

**Lexical Bundles and Disciplinary Variation in
Master Theses: The Case of Native versus Non-
native Corpora**

Sonia Jahangirian

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Arts
in
English Language Teaching

Eastern Mediterranean University
January 2019
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Assoc. Prof. Dr. Ali Hakan Ulusoy
Acting Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Arts in English Language Teaching.

Assoc. Prof. Dr. Javanshir Shibliyev
Chair, Department of Foreign Language
Education

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Arts in English Language Teaching.

Prof. Dr. Ülker Vancı Osam
Supervisor

Examining Committee

1. Prof. Dr. Ülker Vancı Osam

2. Asst. Prof. Dr. Fatoş Erozan

3. Asst. Prof. Dr. Eda Işık Taş

ABSTRACT

Writing academically as a formal skill is an essential requirement that graduate students are expected to acquire; however, the mastery of academic writing skill, especially in EFL contexts, is not an easy task for the graduate students. To be able to write effectively in English, the graduate students are required to acquire the academic discourse of their field of study which is usually a distinct genre with its own special discursal and linguistic elements.

Lexical bundles, as a certain linguistic structure, have been reported to constitute a large percentage of written academic text. Several studies have found that learners experience difficulties while using bundles by misusing, overusing or underusing these bundles in their writing (Adel & Erman, 2012; Hyland, 2008b; Wei & Lei, 2011). Nevertheless, although extensive studies have been done on the use of lexical bundles in various fields, there is still paucity of research into the use of lexical bundles across various disciplines especially in the Iranian context.

Therefore, by adopting a corpus-based analysis approach, this study aimed to explore the frequency, structure and functions of lexical bundles across four different fields of study (business and tourism as soft sciences, and mechanical engineering and civil engineering as hard sciences) used by native English speakers and Iranian EFL master students. For this purpose, 120 theses from two corpora of master theses were selected. The first corpus contained 60 theses written in English by Iranian graduate students who studied at Eastern Mediterranean University (EMU), Famagusta, North Cyprus, and the other corpus included 60 theses written by native writers of English at three

American universities: California State University, Iowa State University and University of Nevada.

The findings of the study revealed that in soft sciences (business and tourism) and in one of the hard sciences (mechanical engineering), Iranian writers used more lexical bundles than their native English counterparts. Moreover, structural and functional differences were reported in each sub-corpus. This study can have several implications for writers of different disciplines to become familiar with different conventions and norms governing different fields of study.

Keywords: lexical bundles, corpus, academic writing, master theses, Iranian EFL learners, native English writers

ÖZ

Akademik bir beceri olan İngilizce yazma becerisi, lisansüstü öğrencilerin edinmeleri gereken temel bir beceridir. Ancak, özellikle İngilizce'nin yabancı dil olarak kullanıldığı ortamlarda akademik yazma becerisinde ustalık kazanmak, lisansüstü öğrencileri için hiç de kolay bir iş değildir. Lisansüstü öğrencilerin İngilizce dilinde başarılı bir şekilde yazabilmeleri için, çalıştıkları alana has söylem ve dilbilimsel özelliklere sahip farklı akademik söylemleri de edinmeleri gerekir.

Bir akademik metnin büyük kısmı sözcük öbeklerinden oluşmaktadır. Birçok araştırmada öğrencilerin sözcük öbeklerini kullanırken zorluk çektikleri, bu nedenle İngilizce dilinde yazarken sözcük öbeklerini yanlış kullandıkları veya gereğinden fazla ya da az kullandıkları ortaya konmuştur. Çeşitli alanlarda sözcük öbeklerinin kullanımı konusunda kapsamlı çalışmalar yapılmış olmasına rağmen, özellikle İran bağlamında farklı disiplinlerde İngilizce yazılmış akademik metinlerde sözcük öbeklerinin kullanımı konusundaki araştırmaların halen yetersiz olduğu söylenebilir.

Bu nedenle, bu çalışmada derlem tabanlı analiz yaklaşımı kullanılarak, anadili İngilizce olan yüksek lisans öğrencileri ile İngilizceyi yabancı dil olarak öğrenmiş İranlı yüksek lisans öğrencileri tarafından İngilizce dilinde yazılmış yüksek lisans tezlerindeki sözcük öbeklerinin sıklığı, yapısı ve işlevleri karşılaştırılmıştır. Bu amaçla sosyal bilimler (işletme ve turizm) ve fen bilimleri (makine mühendisliği ve inşaat mühendisliği) alanından toplam 120 tezden oluşan bir derlem oluşturulmuştur. Bu tezlerden 60 tanesi Doğu Akdeniz Üniversitesi'nde (DAÜ) okuyan İranlı yüksek lisans öğrencilerine ait tezler olup, diğer 60 tanesi ise California Eyalet Üniversitesi, Iowa

Eyaleti Üniversitesi ve Nevada Üniversitesi'nden olmak üzere üç Amerikan üniversitesinde anadili İngilizce olan kişiler tarafından yazılmış olan yüksek lisans tezleridir.

