will focus

on a defined type of formulaic language known as lexical bundles.

2.3 Defining lexical bundles

Lexical bundles are defined as recurrent sequences of three or more words that

occur in high frequency across texts (Biber, et al., 1999). While these multiword

sequences can be complete phrases whose meaning can be derived from the parts (e.g. *on

the other hand*), they can also be incomplete units (e.g. *it should be noted* or *the end of

the*). Lexical bundles are defined purely by their frequency of use and distribution across

texts as detailed in the next section. Lexical bundles thus contrast with idioms, which are

whole phrases with a meaning unrelated to the parts (Wei and Lei, 2011).

Lexical bundles have been described as "important building blocks of discourse"

(Biber and Barbieri, 2007 p. 263). This study suggests that these bundles are not used to

introduce new propositions, rather that they serve a functional purpose in written and

spoken discourse, and can be used to express stance, organize the discourse, and frame

referents. This study found that bundle frequency varies with register and whether the

bundles are spoken or written. While it is generally true that spoken language has a

higher frequency of lexical bundles than written (Biber, et al., 2002), Biber and Barbieri

(2007) found that when examining bundles in specific university registers, writing in

course management and instructional registers had a higher frequency of bundles than the

spoken registers they examined.

Previous studies have categorized lexical bundles by their structural characteristics. They vary in their specific categories, but all are generally focused on noun phrase, prepositional phrase, and verb phrase structures. These studies have found that structural use of bundles can vary according to discipline; for example, writers in the fields of biology and electrical engineering use bundles including a passive verb with higher frequency than do writers in applied linguistics and business (Hyland, 2008). Another study found that expert writers used noun phrase bundles with greater frequency than novice writers at the university level (Chen and Baker, 2010). That same study found that while novice writing varied greatly from expert writing, within the two novice writing groups (native and non-native English speakers), the use of structural bundles is remarkably similar.

Lexical bundles are also categorized by their functional purposes. Studies have found the three primary functions of lexical bundles are as follows: participant-oriented (also known as stance expressions), which are used to express author attitudes or assessments of another proposition; text-oriented (or discourse organizers), which serve to demonstrate a relationship between sections of the writing; and research-oriented (or referential expressions), which are used to make direct or indirect reference to other texts (Biber and Barbieri, 2007; Hyland, 2008; Wei and Lei, 2010). As they fulfill these functions, the appropriate use of lexical bundles can also be determined by the discipline in which the author is writing (Hyland, 2008).

Taken together, the previous studies have shown that lexical bundles vary in spoken and written discourse, in different registers in university settings, between novice

and expert writers, and among native and non-native writers. So while lexical bundles are important building blocks of discourse, as mentioned previously, the bundles required for different types of spoken and written discourse differ greatly. Identifying the types, structures, and functions of bundles in different registers, disciplines, and proficiency areas can help increase our understanding of how these bundles vary in their foundational aspects.

2.4 Identifying lexical bundles

Lexical bundles can be identified only by employing techniques from corpus linguistics, using computer-assisted identification and then categorizing functions by examining bundles in their discourse contexts. For identifying lexical bundles, two aspects of frequency are important: the frequency of occurrence and the range of texts the bundles are found in. Every potential bundle must be identified and every occurrence counted in every text. It is unfeasible to do this over a large number of texts without using computer-assisted techniques. Software tools called concordancers can locate and count all occurrences of potential lexical bundles (Cortes, 2004).

The threshold for frequency of occurrence depends on the type of corpus examined. Since bundles are more common in spoken registers than written, in examining spoken data in large corpora (over one million words), researchers usually look for bundles that occur at least 40 times per million words and across at least five texts (Biber, et al, 1999). This range is necessary, since if a lexical bundle occurs across many texts, it is more likely to be a formulaic sequence than simply an idiosyncrasy of the author (Biber, 2009). Since writing tends to have more diverse word use (Biber et al,

1999) the frequency tends to be set lower, looking for those that occur at least 20 to 25 times per million words, but still across a range of at least five texts.

2.5 Previous studies of lexical bundles in academic writing

Previous studies of lexical bundles in writing have shown differences between L1 and L2 writers (Ädel & Erman, 2012), between novice and expert writers (Chen & Baker, 2010), and between different disciplines (Cortes, 2004; Hyland, 2008). Comparing across groups in academic writing revealed differences in overall frequency of use, the structural types of bundles used, and the functions that these bundles served.

Ädel and Erman (2012) examined lexical bundles in the L2 English writing of native Swedish speakers, and compared those to lexical bundles found in native English speaker writing at a similar level. They reported that in their data, non-native speakers used lexical bundles with less frequency and overall type variation than native speaker writers did. There were also differences in the functional use of lexical bundles across both writing samples, where native speakers used more stance (or participant-oriented) bundles and fewer discourse organizers (text-oriented bundles) than did the non-native speakers.

Lexical bundles can also vary across discipline, as Hyland reported in his 2008 study. He looked at lexical bundles across research articles in biology, electrical engineering, applied linguistics, and business studies. He found that the types of bundles differed across disciplines. Additionally, he found that within disciplines, the distribution of functional uses of bundles varied, so, for example, applied linguistics studies used participant-oriented bundles more frequently than those in biology. Hyland suggested that

these bundles may play an important role in differentiating disciplinary-specific writing, and that mastery of them can help writers gain competence in the discourse in their field of study.

Cortes (2004) found that students in history and biology used few of the same bundles as did professionals in those fields. In history writing, students used fewer text-oriented bundles than did the professionals, and in general used the bundles with more repetition in a single text. Student writers in biology also had few shared bundles with professional writers, though this could be explained by many biology-related bundles being topic-focused and specific to high levels of research in the field. Here, students also used fewer quantification bundles.

Overall, previous research on lexical bundles has found that the lexical bundles are used differently by writers of different ability levels, language backgrounds, and across disciplines. These variations are found in the overall frequency of bundles used, the types of bundles used, the structures of these bundles, and their functional use.

Increasingly, researchers in China are interested in examining the lexical bundles present in the academic writing of tertiary students and published writers. Wei and Lei (2011) reviewed previous research conducted in this area. In one study, the researchers compared the 191 most frequently found three-word bundles from native English speakers' expository writing to those found in timed essays written by Chinese EFL learners. They found that while about a third of the bundles were used in similar proportions across native and non-native speakers' writing, another two-thirds were used with less frequency in the Chinese students' writing than in the native speakers' writing.

Conversely, they reviewed another study (Pang, 2009) comparing Chinese EFL learners'

argumentative writing to that of student native English speakers. That study found that

Chinese students used bundles more frequently than the native speakers, but that they

varied in terms of functional use. In their own study, Wei and Lei compared bundles

found in the doctoral dissertations of Chinese EFL learners to those found in professional

writing in the same field. They found that learners used bundles more frequently than

professional writers, and that the type and function of these bundles used by learners

differed from professional writers.

Chen and Baker (2010) also compared Chinese EFL learners' use of lexical

bundles to those used in native speaker writing at the undergraduate and professional

levels. They found that not only did the bundles in student writing of both non-native and

native English writers differ from professional writers' in terms of frequency, type and

function, but also that the writing of the novice writers was quite similar in these same

areas.

Previous research demonstrates that while learners use bundles with different

frequency and functions than native speaker writers, there is not one description of what

learners do. It is difficult to tease out the difference between proficiency levels and types

of texts, since the writing that has been examined also varies.

2.6 Research question

My study seeks to fill the gap of knowledge about the type and function of lexical

bundles present in Chinese EFL learners' undergraduate theses and how they compare to

the lexical bundles used by professional writers. I compare the frequency, structures, and

functions of lexical bundles between Chinese university study writing and expert writing, and across the disciplines of applied linguistics and literature. The present study aims to answer the following research question:

> What are the differences between the lexical bundles used by Chinese university students and expert writers in the fields of applied linguistics and English literature with respect to the bundles (a) frequency, (b) structural types, and (c) functions?

Chapter 3 – Methodology

This study compared the frequency, structure, and function of lexical bundles found in literature and applied linguistics writing produced by published writers and theses from undergraduate students at a university in China. This chapter describes the methodology of the present study. The first section will introduce the university from which the undergraduate writing was taken. Within this section I will introduce the general characteristics of Longdong University and describe the thesis writing requirements for Foreign Language students. The second section reports the design of the corpora used in the study and corpus compilation procedures. The final section describes the data analysis methods.

3.1 Context of the study

This study included undergraduate theses from students in the School of Foreign Languages and Literatures at Longdong University, a four-year university located in the eastern part of Gansu province in north-central China. As of the 2015-2016 academic year, this department had 578 students total, with 20 "classes" of students over all four years. These "classes" are organizational units in Chinese schools with each class consisting of approximately 25-30 students. Each class takes core courses together, lives in dormitories together, and participates in weekly class activities together.

At Longdong University each grade level has five classes; of these, four are preparing to become primary or middle school teachers, and one is preparing for careers in translation or interpretation (Waiguoyu Xueyuan Jianjie, 9/1/2015). The core courses include English language and literature throughout the four years. All students are

required to write a thesis in their fourth year. Most of these theses are written in English. Those written in Chinese usually analyze some detail of either Russian or Japanese, which are foreign languages students are required to take for two years.

By the senior year, most students are at an intermediate proficiency level in English, having studied English for 8-12 years prior to university admission and then completed four years of college course work in English. Students take three semesters of writing courses, including a practical business writing course during their freshman year and a two-semester writing course spanning the final semester of the sophomore year and first semester of the junior year. The practical business writing and general writing courses are most commonly taught by foreign teachers. The content of these courses varies depending on the teacher, and these courses may or may not serve as preparation for writing a thesis.

3.1.1 Requirements for a senior thesis

During their senior year, students are given printed instructions that detail the university's requirements for a thesis. These have been prepared by faculty in the College of Foreign Languages and Literature, in accordance with the guidelines set forth by the Ministry of Education, which requires all students at four-year universities to write a thesis in their senior year. At this university, there is no course associated with this document. The document is in Chinese and details the required sections of the thesis. It specifies the thesis should include a cover page, a statement of academic integrity, an abstract with at least three key words, and the main text of the document. With the exception of the main text, all components should have an English and Chinese version.

Within the document two types of theses are described: theoretical and empirical.

Literature theses are theoretical in nature, while applied linguistics theses follow the

empirical research writing guidelines. While both types require an introduction, body,

and conclusion, what is to be contained in each section varies by type. A summary of

these requirements is shown in Table 1. The document also details general formatting

requirements, such as how to number chapters, which font and font size are appropriate,

and how to add page numbers. It is important to note that while the document details

what should be included within each section, it gives no guidance on *how* to present the

information.

Table 1. Guidelines for senior thesis writing.

Section	Theoretical	Empirical
Introduction	Address the subject source, significance of the subject, review literature, and current situation of subject	General description of the study, including research question(s)
Body	Address different viewpoints and give examples as needed; should follow a logical structure.	Three parts: 1. Literature review – situate the current study based on previous research in the area. 2. Research methods – detail research questions, the context of the study and participants, the research tools, data collection, data analysis, and limitations. 3. Results and discussion – report findings following order of research questions, supported by tables and charts as needed; discussion is for writer to express their own viewpoints
Conclusion	Review the main points discussed in the body of the paper.	Three parts: 1. Main research findings – re-state the main findings succinctly. 2. The significance of the study, and whether or not findings were as predicted 3. Recommendations for future research

In addition to the thesis requirements document, students also receive two sample thesis outlines, which serve to illustrate the general formatting requirements. They also are given a document detailing APA citation styles, which is presented in English.

When preparing to write the thesis, students select a topic from a faculty-prepared selection of approximately 200 topics that cover applied linguistics, literary analysis, cultural studies, or translation theory. Some example topics are as follows: Genre-Based Approach to Teaching Reading in High School in China, On Cultural Impact of College English Teaching in China, The Philosophy of Life in *The Old Man and the Sea.*

The final requirements are that a thesis is at least 5000 words in length, contains all required sections of theoretical or empirical research, and follows both APA citation style and the university's formatting guidelines. Students have the final weeks of the winter semester and the start of the spring semester to complete their theses. During this time, they work with an advisor to finish and revise drafts. Students are required to submit and revise at least three drafts of the thesis. In total, from topic selection to thesis defense, students have about four months.

3.2 Corpus design and compilation procedures

Two corpora were compiled for this study. One was a learner corpus, consisting of applied linguistics and literature texts published during the 2012-2013 and 2013-2014 academic years at Longdong University. The second was a corpus of professional writing for comparison, consisting of applied linguistics and literature articles published in refereed journals between 2012 and 2013. The following sections detail the design and

compilation procedures for both corpora.

3.2.1 Learner corpus

Any student thesis that was written in English and published during the 2012-2013 and 2013-2014 academic years at Longdong University was eligible for inclusion. These theses were published and made publically available in the university library; no electronic versions existed. I compiled a list of theses published between those years, following class order and year of publication, so the list began with Class 1 theses from the 2012-2013 academic year and continued sequentially through Class 5 from the 2013-2014 year. Student theses written in Chinese were not included on this list. Student thesis topics included applied linguistics, literature, translation theory, and cultural studies. For the purpose of this study, I selected two research areas, applied linguistics and literature to compare. In total, there were 108 theses in applied linguistics and 55 in literature. From this list, I selected every third thesis in each topic area. This resulted in a total of 34 applied linguistics texts and 17 literature texts to be included in the learner corpus. The small number of texts, especially in literature, limits the generalizability of the findings, but still can provide useful initial information about these Chinese students' lexical bundle use.

I re-typed the abstracts and main text of these theses, excluding the title page, acknowledgements, tables and charts, and reference sections, since those are not within the scope of the present analysis. Since the theses were retrieved from the library, no information about student performance was available.

3.2.2 Professional corpus

The comparison corpus was compiled from published articles from representative academic journals of applied linguistics research and literary analysis. Professional articles are not a precise equivalent to the student writing corpus, but I made the choice to use them, rather than student theses from a country where English is widely spoken, because they are the closest representation of the ultimate target of the Longdong students. A previous study examining frequency and structural types of bundles showed that both L1 and L2 university level writers differed in their usage of bundles from expert writers (Chen and Baker, 2010). Since that study showed that L1 writers at a similar level of disciplinary expertise to the L2 writers still were using bundles differently from the experts, comparing the present learner writing samples to those of writers at a similar level would only yield a description of the differences. The present study is interested in not only a description of the learner differences, but also in providing recommendations and suggestions for areas in which learner writing could ultimately be improved. In support of this aim, and similar to Cortes' (2004) study of lexical bundles in native-speaking student and professional writing, I made the choice to select expert writing for comparison, since that would show differences between my students and their target, revealing areas in the learner writing which could be improved.

The journals used are listed in Table 2. The journals were recommended by the Portland State University library website as prominent in their research areas. In order to include journals that varied in style, I selected journals published by different publishers. I also aimed to include journals that focused on a variety of topic areas in order to avoid

repetition of content-specific bundles. Because the learner corpus in literature was smaller than that of applied linguistics, one less literature journal was selected.

Table 2 Journals used for the professional corpora

Applied Linguistics Journals
Language Learning and Technology
Studies in Second Language Acquisition
System
TESOL Quarterly
Literature Journals
American Literature
Studies in Romanticism
Studies in the Novel

I compiled a list of all articles from those journals in the same manner as that of the learner corpus; that is, texts from the first volume published in 2012 were listed in the order of appearance and the list ended with the final volume published in 2013. Although the learner corpora were relatively small in size, I wanted to have a larger professional corpus for comparison because the professional articles are likely to have greater variation within them. For example, the students use a limited number of topics, while the professional topics vary greatly. The students are also all from a similar background, while the article authors represent more diverse background. Capturing the variation and seeing lexical bundles within professional writing thus requires a larger corpus. I therefore decided to select approximately double the number of texts from professional writing over learner writing. This resulted in selecting every fourth journal article from the compiled lists. In total, 58 texts were selected from the applied linguistics journal and 31 texts were taken from the literature journals.

The size of each corpus is presented in Table 3. Not surprisingly, since the

number of texts are higher, the professional corpora are much larger than those of the students. Previous studies similar to mine have also been conducted comparing corpora of different sizes (e.g. Adel and Erman, 2012; Cortes, 2013; Hyland, 2008). Because of this disparity in size, all frequencies reported in the data are not the raw frequencies but have been normed to their occurrence per 100,000 words.

Table 3. Corpora used in the present study

Research Area	Word count - Professional	Number of texts - Professional	Word count - Learner	Number of texts - Learner
Applied Linguistics	399,212	58	138,045	34
Literature	253,566	31	71,752	17
Total	652,778	89	209,797	51

3.3 Data analysis

The analysis for the study has three parts. The first concerns frequency, the second focuses on structural type, and the third concerns function. Before any analysis, however, I had to decide on the length of bundles to investigate. Previous work has ranged from three-word to five-word bundles (Cortes, 2004). However, most studies examining lexical bundles focus on four-word bundles because they occur with greater frequency than five-word bundles, and many three-word bundles are in fact part of longer, four-word bundles (Chen and Baker, 2010; Cortes, 2004). I therefore decided to focus primarily on four-word bundles, including five-word bundles which represented an overlap in two four-word bundles (e.g. *at the beginning of* and *the beginning of the* become *at the beginning of the*).

3.3.1 Frequency analysis

In identifying lexical bundles, two aspects of frequency are important: the frequency of occurrence and the range of texts the bundles are found in. The threshold for frequency of occurrence depends on the type of corpus examined. As mentioned in the literature review, lexical bundles are less frequent in written language than spoken. While the lowest frequency set for spoken data tends to be 40 times per million words (e.g. Biber, et al. 1999), for written data this threshold can be set lower, looking for bundles occurring at least 10 times per million words, though some studies set the frequency higher at 20 times per million words (Cortes, 2004). Bundles also need to occur across a range of texts. This range is necessary because if a lexical bundle occurs across many texts, it is more likely to be a formulaic sequence than simply an idiosyncrasy of the author (Biber, 2009). The range that is set varies depending on the study, but is generally set at between three and five texts (Chen and Baker, 2010).

For the present study, since some corpora are relatively small in size, I chose to look at bundles that occurred over a range of four texts and that occurred at least 20 times per million words. This is in line with previous research comparing learner and professional writing (e.g. Chen and Baker, 2010; Hyland, 2008), though there is no absolute agreement in research about frequency and range cut-offs.

Once these parameters were set, I extracted potential bundles using AntConc version 3.4.3, a free concordancing program developed by Lawrence Anthony (2014). I then manually checked for overlaps in the data and refined the raw data set. For example, *at the beginning of* and *the beginning of the* both occurred with the same frequency and

range, and when checked were both part of a larger five word bundle *at the beginning of the*. In other cases, some four-word bundles were part of a larger five word bundle, but also met inclusion criteria as a four-word bundle. For example, *pay more attention to* was also included in the five word bundles *should pay more attention to* and *pay more attention to the*, in addition to existing as a four-word bundle. In these cases, the five-word bundles were extracted and the counts of the original four-word bundles were adjusted in order to ensure each occurrence of a bundle was only represented in the data once.

My final list of lexical bundles therefore was all the four- and five-word lexical bundles that occurred at least 20 times per million words in one of the corpora and in at least four texts in that corpus. The first part of my research question was answered by comparing the frequency of lexical bundles across the disciplines and across the writer groups.

3.3.2 Structural analysis

I then classified bundles according to structural type. I followed the guidelines developed by Biber, et al. (1999) as refined in Hyland (2008). The eight categories are presented in Table 4. The second part of my research question was answered by comparing frequency and use of structural types between expert writers in both fields, and between learners and experts within each field.

Table 4. Structural categorization of lexical bundles

Structure	Example
Noun phrase + *of*	*the end of the*
Other noun phrases	*the fact that the*
Prepositional phrase + *of*	*as a result of*
Other prepositional phrases	*in the present study*
Passive + prepositional phrase fragment	*is based on the*
Anticipatory *it* + verb/adj	*it is important to*
Be + noun/ adjectival phrase	*is the same as*
Others	*as shown in figure*

3.3.3 Functional analysis

I then analyzed the bundles by function. The functional categorization followed the system developed and implemented in previous research (Hyland, 2008, pp. 13-14). It includes three broad functional categories, described and exemplified in Table 5.

Table 5. Functional categories of bundles

Research – report on the content of the text	location	*at the end of*
	procedure	*the role of the*
	quantification	*one of the most*
	description	*the structure of the*
	topic	*Chinese learners of English*
Text – organize and present the text	transition – additive or contrastive signals	*on the other hand*
	resultative – marking inferential or causative relationships	*it was found that*
	structuring – organize the structure of text	*in the next section*
	framing – frame arguments	*with respect to the*
Participant – engage audience or display writer's stance	engagement	*it should be noted*
	stance	*it is possible that*

I looked at occurrences of each bundle in context to categorize its function within the text. Some bundles were more difficult to categorize, as systems of categorization were not clear on the distinction between description and topic. I made my best efforts to categorize the bundles according to descriptions given in previous research. For bundles which seemed to have multiple functions, such as *at the same time,* I categorized them according to which function seemed most dominant based on their use in context. So while Hyland (2008) categorized *at the same time* as locational bundles, I found that the majority of the ones present in my corpus were transition bundles, so I analyzed the data as if they were all transition bundles. I did this in lieu of analyzing them according to the proportional use of each function, due to the quantity of data and time limitations. I will discuss the limitations of using this categorization method further in Chapter 5. The third part of the research question was answered by comparing the functions of the bundles within the two subjects and between experts in each subject area.

Chapter 4 - Results and Discussion

In this chapter I present and discuss the findings of the study. First, I focus on overall frequency of bundles found across the learner and professional corpora. Next, I discuss the structural types of bundles found within each corpus, and compare those found between and within subject areas. Finally, I present and discuss the functional categories bundles, and compare the functional uses of the bundles between the subject areas and between professional and learner writing within the same subject area.

4.1 Overall frequency of lexical bundles

This section answers the first part of the research question, which is how frequency of bundles vary between proficiency levels and subject areas. In reporting the frequency of lexical bundles, all numbers reported are normed to their frequency per 100,000 words. For reference, a complete list of all bundles from the professional and learner corpora can be found in the Appendix.

Figure 1 summarizes the overall counts of bundles used within each group. It shows that writing in applied linguistics at both the professional and learner level contains more types of bundles and more tokens of those bundles than writing in literature. However, learners writing in applied linguistics used more bundles with greater frequency than the professionals did. The findings were opposite for writing in literature, where learners used fewer types of bundles and with less frequency than the professional writers did.

The greater use of bundles overall in applied linguistics writing may be due to the way applied linguistics writing is structured, where writing in the methods and results

sections of the articles can make use of more formulaic language. Additionally, the

applied linguistics writing had a focus on teaching topics and second language acquisition,

and there may be more formulaic language to describe these focuses.
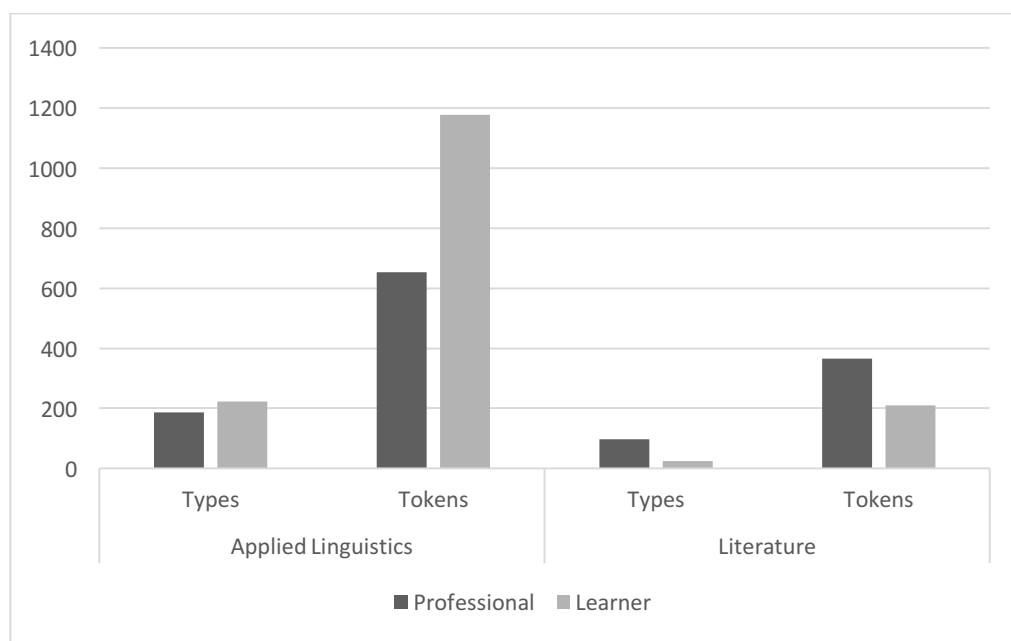
Figure 1. Counts of types and tokens of bundles within groups (per 100,000 words)

4.1.1 Frequent bundles between professional groups

Within the professional writing, literature used fewer overall tokens and types of

bundles than linguistics. The twenty most frequent bundles used by all groups are

displayed in Table 6, where the two professional groups shared six bundles. This shows

that while bundles are used over similar ranges of texts in both subject areas, these

bundles are used with far greater frequency in linguistics writing. One exception was *at*

the same time, which was used with greater frequency in literature. Unusually frequent

use of this bundle appeared to be related describing what the literature does and how it affects the readers.

In this way, wonder holds us spellbound, while *at the same time* transporting our sense to new heights of delight. (Literature, 2013, 4).

Wallace writes posthumanist fiction that retains and affirms humanism *at the same time* it acknowledges and grapples with twenty-first-century social and cultural milieux that are hostile to that same humanism. (Literature, 2012, 26).

The writing in linguistics also used this structure, but primarily used it in order to report on activities happening simultaneous during the course of the research.

A similar small percentage strongly agreed or agreed that it was difficult listening to aural input and reading test items *at the same time*, with the majority of students tending to be neutral. (Linguistics, 2013, 32).

Overall, the greatest difference between the groups of writers at the professional level is the frequencies of bundles used. While professionals share a limited portion of bundles, those in linguistics used them with greater frequency and over a larger range of texts than did the writers in literature.