Çalışmanın bulguları, İngilizceyi yabancı dil olarak kullanan İranlı yüksek lisans öğrencilerinin sosyal bilimlerin her iki alt alanı (işletme ve turizm) ile fen bilimlerin bir alt alanında (makine mühendisliği alanında) yazdıkları tezlerde, anadili İngilizce olan yüksek lisans öğrencilerine oranla daha fazla sözcük demeti kullandıklarını ortaya koymuştur. İki gruba ait alt derlemlerde, yapısal ve fonksiyonel farklılıklar da görülmüştür.

Bu çalışmada elde edilen bulgular, farklı bilim alanlarında akademik yazı yazacak olan kişilerin, o alanlarda sık kullanılan sözcük öbeklerini öğrenmelerine yardımcı olacak ipuçları içermektedir.

Anahtar kelimeler: sözcük öbekleri, derlem, akademik yazma, yüksek lisans tezleri, anadili İngilizce olmayan İranlı öğrenciler, anadili İngilizce olan öğrenciler

ACKNOWLEDGEMENT

First, I would like to deeply appreciate my thesis supervisor, Prof. Dr. Ülker Vancı Osam for all the great support, guidance and insight that I have received from her throughout my thesis writing.

In addition, I would also like to thank Asst. Prof. Dr. Fatoş Erozan, and Asst. Prof. Dr. Eda Işık Taş for their invaluable and important feedback on the thesis.

Finally, I sincerely thank my husband for his help throughout the data collection and data analysis.

TABLE OF CONTENT

ABSTRACT.....	iii
ÖZ.....	v
ACKNOWLEDGEMENT	vii
LIST OF TABLES.....	xi
LIST OF FIGURES	xiii
1 INTRODUCTION	1
1.1 Background to the Study.....	1
1.2 Statement of the Problem.....	3
1.3 Purpose of the Study.....	4
1.4 Significance of the Study.....	4
1.5 Definition of Key Terms.....	5
1.5.1 Lexical Bundles	5
1.5.2 Corpus Linguistics and Corpora.....	5
1.5.3 Discourse and Academic Discourse	6
1.6 Summary	6
2 REVIEW OF LITERATURE.....	7
2.1 Formulaic Language	7
2.2 The Concept of Lexical Bundles	8
2.3 Taxonomies of Lexical Bundles.....	10
2.3.1 Biber et al.'s (1999) Structural Taxonomy of Lexical Bundles	11
2.3.2 Biber and Barbieri's (2007) Functional Taxonomy of Lexical Bundles ..	12
2.3.3 Hyland's (2008a) Functional Taxonomy of Lexical Bundles	14
2.4 Previous Studies on Lexical Bundles	15

2.4.1	Frequency-based Analysis.....	16
2.4.2	Structural Analysis.....	19
2.4.2.1	Studies on Structural Analysis	19
2.4.3	Functional Analysis	20
2.4.3.1	Studies on Functional Analysis.....	21
2.5	Summary	22
3	METHODOLOGY	23
3.1	Research Design.....	23
3.2	Corpus Building Procedure.....	23
3.2.1	The Nonnative Speaker Corpus (NNSC).....	26
3.2.2	The Native Speaker Corpus (NSC).....	26
3.3	Bundle Identification	27
3.4	Concordancing Software: AntConc.....	29
3.5	Summary	31
4	RESULTS.....	32
4.1	Bundle Frequency in the Corpora	32
4.2	Structural Categorization of Lexical Bundles	37
4.2.1	Structural Comparison of Lexical Bundles	37
4.3	Functional Categorization of Lexical Bundles.....	43
4.3.1	Functional Analysis of Lexical Bundles	43
4.4	Summary	48
5	DISCUSSION AND CONCLUSION.....	50
5.1	Discussion	50
5.1.1	Frequency of Bundles	51
5.1.2	Structural Analysis of Lexical Bundles	52

5.1.3 Functional Analysis of Lexical Bundles	53
5.2 Conclusion	54
5.3 Implications.....	56
5.4 Limitations	57
5.5 Suggestions for Further Studies	58
5.6 Summary	59
REFERENCES	60
APPENDICES	70
Appendix A: Structural Classification of Lexical Bundles	71
Appendix B: Functional Classification of Lexical Bundles	76