Learners shared five bundles between the two groups (Table 6), but what is notable about their twenty most frequent bundles is that many of the bundles are focused on the topic (e.g. *in junior middle school; falls in love with; students interest in learning).* The bundles give an indication of what they were writing about, whereas the professional bundles do not give a clear indication of what the study focused on.

Table 6. Twenty most frequent bundles in professional writing across both subjects

Professionals - Linguistics	Freq	Range (58)	Professionals - Literature	Freq	Range (34)
as well as the	58	15	**at the same time**	**38**	**15**
on the other hand	50	13	**in the united states**	33	**13**
the extent to which	49	12	**as well as the**	**20**	**8**
the results of the	46	12	in the face of	20	**8**
in the context of	45	11	in the first place	19	7
in the present study	42	11	on the one hand	19	7
in the united states	40	10	in the midst of	18	7
in the current study	37	9	the figure of the	18	7
at the same time	36	9	**in the context of**	17	7
there was a significant	31	8	**at the end of the**	**16**	**6**
in the target language	30	8	the end of the	16	**6**
of the present study	30	8	in the words of	14	**6**
the fact that the	27	7	as a form of	13	**5**
as a result of	26	7	in the form of	13	**5**
it is important to	26	7	the story of the	13	**5**
in terms of the	24	6	as a kind of	12	**5**
at the end of the	23	6	in a way that	12	**5**
At the beginning of the	22	6	in the case of	12	**5**
the use of the	22	6	in this way the	12	**5**
in the field of	21	5	**on the other hand**	**12**	**5**

Learners – Linguistics	Freq	Range (34)	Learners – Literature	Freq	Range (17)
task based language teaching	66	4	**at the same time**	**24**	**13**
in the process of	37	15	falls in love with	10	6
at the same time	**24**	**13**	in the novel the	10	5
as well as the	**23**	**12**	at the same time she	10	4
that is to say	20	13	we can see that	10	5
in junior middle school	19	6	**is one of the**	**8**	**5**
is one of the	**16**	**10**	all over the world	8	4
on the other hand	**16**	**12**	**plays an important role in**	**8**	**5**
is a kind of	15	10	**on the other hand**	**8**	**4**
to communicate in English	15	4	one of the most	8	4
between teachers and students	14	7	in the 19th century	7	5
of the target language	14	9	**as well as the**	**7**	**4**
students interest in	14	5	for the first time	7	4

learning					
plays an important role in	**13**	**7**	in the nineteenth century	7	5
in foreign language learning	12	6	in the process of	7	4
in order to make	12	10	the heroine of the	7	5
in the field of	12	10	at the same time the	7	4
junior middle school students	12	4	with the development of	7	5
on the basis of	12	9	at the end of the	6	4
the background of the study	12	11	the beginning of the novel	6	4

4.1.2 Shared bundles in applied linguistics

Within the applied linguistics corpora, 31 bundles occurred in both corpora, displayed in Table 7 below. In total, 14% of the bundles in the learner corpus were also present in the professional corpus.

Of these bundles, most occurred with similar frequency and over a similar range of texts; however, there were a few exceptions. The novice writers in the university corpus used the bundles *of the target language*, *in the process of*, *is one of the, to participate in the*, and *it is necessary to* with higher frequency and over a wider range of texts than did the professional writers. At the same time, professional writers used the bundle *the results of the* more frequently and over a greater range of texts than did the learners.

Table 7. Shared bundles across professional and learner writing in applied linguistics

Professional	Freq	Range	Learner	Freq	Range
as well as the	15	55%	as well as the	17	33%
on the other hand	13	38%	on the other hand	12	33%
the results of the	12	40%	the results of the	6	22%
at the same time	9	33%	at the same time	17	36%
it is important to	7	29%	it is important to	4	14%
in terms of the	6	29%	in terms of the	6	17%
at the end of the	6	24%	at the end of the	4	11%
at the beginning of the	6	17%	at the beginning of the	6	19%
the use of the	6	22%	the use of the	5	17%
in the field of	5	24%	in the field of	9	28%
on the one hand	5	19%	on the one hand	8	25%
so that they can	4	16%	so that they can	6	11%
to be able to	4	22%	to be able to	3	11%
that the use of	4	16%	that the use of	3	11%
of the target language	4	17%	of the target language	10	25%
the purpose of the	4	16%	the purpose of the	8	14%
the quality of the	4	14%	the quality of the	3	11%
on the basis of the	3	14%	on the basis of the	4	17%
english as a foreign language	3	17%	english as a foreign language	5	17%
on the basis of	3	17%	on the basis of	9	25%
of teaching and learning	3	10%	of teaching and learning	3	11%
a better understanding of	3	16%	a better understanding of	3	11%
in the process of	2	16%	in the process of	27	42%
is one of the	2	10%	is one of the	12	28%
it can be seen	2	9%	it can be seen	4	14%
it is clear that	2	14%	it is clear that	4	14%
to participate in the	2	14%	to participate in the	8	28%
to the development of	2	16%	to the development of	5	19%
it is necessary to	2	9%	it is necessary to	8	19%
on the other hand the	2	12%	on the other hand the	5	17%
to make use of	2	9%	to make use of	5	14%

Learners tended to follow *is one of the* with the word *most*. In fact, *most* followed that bundle in over a third of the occurrences. Professionals used the string *is one of the most* only once, instead following with more descriptive adjectives like *earliest, few,* or *key*. Examples of the usage are shown below.

Role play *is one of the* most important ways to learn English in middle school. (Learner, 2013, 8).

To be specific, most students do not take an interest in learning English; most of them consider English *is one of the* most important compulsory courses in school curriculum. (Learner Ling, 2014, 2).

Because (language) education *is one of the* key avenues for social change, these ideologies are as potent as textually and visually countable instances of cultural bias or stereotypes. (Professional, 2013, 12).

This claim *is one of the* central tenets of Vygotskian approaches to developmental education. (Professional, 2013, 28).

Conversely, professional writers showed little variation with the word following the bundle *to participate in the*. The word *study* followed the bundle in two thirds of all occurrences, and *project* following the remaining third. All but one occurrence of *to participate in the* was followed by one of those two words. Of the occurrences in the learner corpus, *study* only followed this bundle once. Both *activit(y/ies)* and *life* occurred after this bundle more than one time, with three and two cases respectively.

A total of 51 students elected *to participate in the* study initially, but since data

was collected on five separate occasions, there was some attrition. (Professional,

2013, 40).

Courseware can not be too abused, otherwise using courseware too much will

reduce students' opportunities *to participate in the* activities of language practice

and neglect the cultivation of language ability. (Learner, 2013, 8).

The use of *in the process of* and *it is necessary to* will be discussed further in later

sections, where the former will be addressed in the structural analysis, and the latter

addressed in the functional analysis.

Within applied linguistics writing, learners used a wider range of bundles, and

they used them with more frequency than did the professionals. The most frequently used

bundle in the learner corpus, *task-based language learning*, occurred 66 times per

100,000 words. However, this bundle only occurred over a range of four texts. Students

who used this bundle frequently relied on its repetition, rather than using a wider range of

language to refer to this method in different ways. An excerpt taken from one learner

thesis showed repetition of this bundle three times over three adjacent sentences:

> *Task-based language teaching* is one of the means of situational teaching which is
>
> assessed by the results of the tasks but not by the scores of the tests. In *task-based*
>
> *language teaching,* students have specific motivation to learn in dependently and
>
> can develop their language skills effectively. There are some points that need to
>
> be paid some attention when implementing *task-based language teaching*. (2013,
>
> Text 8)

This bundle did not occur with high frequency in the professional corpus, though seven of the articles did mention task-based approaches to teaching.

In professional applied linguistics writing, the most frequent bundle was *as well as the*, which occurred at a frequency of 15 times per 100,000 words, and had a wide range of occurrence, where it was present in a range of 32 texts (55% of total sample). This bundle also occurred in the learner corpus, where it was used with greater frequency (17/100,000 words), though with a more restricted range, only 33% of the texts, as shown in Table 3). In other words, fewer students use this bundle, but they use it more.

4.1.3 Frequent bundles in literature

Nine bundles total were shared between professional and novice writers, summarized in Table 8. These shared bundles account for 35% of the total types used by learners. Of these, most were used across a similar range of texts as those in the professional corpus. However, the learners tended to use these bundles more frequently than the professional writers.

Table 8. Shared bundles across professional and learner writing in literature

Professional	Freq	Range (% of texts)	Learner	Freq	Range (% of texts)
at the same time	15	77%	at the same time	24	76%
as well as the	8	42%	as well as the	7	41%
at the end of the	6	29%	at the end of the	6	24%
on the other hand	5	26%	on the other hand	8	29%
one of the most	5	32%	one of the most	8	24%
at the same time the	3	22%	at the same time the	7	24%
in the novel the	3	16%	in the novel the	10	29%
in the process of	3	13%	in the process of	7	24%
for the first time	2	13%	for the first time	7	24%

Of the shared bundles, the one with greatest difference in usage was *in the novel the*. The learners used this at a rate of 10/100,000 words, while the professional writers used it only 3/100,000 words. It appeared over a greater range of texts for the learners. The learners were more likely to use this bundle at the start of a sentence, while for professional writers, this bundle spanned two clauses:

> *In the novel, the* author portrays the image of an old Cuban fisherman who has gone 84 days with catching nothing. (Learner, 2014, 4).

> It holds true for almost everybody *in the novel: the* aesthetic attitude is the default life-view of the contemporary society portrayed therein. (Professional, 2012, 25)

While within applied linguistics writing, learners used more types and tokens of bundles than did professionals, the opposite was true for the literature corpora. Here, professionals used a greater number of types and tokens of bundles than did the learners. Professional writers had 98 distinct bundles, whereas the learners had only 26 different bundles. A complete list of these bundles is presented in the Appendix.

The most frequent bundle in literature writing was *at the same time,* which occurred in about three quarters of the texts of both professional and learner writers, though learners used this bundle with greater frequency, with a frequency of 24/100000 words as compared to the professionals use of 15/100000 words. Examples include the following:

> The hunting nightmare, the lighting-laden storms, the cadaverous apparitions, the desire of sleeping with dead body, and the wandering ghost and so on, these things will definitely scares us and yet *at the same time* spur us to read on.

(Learner, 2013, 4)

They would belong to a world in which loss and woe would not be possible, while

at the same time admitting that such a world would have to do without human life.

(Professional, 2012, 2)

4.1.4. Summary of frequency findings

In comparing professional writing in both disciplines, applied linguistics writing

contained more distinct types and used those with greater frequency than did writers in

literature. Comparing within fields, learners in applied linguistics used more types of

bundles with greater frequency than the professionals. Of these bundles, learners had a

high repetition of certain bundles, with *task-based language learning* occurring with the

greatest frequency. This could reflect a reliance on repetition of certain bundles rather

than using a wider variety. Learners and professionals also varied in word choice

following certain lexical bundles. With *is one of the*, learners had a more restricted set of

words that followed the bundle, while with *to participate in the*, learners employed a

wider variety of words. Within literature writing, learners used fewer distinct types of

bundles overall, but used them with higher frequency, which also suggests that they may

rely on repetition of known bundles instead of using a wider variety of bundles.

4.2 Overall distribution of structural types in applied linguistics and literature

Table 9 displays the distribution of structural types across professional and learner

writing in both subject areas. The most frequently used structure was *other prepositional*

phrase in all corpora with the exception of the learner applied linguistics corpus, where

other noun phrase was the most frequently used. Within the applied linguistics corpora,

the professionals and learners used *other prepositional phrase* with similar frequency, where it accounted for 24% and 23%, respectively, of all bundles used. Within the literature corpora, the learners used this structure more frequently, as it accounted for 43% of all bundles used in their writing. This structure accounted for 30% of the bundles used in the professional corpus of literature.

Table 9. Structural distribution of bundles across groups

Structure	Applied Linguistics		Literature	
	Professional	Learner	Professional	Learner
Noun phrase + *of*	17%	9%	16%	10%
Other noun phrase	19%	24%	5%	7%
Prepositional phrase + *of*	17%	11%	34%	15%
Other prepositional phrase	24%	23%	30%	43%
Passive + PP	4%	3%	0.5%	0%
Anticipatory *it* + verb/adj	6%	5%	2.5%	0%
Be + noun/adjectival phrase	2%	3%	0%	4%
Others	11%	22%	12%	21%
	100%	100%	100%	100%

In the following sections, I will discuss the usage of the *noun phrase + of*, *other noun phrases*, and *prepositional phrase +of*. These structures had differences both within and between subject areas. While there is some variation among the bundles in other structural categories, those structures were overall used infrequently, accounting for less than 10% of all bundles. The bundles in the *other* and *other prepositional phrases* categories, while accounting for a sizable portion of learner bundles across both subject areas, are not of interest for discussing structure, as no strong patterns emerged; however, certain bundles with these structures will be discussed in the section that focuses on functional purposes of bundles.

4.2.1 *Noun phrase + of* in applied linguistics

Bundles following the structure of *noun phrase + of* are displayed in Table 10 below. Those that occurred in both sets of data are bolded. This structure accounted for 17% of all bundles used in professional writing, and 9% of bundles used within learner writing.

Table 10. Lexical bundles using the structure *noun phrase + of* in applied linguistics

Bundle	Freq.	Bundle	Freq.
a better understanding of	3	**a better understanding of**	3
a great deal of	4	a great number of	3
a small number of	3	a large amount of	5
a wide range of	3	half of the students	3
native speakers of English	3	more than half of	4
one of the most	4	most of the students	7
the average number of	2	the background of the study	9
the content of the	4	**the content of the**	3
the effect of the	2	the cultivation of students	3
the effectiveness of the	2	the culture of the	4
the effects of the	4	the development of language	3
the end of the	3	the development of the	7
the findings of the	3	the importance of the	3
the focus of the	2	the influence of the	3
the impact of the	2	the process of learning	4
the interpretation of the	3	**the purpose of the**	8
the majority of the	3	The purpose of the study	4
the meaning of the	3	**the quality of the**	3
the members of the	4	**the results of the**	6
the nature of the	5	the role of the	3
the number of words	4	the significance of the study	8
the purpose of the	4	the ultimate goal of	3
the purpose of this	2	**the use of the**	5
the quality of the	4		
the rest of the	3		
the results of a	3		
the results of the	12		
the results of this study	5		
the second part of	2		
the total number of	4		
the use of the	6		
Total Type	31	Type	23
Token	111	Token	105

As Chen and Baker (2010) reported, most structures of this type follow the pattern of *the + N + of the*. In professional writing, 80% of the bundles used followed this pattern, whereas only 60% of the learner's bundles followed this structure. While both the learners and professionals used these bundles with similar frequency, where this structure occurred 69 times in the professional writing and 53 times in the learner writing, the learners used fewer types of bundles in this structure overall, having only 11 distinct bundles as compared to the professionals' 18. The words used to fill those slots are present in table 11 below.

Table 11. Nouns in the structure *the + N + of the* in applied linguistics

		Types	Tokens
Professional	**content**$_{(4)}$, effect$_{(2)}$, effectiveness$_{(2)}$, effects$_{(4)}$, end$_{(3)}$, findings$_{(3)}$, focus$_{(2)}$, impact$_{(2)}$, interpretation$_{(3)}$, majority$_{(3)}$, meaning$_{(3)}$, members$_{(4)}$, nature$_{(5)}$, **purpose**$_{(4)}$, **quality**$_{(4)}$, rest$_{(3)}$, **results**$_{(12)}$, **use**$_{(6)}$	18	69
Learner	**content**$_{(3)}$, culture$_{(4)}$, development$_{(7)}$, importance$_{(3)}$, influence$_{(3)}$, **purpose**$_{(8)}$, **quality**$_{(3)}$, **results**$_{(6)}$, role$_{(3)}$, significance$_{(8)}$, **use**$_{(5)}$	11	53

Of those nouns shared between the two groups, most were used with similar frequency. The learners used the structure *the purpose of the* more frequently than did the professionals, while the professionals used *the results of the* with greater frequency. For *the purpose of the*, this bundle occurred more frequently in the learner corpus simply because it was mostly frequently used as a section heading, whereas in the professional corpus it occurred within the text.

For *the result of the*, this bundle was used more frequently by the professionals,

where it often served the purpose of introducing results of statistical tests or directing the reader to a table where the results were presented, an example of which is given below. The learner theses generally did not rely on reporting statistical results, so this could be a reason why this bundle appears with less frequency in their texts.

> *The results of the* statistical analysis (ANCOVA) showed that the students who received advance organizers demonstrated significant improvement on the listening comprehension posttest while the control group did not. (Linguistics, 2012, 18)

4.2.2 *Noun phrase + of* in literature

This structure accounted for 16% of all bundles used within professional writing, and 10% of those used in learner writing. The types and frequencies of these bundles are reported in Table 12. The learners used a narrower range of bundles than the professional writers did. One bundle was shared between the two groups, though the learners used it with higher frequency than did the professionals. It occurred 8 times per 100,000 in the learner corpus, while only occurring 5 times per 100,000 words in the professional corpus.

Table 12. *Noun phrase + of* bundles in literature

Professional			Learner	
Bundle	Freq		Bundle	Freq
one of the most	5		**one of the most**	8
our understanding of the	3		the heroine of the	7
the emergence of the	2		the beginning of the novel	6
the end of the	6			
the figure of the	7			
the form of a	2			
the history of the	4			
the loss of the	2			
the materiality of the	4			
the perspective of the	2			
the rest of the	5			
the result of the	2			
the role of the	3			
the story of the	5			
the words of the	3			
the world of the	4			
Total Type	16		Type	3
Token	59		Token	21

Here, like in the applied linguistics corpora, the bundles following the structure *the + N + of the* accounted for most of the bundles. The nouns used to fill those slots are presented in Table 13. There were no shared bundles of this structure between the two groups. The professional writers used this structure more frequently and with more distinct types than did the learners, who only had two types accounting for 13 total tokens.

Table 13. Nouns in *the + N + of the* in literature

		Types	Tokens
Professional	emergence$_{(2)}$, end$_{(6)}$, figure$_{(7)}$, form$_{(2)}$, history$_{(4)}$, loss$_{(2)}$, materiality$_{(4)}$, perspective$_{(2)}$, rest$_{(5)}$, result$_{(2)}$, role$_{(3)}$, story$_{(5)}$, words$_{(3)}$, world$_{(4)}$	14	51
Learner	heroine$_{(7)}$, beginning$_{(6)}$	2	13

4.2.3 Comparison of *the + N + of the* in applied linguistics and literature

In examining the use of the pattern *the + N + of the* between the two professional groups, the nouns that filled the slots varied tremendously, as shown in Table 14. The two groups had only one shared bundle between them, *the rest of the*. This pattern was used with similar frequency and range between the two groups, with the primary difference being the noun chosen to fill the slot. This suggests that writers in each genre, while each relying on formulaic language, use genre-specific bundles. This could be because writers in applied linguistics need to report on the methods and results of the study, while writers in literature are describing events that occur in the prose being analyzed.

Table 14. Nouns in *in + N + of the* in professional corpora

		Type	Token
Linguistics	content$_{(4)}$, effect$_{(2)}$, effectiveness$_{(2)}$, effects$_{(4)}$, end$_{(3)}$, findings$_{(3)}$, focus$_{(2)}$, impact$_{(2)}$, interpretation$_{(3)}$, majority$_{(3)}$, meaning$_{(3)}$, members$_{(4)}$, nature$_{(5)}$, purpose$_{(4)}$, quality$_{(4)}$, **rest**$_{(3)}$, results$_{(12)}$, use$_{(6)}$	18	69
Literature	emergence$_{(2)}$, end$_{(6)}$, figure$_{(7)}$, form$_{(2)}$, history$_{(4)}$, loss$_{(2)}$, materiality$_{(4)}$, perspective$_{(2)}$, **rest**$_{(5)}$, result$_{(2)}$, role$_{(3)}$, story$_{(5)}$, words$_{(3)}$, world$_{(4)}$	14	51

4.2.4 Other noun phrases in applied linguistics

This structure accounts for 19% of the bundles used in professional writing, and 24% in the learner writing. The bundles used and their overall frequencies are reported in Table 15. This structure accounted for the largest proportion of bundles in the learner writing. Comparing between learners and professionals, the two groups only shared one bundle, *English as a foreign language*, which was used 3 times per 100,000 words in the professional writing, and 5 times per 100,000 words in the learner's writing. The learners used a wider range of bundle types than did the professionals, where learners had 47 distinct types and the professionals had 38 distinct types. Within professional writing, these types were represented by 124 different tokens of use, while learners used this structure more frequently, with 282 total tokens.

Table 15. *Other noun phrases* in applied linguistics writing

Professional		Learner	
Bundle	Freq.	Bundle	Freq
a one way anova	3	both at home and abroad	3
a significant difference between	2	**English as a foreign language**	5
a two way anova	2	English learning it is	3
Differences between L1 and L2	2	English majors in longdong university	7
English as a foreign language	3	English teaching and learning	6
English as a foreign language (EFL)	4	foreign language in the	3
English as a second language	3	foreign language teaching and	6
et al found that	2	interest in learning English	5
language learning and teaching	3	interest is the best teacher	5
participants were asked to	3	junior middle school students	9
previous studies e g	3	language teaching and learning	7
research is needed to	3	language teaching is to	4
second language (L2) learners	2	level of the students	3

Professional		Learner	
Bundle	Freq	Bundle	Freq
Significant difference between the two	3	listening speaking reading and writing	4
significant differences between the	2	listening speaking reading writing	4
students were able to	2	longdong university as its	3
students were asked to	4	middle school English teaching	7
students were required to	2	procedures and data collection	4
the degree to which	4	process of learning English	5
the descriptive statistics for	2	students ability to use	3
the extent to which	12	students are able to	4
The extent to which the	2	students are required to	3
the fact that the	7	students interest in English	5
the first research question	3	students interest in learning	10
the following research questions	3	students to improve their	3
the participants in the	5	students to use the	4
The participants in this study	4	table shows that students	7
the present study is	3	task based language teaching	66
the present study was	4	teaching and learning in	3
the results showed that	3	teaching and learning is	3
the second research question	3	teaching and learning the	3
the students in the	3	the first part is	4
there was a significant	8	the following questions what	3
there was no significant	2	the foreign language teaching	4
these results suggest that	2	the learning process and	4
they were able to	2	the new curriculum standards	3
they were asked to	2	the questionnaire is designed	3
you do not know	2	the relationship between the	8
		the students English learning	4
		the students in the	4
		the students interest in learning	4
		the target language and	5
		the teacher and the	4
		the teachers and students	3
		the traditional English teaching	3
		their interest in English	5
		we can see that	7
Total	Type	38	Type	47
	Token	124	Token	282

Among the bundles that start with *students+V,* it is interesting to note that while

the professional writers used the past tense of the *be* verb, students used the present tense.

Looking at these bundles in context shows that while professionals are using these

bundles for the purpose of reporting what happened, the learners are using these bundles

in order to discuss the applications of their research. Examples include the following:

> The present study shows that *students were able to* provide a balanced evaluation
>
> of LETs and NETs as English teachers. (Professional corpus, text 3)
>
> If the teacher can also provide creative situations in which *students are able to*
>
> speak any English they know in a natural setting, it will encourage youngsters to
>
> expand their vocabulary. (Learner corpus, text 1)

4.2.5 Prepositional phrase + *of* in applied linguistics

This structure accounted for 17% of the bundles used in the professional corpus,

and 11% of those used in the learner corpus. The bundles using this structure are reported

in Table 16, where bundles that occur in both professional and learner writing appear in

bold text. The professional writers used a larger variety of bundles within this structure,

with 31 types total and 112 total tokens. The learners used fewer types of bundles (23)

total, but overall used them with higher frequency (133 tokens total). A total of eight

bundles are shared between the two proficiency levels, accounting for over a quarter of

the learners bundles.

Table 16. *Prepositional phrases + of* in applied linguistics

Professional		Learner	
Bundle	Freq.	Bundle	Freq
as a result of	7	as a kind of	3
as part of a	2	as one of the	3
as part of the	3	**at the beginning of the**	6
as part of their	2	at the end of	3
at the beginning of	3	**At the end of the**	4
At the beginning of the	6	for the purpose of	4
at the end of the	6	**in terms of the**	6
At the time of the	3	in the course of	5
At the time of the study	2	**in the field of**	9
for each of the	3	**in the process of**	27
for the purposes of	3	In the process of English	4
in a number of	3	in the process of learning	6
in a variety of	4	In the process of teaching	4
in light of the	2	of the use of	4
in terms of the	6	**on the basis of**	9
in terms of their	3	**On the basis of the**	4
in the case of	5	**to the development of**	5
In the case of the	4	to the study of	5
in the context of	11	to the use of	3
in the development of	2	under the influence of	4
in the field of	5	with the development of	5
in the form of	4	with the help of	6
in the process of	2	with the rapid development of	4
of the process of	2		
on the acquisition of	3		
on the basis of	3		
On the basis of the	3		
on the part of	3		
through the use of	3		
to the development of	2		
with the use of	2		
Types	31		23
Tokens	112		133

Chen and Baker (2010) also found that the structure *in the + N + of* accounted for most of the *prepositional phrase + of* structures, and my findings are consistent with that. Table 17 presents the nouns that fill that slot. While the professional writers have more

distinct types following this structure, the learners have more tokens overall, which is influenced by their reliance on the bundle *in the process of,* which was present 27 times/100,0000 words.

Table 17. Nouns in structure *in the + N + of*   in applied linguistics

		Type	Token
Professional	case$_{(9)}$, context$_{(11)}$, development$_{(2)}$, **field**$_{(5)}$, form$_{(4)}$, **process**$_{(2)}$	6	34
Learner	course$_{(5)}$, **field**$_{(9)}$, **process**$_{(27)}$	3	41

The learners used the bundle *in the process of* with much greater frequency than did the professional writers. This may be in part because several student writers repeated this bundle at least five times in their writing, choosing to rely on one bundle rather than use a wider range of bundles. Some examples of its usage from one writer are given below:

> *In the process of* teaching, firstly, the teacher should stimulate students' study interest and motivation with full teaching enthusiasm and vivid teaching method. (Learner, 2014, 11)

> At the same time, students' activity should reflect the teaching objectives *in the process of* practice. (Learner, 2014, 11).

4.2.6 Prepositional phrase + *of* in literature

Within the literature corpora, this structure accounted for 34% of the bundles in the professional corpus, but only 15% of those in the learner corpus. These bundles are listed in Table 18. The professionals used more distinct types and tokens than did the learners, who used five types of bundles, in contrast with professionals use of 31 types.

Table 18. *Prepositional phrase + of* in literature

Professional		Learner	
Bundle	Freq.	Bundle	Freq.
as a figure of	3	as a matter of	6
as a form of	5	**at the end of the**	6
as a kind of	5	in the process of	7
as a means of	3	under the influence of	6
as a result of	2	with the development of	7
as one of the	3		
at the beginning of the	3		
at the center of	4		
at the end of the	6		
by the end of the	4		
for the sake of	2		
in a state of	2		
in one of the	2		
in terms of the	4		
in the absence of	4		
in the case of	5		
in the context of	7		
in the course of	4		
in the face of	8		
in the figure of	2		
in the form of	5		
in the middle of the	3		
in the midst of	7		
in the process of	3		
in the wake of	4		
in the words of	6		
on the basis of	3		
on the part of	3		
On the part of the	3		
over the course of	4		
to the point of	3		
Total Type	31	Type	5
Token	122	Token	32

Similar to the linguistics writing, a pattern emerged following the *in the + N + of*

structure. The professionals used this structure with greater frequency than learners did,

where the latter group only had one bundle that followed this pattern. The words filling

the N slot are displayed in Table 19:

Table 19. Nouns used in the structure *in the + N + of* in literature writing

		Types	Tokens
Professional	absence$_{(4)}$, case$_{(5)}$, context$_{(7)}$, course$_{(4)}$, face$_{(8)}$, figure$_{(2)}$, form$_{(5)}$, middle$_{(3)}$, midst$_{(7)}$, **process**$_{(3)}$, wake$_{(4)}$, words$_{(6)}$	12	58
Learner	**process**$_{(7)}$	1	7

4.2.7. A comparison of *in the +N + of* in applied linguistics and literature

Across the two groups of professional writers, those in literature used this structure with a greater frequency than did those in linguistics, as can be seen in Table 20. Literature writers also employed a greater number of nouns to fill the slot in this structure. There were four bundles shared between the two subject areas. The greater number of nouns utilized in this structure within literature may be due the writers' need to report on and analyze the prose that is the focus of the article.

Her mourning process increasingly gives way to traumatic expressions that reflect Kitty's inability to cope with trauma *in the absence of* partnership. (Literature, 2013, 27).

The Pale King's characters constantly struggle to locate themselves *in the face of* an excess material that they can be sure is not the self, which, in the late twentieth century, often takes the shape of data, information, entertainment, or some cross-section thereof. (Literature, 2012, 26).

In contrast, the more restricted set of nouns used in linguistics appears to be due to more conventionalized forms of reporting information in that subject area. The bundle *in*

the field of was often used to introduce fields in which ideas or tools originated.

Among the different measures used to determine personality traits, the BFI is a

respected and popular instrument used *in the field of* psychology. (Linguistics,

2012, 21).

It extended the investigation of TV shows and presented important implications

for practitioners and researchers *in the field of* English language teaching.

(Linguistics, 2012, 06).

Table 20. Nouns used in the structure *in the + N + of* in professional writing

		Types	Tokens
Literature	absence$_{(4)}$, **case**$_{(5)}$, **context**$_{(7)}$, course$_{(4)}$, face$_{(8)}$, figure$_{(2)}$, **form**$_{(5)}$, middle$_{(3)}$, midst$_{(7)}$, **process**$_{(3)}$, wake$_{(4)}$, words$_{(6)}$	12	58
Linguistics	**case**$_{(9)}$, **context**$_{(11)}$, development$_{(2)}$, field$_{(5)}$, **form**$_{(4)}$, **process**$_{(2)}$	6	34

4.2.8. Summary of structural findings

Overall, there were differences in structural use between professionals in both

fields, as well as within each field. Two structures, *the + N + of the* and *in the + N +of,*

showed great variation both between disciplines and with how they were used within

each discipline. For *the + N + of the,* professional writers in applied linguistics used this

structure with greater frequency; however, the nouns chosen to fill the *N* slot showed

great variation between the two disciplines, where there was only one shared bundle

between the two groups. Learners in both disciplines had a more restricted set of nouns

used to fill the slot, though learners in applied linguistics had a greater variety.

Differences between learner and professional use in applied linguistics may be due to

differences in the types of data being reported. With *in the + N + of,* professional writers

in literature had a greater variety of words used to fill the slot than did the writers in

applied linguistics. Within the fields, professionals in both literature and applied

linguistics used a larger variety of words than did the learners. Learners in both subjects

used *process* to fill the slot, and this was the only noun used in this structure in the

learner literature corpus. With *other noun phrases*, learners had several bundles using the

present tense of the verb, while similar bundles in the professional corpus used the past

tense.

4.3 Functional use of bundles

This section will report the results of functional distribution of bundles following

the categorization described in Chapter 3.

Figure 2 displays the distribution of the different functional bundles between all

four corpora. Overall, participant oriented bundles are the least frequently used, though

they are more common in applied linguistics writing at both levels of proficiency than

they are in literature writing. With respect to text-oriented bundles, professional writing

in both areas use this functional type with greater frequency than the learner writing in

those areas. Meanwhile, learners in these areas have a greater number of research-

oriented bundles than do the professionals.

Functional Use of Bundles (% of Tokens)
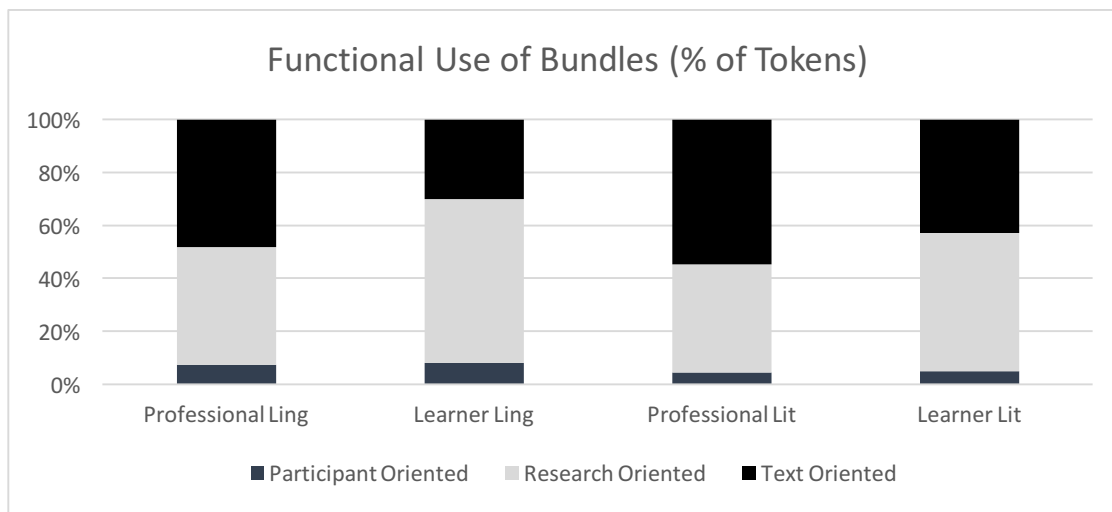
Figure 2. Functional use of bundles in each corpus as a percentage of total tokens

Professional writers in applied linguistics and literature had a similar distribution of functional bundles. Both groups used participant-oriented bundles the least, accounting for less than 10% of all bundles in each sample. The remaining bundles were split between research- and text-oriented.

Professionals in literature used text-oriented bundles with the greatest frequency, where they accounted for about half of all bundles used. In applied linguistics, professionals used both research- and text-oriented bundles with similar frequencies, accounting for around 45% each of total functional bundles.

Within the subject areas, learners behave similarly to the professional writers with respect to participant-oriented bundles. However, learners in applied linguistics use a higher percentage of research-oriented bundles and a lower percentage of text-oriented bundles than do the professionals. The following sections will examine each functional sub-type in depth and discuss possible reasons why this distribution occurs.

4.3.1 Research-oriented bundles in applied linguistics and literature

Research-oriented bundles are used to help structure and report accounts of the research activities and the world they took place in. Within this category, there are five subtypes. Locational bundles report on time and place, procedural bundles are used to recount the research process, quantitative bundles are used to report size or scale, descriptive bundles are used to describe the setting, and topical bundles are used to report the main focuses of the study. The distribution of these subtypes across all corpora are displayed in Table 21. Description bundles were the most frequent in applied linguistics corpora, while location bundles were most frequent in literature corpora.

The literature writers did not use procedure or topic bundles in their writing. The lack of procedure bundles could be because linguistics writing relies on reporting procedures while this is not a component of writing in literature. The reason for the lack of topic bundles seems less clear, but could be due to more varied topics in the literature writing, whereas the linguistics writing dealt with a more restrained group of topics. All writing in these corpora were focused on applied linguistics, a subfield of linguistics focused on discussing the applications of linguistics research to teaching, while the literature corpora had a broader range of topics. A larger literature corpus could yield different results.

Table 21. Distribution of research-oriented bundles across all corpora (% of tokens)

	Linguistics		Literature	
	Learner	Professional	Learner	Professional
Location	6%	22%	65%	53%
Procedure	9%	17%	0%	0%
Quantification	10%	14%	10%	5%
Description	39%	29%	25%	42%
Topic	36%	18%	0%	0%
	100%	100%	100%	100%

Within applied linguistics writing, learners used the description and topic bundles with higher frequency than did the professionals. The higher use of the topic bundles could be explained by the learners' repetition of a few key bundles, such as *task based language teaching*. Among the procedural bundles, professionals used these with greater frequency than did the learners, and used more bundles that explicitly described the study. Professionals may have better command over common formulaic structures used to describe research procedures, which could account for their more frequent use of these bundles. Some examples of these in use are given below.

More specifically, *students were asked to* rate themselves as readers of Spanish. (Professional, 2012, 16).

During the quarter, *students were asked to* watch one cultural news segment a week and discuss in class and during chat sessions what they had learned. (Professional, 2013, 58).

The *participants were asked to* choose an appropriate reading of each kanji on the test. (Professional, 2013, 36).

Within literature writing, learners used a higher percentage of description bundles than did the professional writers; however, the learners only used five different bundles for this function. The professional writers used a wider variety of bundles, twenty-two in total. This disparity could possibly be due to the small size of the learner corpus in this subject area. There were no shared bundles between the two groups.

Most of the descriptive bundles in the professionals writing followed the *the + N+ of the* structure, which was discussed in a previous section. Two of the five bundles in the learner corpus followed this structure, with the remaining three being *falls in love with, advocate the spirit of* and *get rid of the*. For *advocate the spirit of,* there were four tokens total, across four texts, though they were repeated twice, which suggests that students who used this phrase copied it from another source. Two such examples are shown below:

This paper will analyze the image of the old man, reveal the characteristics of the code hero, and *advocate the spirit of* never giving up. (Learner, 2013, 8).

This paper will analyze the image of the old man, reveal the characteristics of hero, and *advocate the spirit of* never giving up. (Learner, 2014, 4).

The issue of plagiarism and how it could affect analysis of lexical bundles in the present data set will be discussed further in Chapter Five.

4.3.2 Text-oriented bundles in applied linguistics and literature

Text-oriented bundles are used to organize the text and contain four subtypes. Transition signals establish additive or contrastive links between elements in the text, resultative signals mark relationships between elements that are causative or inferential; structuring signals organize the text or direct readers elsewhere within the text; framing

signals serve to situate and limit arguments. Table 22 shows the overall distribution of these subtypes across the four corpora.

Table 22. Distribution of text-oriented bundles across all corpora (% of tokens)

	Applied Linguistics		Literature	
	Learner	Professional	Learner	Professional
Transition	25	24	62	31
Resultative	19	25	0	5
Structuring	13	20	0	9
Framing	43	31	38	55
	100	100	100	100

In looking at professional writing between the two subject areas, those writing in literature used framing and transition bundles with higher frequency than did the applied linguistics writers. The writers in linguistics used resultative and structure bundles with higher frequency. These differences could be due to the rhetorical style of each subject area. Writing in applied linguistics may more often discuss links between elements in the text and need to orient reader's attention towards tables and charts, and seems to be generally more formulaic in nature than literature writing. Conversely, writers in literature may need to situate their arguments more, and discuss how the text affects the reader.

For framing bundles, most of these follow the structure of *in the + N +of*, discussed above. There, writers in literature employ a wider range of nouns to fill these slots (see Table 16).

In applied linguistics, many of the resultative bundles use the words *results, shown,* or *found*, which corresponds to reporting or interpreting results, which would be

expected for the subject area. The bundles which use *results* in the literature corpus tend to use these to connect the relationship between actions that occurred in the text being analyzed. Some examples are given below.

The results showed that the acquisition pattern for the English plural morpheme by native Mandarin Chinese-speaking children was similar to English monolingual children in general. (Linguistics, 2012, 17).

Based on *the results of the* questionnaire, 45% of the teachers in his study reported a rather strong influence on their teaching while 31% believed the impact was not strong. (Linguistics, 2013, 31).

Boredom in the Kierkegaardian sense is *the result of the* negative freedom of the aesthete's irony: it follows from the conclusion that there is nothing left in the world with which the aesthete is connected, which is of value to him. (Literature, 2012, 25).

The two vectors of the Gothic help us to see these interpretations not as the excesses of literary criticism, its fashions and follies, but as *the result of a* dualism inherent in theories of the Gothic. (Literature, 2013, 8).

In looking at the differences between learners and professionals within subject areas, not much can be said about the literature writing, since the learner corpus contained so few types of bundles overall. In applied linguistics, learners used framing bundles with greater frequency than did the professional writers. This may be because of differences in frequency of use of select bundles; for example, learners used the bundle *in the process of* 27 times per 100,000 words whereas the professionals used it only 2 times

per 100,000 words.

Learners used two patterns for framing bundles that were not found in the professional corpus. These were *with the (rapid) development* and *in this way (the) students*. The bundles were found most commonly in the abstract or as the opening sentence of the introduction, which suggests that this could be a common pattern for these learners as a way of situating the argument of the paper. Both bundles are exemplified here:

> *With the development of* society and economics, English becomes more and more
> important. (Learner, 2013, 1).
> *With the rapid development* of society and technology, the international contacts
> between different people facilitate the cultural and economic exchange. (Learner,
> 2013, 15).
> *In this way, students'* sensitivity and identification to the distinction between
> Chinese and Western culture can be promoted effectively so that it can cultivate
> the intercultural communicative competence. (Learner, 2013, 13).
> *In this way, students* could be deeply impressed of what they've learnt. (Learner,
> 2013, 9).

4.3.3 Participant-oriented bundles in applied linguistics and literature

Participant-oriented bundles are those that focus on either the writer or reader, within this functional category there are two subtypes. Table 23 shows the distribution of these subtypes across all four corpora. These were relatively uncommon as compared to research-oriented bundles or text-oriented bundles, though they were more frequent in the

applied linguistics corpora than the literature corpora.

Table 23. Distribution of participant-oriented bundles across all corpora

	Applied Linguistics		Literature	
	Learner	Professional	Learner	Professional
Stance	74%	34%	0%	64%
Engagement	26%	66%	100%	36%
	100%	100%	100%	100%

The majority of participant oriented bundles in the applied linguistics learner corpus are stance bundles, while the professional corpus had more engagement bundles. Writing in the learner literature corpus have no stance bundles, while in the professional corpus the distribution of participant-oriented bundles is 66% stance bundles and 36% engagement bundles.

4.3.4 Stance bundles in applied linguistics and literature

This type of bundle expressed the writer's beliefs and attitudes. Across all corpora, most of these bundles follow the structure *it +V/ADJ*, for example *it is important that* (found in all corpora containing stance bundles). The words that can fill the ADJ position vary between proficiency levels, however, as displayed in Table 24. The adjectives in the learner corpus use a wider range of words to express similar meanings, whereas those in the professional corpora tend to each have a distinct meaning. This is shown where the learner corpus includes both *clear* and *obvious*, as well as *essential* and *important*.

Table 24. Adjective contained in *it + V/Adj* structure stance bundles

	It is ...
Learner	**clear**, essential, hoped, **important**, **necessary**, obvious
Professional (ling)	**clear**, **important**, likely, **necessary**, not clear, **possible**
Professional (lit)	**important,** not surprising, **possible**

The bundles shared between learner and professional writing all take an assertive stance; those not shared differ in the weight of their stance, with the exception of *it is hoped that*. Most learner stance bundles take a strong position while professional bundles show a mix of strong assertion and hedging. This is also true in those stance bundles in literature, where there is a mix of strong assertion and hedging.

From the data above, *it is obvious that* non-English major's are well aware of the influencing factors of unsatisfactory English achievement. (Linguistics, Learner 2014, 2)

Nevertheless, *it is likely that* this factor does play some part in learners' difficulties when reading. (Linguistics, Professional, 2013, 40)

It is not surprising, then, that Ask Your Mama features the interaction of African cultures in the Americas and Africa through its evocation of Afro-Caribbean as well as African American music and its movement among different sites of black revolutionary struggle. (Literature, Professional, 2013, 13)

The exception in the learner corpus, *it is hoped that*, was primarily used by students in order to express the intended implications of their papers. Some examples of this are given below:

It is hoped that this paper can give certain guidance to English vocabulary learning and teaching and can offer students advice about vocabulary learning. (Linguistics, Learner, 2013, 7).

It is hoped that the essay can be of great help for English major students and department management system. (Linguistics, Learner, 2014, 6).

Among those found in the learner corpus, there are four bundles that start with *should*, which are *should be given to, should pay attention to the, should pay more attention to,* and *should try their best to*. These bundles were used to recommend courses of action for those reported in the study, for example:

> During the teaching process, teachers *should try their best to* create
>
> communication environment, use authentic materials and design communication
>
> activities. (Linguistics, Learner, 2014, 15).

4.3.5 Engagement bundles in applied linguistics and literature

This subtype is used to engage directly with the readers. Within the linguistic corpora, this type of bundle was more common in the professional corpus, where it accounted for 66% of participant-oriented bundles, while it was only 26% of those in the learner corpus. Among participant-oriented bundles in the literature corpora, this subtype accounted for 36% of the bundles in the professional corpus, and 100% of those within the learner corpus. It should be noted, however, that the learner corpus contained only one token of this type, *we can see that*, that served this function.

The most frequency structure for engagement bundles followed the passive structure *it/as + can be seen*, which occurred in all corpora except the learner literature, where it was manifested as *we can see that*. This type also occurred with frequency (7 times per 100000 words) in the linguistics learner corpus, where it occurred more frequently than did *it can be seen*, which had a normalized frequency of 4 times per 100000 words. Examples of this usage are given below.

> *We can see that* Heathcliff does what he wants, ignoring every moral factors and

laws and ethics. (Literature, Learner, 2014, 5)

Thus *we can see that* it is really high time for the teacher to have an effective application of multimedia and senior English teaching. (Linguistics, Learner, 2013 ,12).

4.3.6 Summary of findings

Variation in functional use of lexical bundles among professionals can be attributed to the needs of the disciplines within which they are writing. Writing in applied linguistics has more procedural bundles than literature writing, since writing in this area is undertaking a research project. Even when bundles have similar structures, the ideas they connect differ; where applied linguistics uses bundles containing *results* to report results, those in literature use this to describe connections occurring in the text being described. Learners use functional bundles that do not occur in the professional corpora, notably bundles containing *development* for framing, and *it is hoped that* for expressing stance. Some frequency counts in learner bundles are due to repetition of a restricted number of bundles, and in at least one case, to plagiarism.

The next chapter will review the main findings of the study and will discuss their implications for teaching.

Chapter 5 – Conclusion

In this chapter I will summarize the main findings of the study, as they answered the research questions. I will then discuss the implications for teaching and curriculum development, and give suggestions for directions for future research. Finally, I will review the limitations of the present study and summarize the benefits of the present research.

5.1 Overview of findings

Applied linguistics writing at both professional and learner level used lexical bundles with greater frequency than did literature writing. This is in part likely due to the formulaic nature of applied linguistics writing, where the use of bundles are tied to specific rhetorical moves of the field, such as reporting results and describing study methodology. Learners in applied linguistics relied on a wider range of bundles than did the professionals, while learners in literature writing used fewer than the professionals. The structural and functional use of lexical bundles also differed between professionals in both fields, and between learners and professionals within each field.

With respect to the first research question, which looked at overall frequency of bundles, the writing in applied linguistics used a wider range of bundles with greater frequency than did literature writing. Within applied linguistics writing, the learners used a wider range of bundles with greater frequency than did the professionals. However, within the literature corpora learners used fewer bundles with less frequency overall than did the professional writers.

The frequently used bundles in applied linguistics learner writing rely heavily on repetition of certain noun phrases, where *task based language learning* occurred with the highest frequency in the learner corpus. As discussed previously, this could be due to learners having learned fewer ways to restate the same idea using fewer words, but it could also be due to the subject focus of the learner papers.

The frequency findings support those found by Chen and Baker (2010), Wei and Lei (2011), and Cortes (2004). Chen and Baker found that Chinese L1 writers, when writing in English, favored repetition of the same bundles. Similarly, Wei and Lei found that Chinese EFL writers used a larger proportion of bundles when compared to a control corpus. Cortes, in comparing student and professional writing in history and biology, found that learners used few of the target bundles identified in the professional fields. Additionally, the student writers tended to rely on repetition of their frequent bundles.

When bundles are shared between learners and professionals in applied linguistics, the two groups use them very differently. In one case, with *is one of the*, professional writers displayed a greater variety of word choices that could follow the bundle. At the same time, words that followed the bundle *to participate in the* were more limited in the professional writing, where most examples were followed with *activit(ies/y)* or *study*. A larger corpus may have shown those to actually be components of five-word lexical bundles; however, the present study examined them as though they were adjacent to four-word bundles.

Within the literature writing, learners used fewer distinct types of bundles overall, but used them with greater frequency than did the professional writers. When comparing

shared bundles between these two groups, learners used those bundles over a similar range as did the professionals, but with much greater frequency.

For the second part of the research question, which was looking at structural types of bundles, the professional writers displayed differences in two structures, and learners used these same structures differently than the professionals in their disciplines.

The structure *in + N + of the* was used less frequently by learners, and with less variation in the nouns chosen to fill the slot. Some of this difference in applied linguistics can be attributed to differences in the types of data being reported on. Professionals used some bundles of this structure to report on statistical analyses, and students did not have a similar use for those bundles within their writing. However, there was disciplinary variation between applied linguistics and literature with respect to the nouns used in this structure, which reflects the stylistic differences between the two disciplines.

Similarly, the structure *in the +N + of* was used less frequently by learners in both fields, and there was a wider variety of nouns used by professionals to complete the structure. As with the previous structure, the nouns chosen varied between the two disciplinary areas. However, in this area, learners in both subjects used *process* with higher frequency than the professionals, where learners in applied linguistics used this structure 27 times per million words. As occurred with frequent bundles, some explanation of this could be that learners have a restricted lexicon and therefore rely on repetition of a few bundles they are familiar with.

These findings support those found in Chen and Baker (2010), where the Chinese L1 writers had a more restricted range of nouns used to fill the noun slot. L1 English

writers at the student level had a slightly wider range of nouns used, but both student writers groups used fewer overall tokens and nouns than did the professional writers.

For the third research question, which looked at how the functional use of lexical bundles varied between professionals and learners within disciplines, differences emerged between how professionals in both disciplines used these bundles, and how learners within each used them. These differences occurred across all functional categories, though overall writers in literature used more text-oriented bundles, and writers in linguistics had a more even split between text- and research-oriented bundles. Participant-oriented bundles were the least frequently used type across all groups, though writers in linguistics used these with greater frequency.

Among research-oriented bundles in applied linguistics, learners favored topic and description bundles, which could be due to their reliance on repetition of certain topic oriented bundles, for example *task based language learning*. Professionals use more procedure bundles, which could be due to their better command of reporting results and procedures. In literature, learners likewise used many description bundles, but used very few distinct types. Some of these bundles appeared because they were repeated across texts via plagiarism. This may have well happened in more learner texts, though only two cases were discovered in the present study.

With text-oriented bundles in professional writing, resultative bundles were used in applied linguistics for the reporting of results, while in literature they were used to link actions that occurred within the text being written about. Learners in both areas used framing bundles with higher frequency than the professionals did, part of which was due

to their frequent use of *in the process of*. Learners in applied linguistics used bundles including the word *development* frequently to frame their argument. Bundles including this word did not occur in the professional corpus.

Participant-oriented bundles, while less common among all corpora, still showed an interesting connection between structure and function. All but learners in literature used the structure *it + V/Adj* for their stance bundles. Professionals in both fields used a mix of assertive and hedging adjectives, while all but one of the learner adjectives displayed an assertive stance. The exception, *it is hoped that* was found to occur near the start of the learner text, where it was used to express the writer's desire for what the reader would take away from their paper.

There are some differences here with respect to previous research. Chen and Baker (2010) found a larger proportion of stance bundles across all proficiency levels, though they did not control for disciplinary variation. However, similar to Wei and Lei (2011), the present study did find that learners used more research-oriented bundles than did the professionals. Additionally, while their study found that learners also used more text-oriented bundles, whereas the present study found that learners used fewer of these overall.

Overall, there was variation between professional writing in both disciplines, and differences in how the learners within each discipline used lexical bundles. These differences can provide insights and opportunities for how to improve the writing curriculum and instruction. These suggestions will be discussed in the following section.

5.2 Implications for teaching and curriculum development

This study highlights the difference in use of lexical bundles between two subject areas and between two proficiency levels. Learners do not seem to be using the bundles effectively, though this is not the only cause for difference in their writing. The results of this study could be used to inform the development of a writing curriculum or lessons to help students bridge the gap between their current writing level and the target. However, more research across more corpora may be needed to make better recommendations for future teaching directions.

An initial suggestion is to focus on disciplinary variation, especially with respect to how the nouns used to fill in the structures *in the + N + of* and the *the + N + of the*. These were areas where there were differences between the professional groups, and where the learners also showed differences from the targets in their fields. Learners showed a restricted range of nouns used to fill the slot in this structure and relied on those forms with greater frequency than did the professionals. For an activity, students writing in each subject area could be exposed to some examples from professional writers in their field and try to infer rules for how these structures are used. They could then complete a guided fill-in-the-blank activity.

Another suggestion is to have the instructor compile a corpus of representative professional writing, take students to computer labs, and guide the students' use of concordancers in order to raise awareness of how specific bundles are used within the discipline. Teachers could also create a cloze-type activity where these bundles are removed from the text and students are asked to replace them in the text. This could be

employed with the previous activity, where first students could focus on noticing, then follow a guided practice worksheet before using them independently.

Learners also tended to have repetitive use of bundles (e.g. *task based language teaching)* where professionals employed a wider range of bundles or used pro-forms. The repetition made learners' writing less effective. Teachers could therefore also focus on helping students not overuse specific bundles. For example, they could teach cohesive devices (such as pronouns) in writing that can take the place of a bundle. Teachers could display a sample student text with repetitive use of one such bundle and guide students through replacing those bundles with typical cohesive devices. Students could then be given a similar text to correct on their own.

With respect to verb tense, some bundles in the learner corpus showed use of the present tense while similar bundles in the professional corpus used the past tense. It would also be useful for learners to take professional writing and examining the tenses of verbs in each section, and perhaps formulating rules for their usage based on this activity. Similarly, students used stance bundles differently than did the professionals, where learners in literature did not use them at all. Materials could be developed to show students how professionals frame their arguments, and give them a chance to correct errors in other students' writing. An example of such an activity is given in Figure 3.

Figure 3. Sample activity for teaching

Reporting results:
These structures are used to report results that are not certain. You can use these structures for results where you are uncertain about the outcomes of the applications.

Yet with technologies that allow many-to-many simultaneous or near simultaneous writing, greater accountability for participation in the writing process, and improved document sharing, *it is likely that* the number of projects involving collaborative writing with three or more writers will grow in the future.

As discussed before, *it is likely that* the complete picture of the difficulties caused by words used in unfamiliar ways is not shown here due to the methodology used.

However, *it is possible that* social support that is provided from mentors or colleagues may come at a price.

It is possible that novice ESOL teachers may face greater challenges as they work with students from varied cultural backgrounds and cultures of schooling.

Re-write these sentences, using one of the phrases from above, correct other errors as needed.
1. What is more, they will very probably be helpful in developing students' positive attitudes toward learning English.
2. If Chinese students do not know the cultural differences, they may have incomprehension and misunderstanding in cross-cultural communication, even make a stupid mistake and produce culture shock.
3. Students should also cultivate positive attitude towards English learning, sustain motivation and confidence, and control anxiety in communication activities, which, in turn, may promote their willingness to communicate in English learning.

Teachers could also emphasize which bundles are used for particular functions. So for example, students used fewer structuring bundles than did the professionals, and teachers could create lessons that emphasize the bundles that serve this function and demonstrate their proper use. Such an activity could be done for any type of functional bundle. Teachers could choose to either present bundles typical of each function, or they

could guide students in discovering these bundles based on descriptions presented to them.

Overall, noticing how lexical bundles are used in writing is important for learners if they are to understand how to use them better themselves. This can be promoted through encouraging more reading of studies in the field that students are writing in, both independently and through guided class activities. If students are able to read and analyze the language used in studies, this may guide them in acquiring proficient use of this language in their own writing.

Additionally, the results of this study can be used to inform the teaching writing component of Peace Corps China pre-service training, especially since writing a thesis is a requirement of all English majors and many volunteers are asked to teach writing courses. Here the training could include a brief introduction to lexical bundles and introduce the trainees to free corpora and concordancers that they could use to inform their lesson preparation.

5.3 Limitations and suggestions for further research

While there are many ideas for how this research could inform classroom teaching and the curriculum, it should be noted that the present study does not provide a comprehensive view of how bundles are used. The results of this study represent what was found in the current corpora, and the results found should not be held as absolute truths. In the following section, I will make suggestions for how future researchers can further our understanding of the use of lexical bundles and how their results can also better inform teaching practices.

While this study examined lexical bundles in applied linguistics and literature writing and two different proficiency levels, it is important to remind the reader of several points. First, this study is not meant to be representative of all Chinese EFL writing, merely a selection of what was produced at one university during a limited time frame. Because the learner corpus was only collected from one university, the results of this study do not describe structural type and functional use of lexical bundles for all Chinese English learners. Additionally, the learner corpora may not be totally representative of the lexical bundles used across all theses in the department because of their limited size.

Further research is needed to both examine these subject areas of writing at universities elsewhere in China, and to collect a wider range of data. Secondly, while lexical bundles have been identified as being particularly important in the production of academic writing, expert use of these alone does not guarantee a proficient essay. Other research on characteristics of Chinese EFL writing needs to be undertaken, and it is important to consider the interaction of many different factors. While the suggestions laid out in this paper will prove beneficial for future or practicing teachers in China, they should not be taken in isolation.

There has also been recent research looking at how lexical bundles are used rhetorically in different sections of academic writing. A greater understanding of that could lead to improved instructional benefits, because it could equip learners and educators with the knowledge of discipline-specific bundles and how they are used, and also of which bundles occur in which sections of the text and how they are used there.

5.5 Conclusion

Despite the need for more research, the results of this study are the first information about what the intermediate language learners at this university are doing in their writing and how it compares to professional writing within the same fields. The study is a valuable contribution to our knowledge of second language writing and student needs. Chinese English learners have not been extensively studied, and examining the writing of these learners can help us understand what learners are able to do with their present level of instruction. In addition, while the primary goal of this study was to describe the differences between professional and learner writing, the results had applications for enhancing the writing curriculum at Longdong University. Finally, while this study could in no way provide a comprehensive description of lexical bundles in these areas, I hope that future research will take up on the suggestions given above in order to further our understanding of disciplinary variation of lexical bundles.

References

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*(2), 81-92.

Ädel, A. & Römer, U. (2012). Research on advanced student writing across disciplines and levels: Introducing the Michigan Corpus of Upper-level Student Papers. *International Journal of Corpus Linguistics, 17*(1), 3-34.

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Bannard, C. & Lieven, E. (2012). Formulaic language in L1 acquisition. *Annual Review of Applied Linguistics, 32*, (1), 3-16.

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

Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286.

Biber, D., Conrad, S., & Leech, G. (2002). *Longman student grammar of spoken and written English.* Harlow, England: Longman.

Biber, D., S. Conrad, R. Reppen, P. Byrd, and M. Helt. 2002. Speaking and writing in the university: A multi-dimensional comparison. *TESOL Quarterly* 36.9-48.

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

Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes, 25*(3), 310-331.

Chen, C. (2006). The use of conjunctive adverbials in the academic papers of advanced Taiwanese EFL learners. *International Journal of Corpus Linguistics, 11*(1), 113-130.

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

Conrad, S.M. (1999). The importance of corpus-based research for language teachers. *System, 27*(1), 1-18.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 39-423.

Coxhead, A. & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing, 16*(3), 129-147.

Ellis, N.C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 24,* 143-188.

Friginal, E., Li, M., & Weigle, S. (2014). Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing, 23*(1), 1-16.

Hinkel, E. (2003). Simplicity with elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly, 37*(2), 275-301.

Hinkel, E. (2004). Tense, aspect and the passive voice in L1 and L2 academic texts. *Language Teaching Research, 8*(1), 5-29.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.

Laufer, B. & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*(6), 647-672.

Lee, D. & Chen, S. (2009). Making a bigger deal of the smaller words: Function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing, 18*(3), 281-296.

Leedham, M. & Cai, G. (2013). Besides... on the other hand: Using a corpus approach to explore the influence of teaching materials on Chinese students' use of linking adverbials. *Journal of Second Language Writing, 22*(4), 374-389.

Liu, D. (2008). Linking adverbials: An across-register corpus study and its implications. *International Journal of Corpus Linguistics, 13*(4), 491-518.

Martinez, I. (2005). Native and non-native writers' use of first person pronouns in the different sections of biology research articles in English. *Journal of Second Language Writing, 14*(3), 174-190.

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

Schmidt, R.W. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 137-174). Rowley, MA: Newbury House.

Wood, D. (2015). *Fundamentals of formulaic language*. London: Bloomsbury Academic.

Wray, A. (2000). Formulaic sequences in second language teaching: Principles and practice. *Applied Linguistics, 21*(4), 463-489.

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

Waiguoyu Xueyuan Jianjie [Brief introduction to the school of foreign languages and literatures]**.** (2015, September 1). Retrieved October 27, 2015. Retrieved from: http://sfll.ldxy.edu.cn/introduction.htm

Wei, Y. & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC Journal, 42*(2), 155-166.

Appendix: Complete list of bundles found in the study

Most frequent bundles in literature

Professional				Learner			
Bundle	RF*	NF	R	Bundle	RF	NF	R
at the same time	**38**	**15**	**24**	**at the same time**	17	24	13
in the united states	33	**13**	7	falls in love with	7	10	6
as well as the	**20**	**8**	**13**	in the novel the	7	10	5
in the face of	20	**8**	11	At the same time she	7	10	4
in the first place	19	7	10	we can see that	7	10	5
on the one hand	19	7	11	is one of the	6	8	5
in the midst of	18	7	8	all over the world	6	8	4
the figure of the	18	7	11	Plays an important role in	6	8	5
in the context of	17	7	13	get rid of the	6	8	4
at the end of the	**16**	**6**	**9**	**on the other hand**	6	8	5
the end of the	16	6	9	**one of the most**	6	8	4
in the words of	14	6	10	in the 19th century	5	7	5
as a form of	13	5	10	**as well as the**	5	7	4
in the form of	13	5	11	**for the first time**	5	7	4
the story of the	13	5	7	in the nineteenth century	5	7	5
as a kind of	12	5	7	**in the process of**	5	7	4
in a way that	12	5	8	the heroine of the	5	7	5
in the case of	12	5	8	**At the same time the**	5	7	4
in this way the	12	5	7	with the development of	5	7	5
on the other hand	**12**	**5**	**8**	**at the end of the**	4	6	4
one of the most	**12**	**5**	**10**	the beginning of the novel	4	6	4
the rest of the	12	5	6	advocate the spirit of	4	6	4
it is possible to	11	4	5	as a matter of	4	6	4
over the course of	11	4	8	for a long time	4	6	4
by the end of the	10	4	7	the characters in the	4	6	4
in the wake of	10	4	7	under the influence of	4	6	4
of the nineteenth century	10	4	9				
the history of the	10	4	7				
the ways in which	10	4	7				
at the center of	9	4	8				
in such a way that	9	4	6				
in terms of the	9	4	7				
in the absence of	9	4	6				
in the course of	9	4	8				
in this essay I	9	4	9				
of the twentieth century	9	4	5				
the materiality of the	9	4	4				
the world of the	9	4	7				
turns out to be	9	4	7				
as a means of	8	3	6				

Professional			
Bundle	RF*	NF	R
as one of the	8	**3**	7
at the beginning of the	8	**3**	5
At the same time the	**8**	**3**	**7**
in the middle of the	8	**3**	6
in the novel the	**8**	**3**	**5**
in the preface to	8	**3**	5
of the novel as	8	**3**	7
on the part of	8	**3**	6
that there is a	8	**3**	5
that there is no	8	**3**	6
the extent to which	8	**3**	4
the novel as a	8	**3**	7
the role of the	8	**3**	7
to make sense of	8	**3**	7
while at the same time	8	**3**	4
a world in which	7	**3**	5
as a figure of	7	**3**	5
as well as to	7	**3**	4
can no longer be	7	**3**	5
in relation to the	7	**3**	5
in response to the	7	**3**	6
in the process of	**7**	**3**	**4**
in the same way	7	**3**	7
it is not surprising	7	**3**	6
of the novel in	7	**3**	6
on the basis of	7	**3**	6
On the part of the	7	**3**	5
our understanding of the	7	**3**	6
that it is not	7	**3**	6
the words of the	7	**3**	6
to the point of	7	**3**	4
what he calls the	7	**3**	5
as a result of	6	**2**	4
as well as a	6	**2**	5
at a time when	6	**2**	5
can be seen as	6	**2**	5
for the first time	**6**	**2**	**4**
for the sake of	6	**2**	5
I have tried to	6	**2**	4
in a letter to	6	**2**	4
in a state of	6	**2**	4
in one of the	6	**2**	6
in the figure of	6	**2**	6
is that it is	6	**2**	4
it is important to	6	**2**	5

Professional			
Bundle	RF*	NF	R
of the fact that	6	**2**	5
of the novel and	6	**2**	4
of the novel the	6	**2**	5
out to be a	6	**2**	5
takes the form of	6	**2**	4
the early nineteenth century	6	**2**	5
the emergence of the	6	**2**	6
the form of a	6	**2**	6
the loss of the	6	**2**	4
the perspective of the	6	**2**	5
the result of the	6	**2**	4
what it means to	6	**2**	5
which is to say	6	**2**	5

* RF = raw frequency, NF = frequency normed to per 100,000 words, and R= range (number of texts)

Most Frequent Bundles in Applied Linguistics

Professional				Learner			
Bundle	RF*	NF	R	Bundle	RF	NF	R
as well as the	58	15	32	task based language teaching	91	66	4
on the other hand	50	13	22	**in the process of**	37	27	15
the extent to which	49	12	15	**at the same time**	24	17	13
the results of the	46	12	23	**as well as the**	23	17	12
in the context of	45	11	20	that is to say	20	14	13
in the present study	42	11	20	in junior middle school	19	14	6
in the united states	40	10	19	**is one of the**	16	12	10
in the current study	37	9	16	**on the other hand**	16	12	12
at the same time	36	9	19	is a kind of	15	11	10
there was a significant	31	8	10	to communicate in English	15	11	4
in the target language	30	8	13	between teachers and students	14	10	7
of the present study	30	8	14	**of the target language**	14	10	9
the fact that the	27	7	22	students interest in learning	14	10	5
as a result of	26	7	17	plays an important role in	13	9	7
it is important to	26	7	17	in foreign language learning	12	9	6
in terms of the	24	6	17	in order to make	12	9	10
at the end of the	23	6	14	**in the field of**	12	9	10
At the beginning of the	22	6	10	junior middle school students	12	9	4
the use of the	22	6	13	**on the basis of**	12	9	9
in the field of	21	5	14	the background of the study	12	9	11
the results of this study	21	5	12	at home and abroad	11	8	5
as shown in table	20	5	12	**it is necessary to**	11	8	7
the nature of the	20	5	13	**on the one hand**	11	8	9
in the case of	19	5	12	pay more attention to	11	8	7
of the current study	19	5	11	**the purpose of the**	11	8	5
in the sense that	18	5	9	the relationship between the	11	8	6
it is possible that	18	5	11	the significance of the study	11	8	10
on the one hand	18	5	11	**to participate in the**	11	8	10
the participants in the	18	5	12	to use the language	11	8	7
in addition to the	17	4	14	At the same time the	10	7	7
the present study was	17	4	12	is based on the	10	7	8
in other words the	16	4	11	it is obvious that	10	7	8
It is important to note that	16	4	9	language teaching and learning	10	7	6
one of the most	16	4	15	most of the students	10	7	8
so that they can	16	4	9	table shows that students	10	7	4
the content of the	16	4	10	the development of the	10	7	9
the degree to which	16	4	8	what they have learned	10	7	7
the members of the	16	4	7	as a result the	9	7	8
to be able to	16	4	13	English majors in Longdong university	9	7	4
used in this study	16	4	12	in the learning process	9	7	7
English as a foreign language (EFL)	15	4	12	middle school English teaching	9	7	6
In the case of the	15	4	10	of students think that	9	7	5

Professional				Learner			
Bundle	RF	NF	R	Bundle	RF	NF	R
that the use of	15	4	9	to a large extent	9	7	4
The participants in this study	15	4	8	to find out the	9	7	6
the total number of	15	4	7	to take part in	9	7	6
a great deal of	14	4	12	we can see that	9	7	5
in a variety of	14	4	11	**at the beginning of the**	8	6	7
in the form of	14	4	13	English teaching and learning	8	6	5
in the same way	14	4	13	for the students to	8	6	7
it should be noted that	14	4	8	foreign language teaching and	8	6	6
of the target language	14	4	10	in order to improve	8	6	7
students were asked to	14	4	8	**in terms of the**	8	6	6
the effects of the	14	4	8	in the process of learning	8	6	6
the number of words	14	4	7	in this way the	8	6	8
the purpose of the	14	4	9	only in this way	8	6	5
the quality of the	14	4	8	pay attention to the	8	6	7
with regard to the	14	4	10	pay much attention to	8	6	5
as a foreign language	13	3	8	**so that they can**	8	6	4
At the time of the	13	3	6	**the results of the**	8	6	8
can be found in	13	3	9	with the help of	8	6	6
English as a second language	13	3	10	a large amount of	7	5	6
for the purposes of	13	3	7	an important role in	7	5	5
in this study the	13	3	11	can be divided into	7	5	6
it is likely that	13	3	9	**English as a foreign language**	7	5	6
native speakers of English	13	3	8	for a long time	7	5	7
of the variance in	13	3	4	in foreign language teaching	7	5	5
of this study was	13	3	10	in the course of	7	5	6
On the basis of the	13	3	8	interest in learning English	7	5	4
previous studies e g	13	3	8	interest is the best teacher	7	5	4
the end of the	13	3	9	it is necessary for	7	5	5
the following research questions	13	3	13	learning a foreign language	7	5	4
the interpretation of the	13	3	7	**On the other hand the**	7	5	6
the meaning of the	13	3	7	process of learning English	7	5	4
a one way anova	12	3	4	students interest in English	7	5	4
a wide range of	12	3	10	the target language and	7	5	5
as part of the	12	3	10	**the use of the**	7	5	6
English as a foreign language	12	3	10	their interest in English	7	5	4
for each of the	12	3	8	to develop the students	7	5	4
in this study we	12	3	8	**to make use of**	7	5	5
in this study were	12	3	9	**to the development of**	7	5	7
it is possible to	12	3	9	to the study of	7	5	6
on the basis of	12	3	10	with the development of	7	5	6
the findings of the	12	3	10	a very important role	6	4	4
the majority of the	12	3	10	and the use of	6	4	5

Professional				Learner			
Bundle	RF	NF	R	Bundle	RF	NF	R
the results of a	12	3	9	can be used to	6	4	6
the students in the	12	3	5	communicate with each other	6	4	5
through the use of	12	3	8	for students to learn	6	4	4
as well as in	11	3	9	for the teacher to	6	4	4
as well as their	11	3	9	In the process of English	6	4	6
at the beginning of	11	3	6	in this way students	6	4	6
been found to be	11	3	8	interested in English learning	6	4	4
in the control group	11	3	5	it is essential to	6	4	5
In the present study the	11	3	9	it is found that	6	4	5
is in line with	11	3	9	it is important for	6	4	4
language learning and teaching	11	3	5	**it is important to**	6	4	5
of teaching and learning	11	3	6	make full use of	6	4	6
participants were asked to	11	3	7	of foreign language learning	6	4	6
the first research question	11	3	8	of foreign language teaching	6	4	5
the rest of the	11	3	7	of learning English and	6	4	4
the results showed that	11	3	7	**On the basis of the**	6	4	6
a better understanding of	10	3	9	should pay attention to the	6	4	4
a small number of	10	3	7	so that students can	6	4	5
accounted for of the	10	3	6	the foreign language teaching	6	4	4
as can be seen	10	3	6	the learning process and	6	4	6
As can be seen in	10	3	6	The purpose of the study	6	4	6
as compared to the	10	3	6	the students interest in learning	6	4	4
at a university in	10	3	8	the teacher and the	6	4	6
be attributed to the	10	3	8	to stimulate students interest	6	4	4
Due to the fact that	10	3	6	when it comes to	6	4	4
in a number of	10	3	6	and the ability of	5	4	4
in terms of their	10	3	5	**At the end of the**	5	4	4
in the language classroom	10	3	8	be taken into consideration	5	4	4
of this study is	10	3	9	closely related to the	5	4	4
on the acquisition of	10	3	6	do not have enough	5	4	4
on the part of	10	3	8	for the learners to	5	4	4
research is needed to	10	3	9	for the purpose of	5	4	5
Significant difference between the two	10	3	8	improve the quality of	5	4	4
the present study is	10	3	7	in accordance with the	5	4	4
the second research question	10	3	8	In the process of teaching	5	4	4
was found to be	10	3	5	in the teaching practice	5	4	4
were randomly assigned to	10	3	8	In this way the students	5	4	5
are shown in table	9	2	6	Is one of the most important	5	4	4
as a means to	9	2	4	**it can be seen**	5	4	4
as part of a	9	2	8	**it is clear that**	5	4	5
as shown in the	9	2	6	it refers to the	5	4	5
as well as a	9	2	6	language teaching is to	5	4	5

Professional				Learner			
Bundle	RF	NF	R	Bundle	RF	NF	R
as well as to	9	2	7	listening speaking reading and writing	5	4	5
at the beginning and	9	2	4	listening speaking reading writing	5	4	5
et al found that	9	2	7	more than half of	5	4	5
for the present study	9	2	7	of a second language	5	4	5
in light of the	9	2	8	of learning English is	5	4	5
in the next section	9	2	8	of second language acquisition	5	4	4
in the process of	9	2	9	of the relationship between	5	4	4
is one of the	9	2	6	of the use of	5	4	5
it can be seen	9	2	5	On the other hand it	5	4	5
it is clear that	9	2	8	Pay more attention to the	5	4	4
it was found that	9	2	7	procedures and data collection	5	4	5
participated in this study	9	2	9	promote the development of	5	4	4
second language (L2) learners	9	2	9	should be given to	5	4	5
significant differences between the	9	2	8	should be regarded as	5	4	4
the effect of the	9	2	7	should pay more attention to	5	4	5
The extent to which the	9	2	8	Should try their best to	5	4	4
the focus of the	9	2	7	Should try their best to	5	4	4
there was no significant	9	2	5	students are able to	5	4	5
they were asked to	9	2	7	students to use the	5	4	4
to answer research question	9	2	5	the culture of the	5	4	4
to participate in the	9	2	8	the first part is	5	4	5
to the development of	9	2	9	the process of learning	5	4	5
with the use of	9	2	5	the students English learning	5	4	4
a significant difference between	8	2	6	the students in the	5	4	4
a two way anova	8	2	4	to a certain extent	5	4	5
are in line with	8	2	6	to focus on the	5	4	5
are likely to be	8	2	6	to learn English well	5	4	4
are more likely to	8	2	5	to pay attention to	5	4	5
as measured by the	8	2	4	under the influence of	5	4	4
as part of their	8	2	8	with the rapid development of	5	4	5
At the time of the study	8	2	5	**a better understanding of**	4	3	4
been shown to be	8	2	6	a great number of	4	3	4
Differences between L1 and L2	8	2	4	and the influence of	4	3	4
examined the effects of	8	2	5	as a kind of	4	3	4
in Hong Kong the	8	2	5	as one of the	4	3	4
in such a way	8	2	7	at the end of	4	3	4
in the development of	8	2	6	both at home and abroad	4	3	4
it is necessary to	8	2	5	by more and more	4	3	4
it is not clear	8	2	6	change the traditional teaching	4	3	4
of the fact that	8	2	7	do not know the	4	3	4
of the process of	8	2	4	English learning it is	4	3	4

Professional				Learner			
Bundle	RF	NF	R	Bundle	RF	NF	R
of whether or not	8	2	5	foreign language in the	4	3	4
On the other hand the	8	2	7	half of the students	4	3	4
students were able to	8	2	5	have more opportunities to	4	3	4
students were required to	8	2	4	in English teaching and	4	3	4
the average number of	8	2	5	in language learning and	4	3	4
the descriptive statistics for	8	2	6	in longdong university as	4	3	4
the effectiveness of the	8	2	6	in order to attain	4	3	4
the impact of the	8	2	5	in this paper the	4	3	4
the purpose of this	8	2	8	in which the language	4	3	4
the second part of	8	2	6	is helpful to improve	4	3	4
these results suggest that	8	2	7	is related to the	4	3	4
they were able to	8	2	6	is the process of	4	3	4
to ensure that the	8	2	8	it is hoped that	4	3	4
to make use of	8	2	5	level of the students	4	3	4
to the target language	8	2	4	longdong university as its	4	3	4
was one of the	8	2	7	motivation to learn English	4	3	4
you do not know	8	2	4	of English teaching and	4	3	4
				of English teaching in	4	3	4
				of language learning is	4	3	4
				of teachers and students	4	3	4
				of teaching and learning	4	3	4
				of the language and	4	3	4
				on students English learning	4	3	4
				pointed out that the	4	3	4
				points out that the	4	3	4
				should be taken into	4	3	4
				so it is necessary	4	3	4
				solve the problems in	4	3	4
				students ability to use	4	3	4
				students are required to	4	3	4
				students to improve their	4	3	4
				teaching and learning in	4	3	4
				teaching and learning is	4	3	4
				teaching and learning the	4	3	4
				that the students are	4	3	4
				that the use of	4	3	4
				the content of the	4	3	4
				the cultivation of students	4	3	4
				the development of language	4	3	4
				the following questions what	4	3	4
				the importance of the	4	3	4
				the influence of the	4	3	4
				the new curriculum standards	4	3	4
				the quality of the	4	3	4
				the questionnaire is designed	4	3	4

| | Learner | | | |
	Bundle	RF	NF	R
	the role of the	4	3	4
	the teachers and students	4	3	4
	the traditional English teaching	4	3	4
	the ultimate goal of	4	3	4
	to answer the question	4	3	4
	to be able to	4	3	4
	to deal with the	4	3	4
	to learn English in	4	3	4
	to the use of	4	3	4
	to use English to	4	3	4

*RF = raw frequency, NF = frequency normed to per 100,000 words, and R= range (number of texts)