LIST OF TABLES

Table 2.1: Biber et al.'s (1999) Structural Classification of Lexical Bundles	12
Table 2.2: Biber and Barbieri's (2007) Functional Classification of Lexical Bundles.	13
Table 2.3: Hyland's (2008a) Functional Taxonomy of Lexical Bundles	15
Table 3.1: Words Counts and Average Word Length of the Sections of the Two Corpora.....	25
Table 3.2: Frequency and Range of the Sub-sections of the Two Corpora	28
Table 4.1: The Frequency of Lexical Bundles Across the Four Sub-corpora	32
Table 4.2: Top 30 Frequent 4-word Bundles in Business Sub-corpus in Rank Order..	33
Table 4.3: Top 30 Frequent 4-word Bundles in Tourism Sub-corpus in Rank Order...	34
Table 4.4: Top 30 Frequent 4-word Bundles in Civil Engineering Sub-corpus in Rank Order.....	35
Table 4.5: Top 30 Frequent 4-word Bundles in Mechanical Engineering Sub-corpus in Rank Order.....	36
Table 4.6: Structural Comparison of Bundles in Business Master Theses of Native English and Iranian Students.....	38
Table 4.7: Structural Comparison of Bundles in Tourism Master Theses of Native English and Iranian Students.....	39
Table 4.8: Structural Comparison of Bundles in Mechanical Engineering Master Theses of Native English and Iranian Students.....	41
Table 4.9: Structural Comparison of Bundles in Civil Engineering Master Theses of Native English and Iranian Students.....	42
Table 4.10: Functional Analysis of 4-word Bundles in Business Master Theses of Native English and Iranian Students	44

Table 4.11: Functional Analysis of 4-word Bundles in Tourism Master Theses of Native English and Iranian Students	45
Table 4.12: Functional Analysis of 4-word Bundles in Civil Engineering Master Theses of Native and Iranian Students	46
Table 4.13: Functional Analysis of 4-word Bundles in Mechanical Engineering Master Theses of Native and Iranian Students	48

LIST OF FIGURES

Figure 3.1: AntConc Program Displaying the Non-native English Business Students' Use of Bundles	30
Figure 3.2: AntConc Screenshot Showing the Concordances	31
Figure 4.1: Structural Analysis of 4-word Lexical Bundles in Business Sub-corpus..	38
Figure 4.2: Structural Analysis of 4-word Lexical Bundles in Tourism Sub-corpus...	40
Figure 4.3: Structural Analysis of 4-word Lexical Bundles in Mechanical Engineering Sub-corpus	41
Figure 4.4: Structural Analysis of 4-word Lexical Bundles in Civil Engineering Sub-corpus	43
Figure 4.5: Functional Analysis of 4-word Lexical Bundles in Business Sub-corpus.	44
Figure 4.6: Functional Analysis of 4-word Lexical Bundles in Tourism Sub-corpus..	46
Figure 4.7: Functional Analysis of 4-word Lexical Bundles in Civil Engineering Sub-corpus.....	47
Figure 4.8: Functional Analysis of 4-word Lexical Bundles in Mechanical Engineering Sub-corpus	48

Chapter 1

INTRODUCTION

This chapter first outlines the background to the study and then presents the statement of the problem. Later, the purpose and the significance of the study are described. Finally, the definition of key terms (Lexical bundles, Corpus, Discourse and Academic Discourse) is reviewed.

1.1 Background to the Study

The growing heterogeneity of academic texts is one of the challenges facing English for academic purposes (EAP) practitioners, teachers and researchers as academic texts, lectures, papers and assignments come to differ widely in terms of their discursal and linguistic conventions (Durrant, 2017). A number of researchers have pointed to the wide disciplinary variation between texts (e.g., Durrant, 2014; Hyland, 2002; Hyland & Tse, 2007) and the varying needs of students from even interrelated fields of studies (Durrant, 2017).

In academia, graduate studies have been very demanding for the students who face numerous challenges in reading various academic texts, articles, theses and dissertations as well as writing their own projects, articles and theses. Writing academically as a formal skill is an essential requirement that graduate students are expected to acquire; however, the mastery of academic writing skill, especially in EFL contexts, is not an easy task for the graduate students. To be able to write effectively in English, the graduate students are required to acquire the academic discourse of

their field of study which is usually a distinct genre with its own special discourses and linguistic elements.

Further, certain linguistic structures have been reported to constitute a large percentage of written academic text (Biber, Johansson, Leech, Conrad, & Finegan, 1999) indicating that highly frequent and recurring formulaic language deserves special attention (Nation, 2013). Such high-frequency word combinations are an indication of disciplinary variation which makes its exploration a worthwhile quest.

Formulaic combinations and patterns of every academic discipline are an important component of a language which should be acquired. Lexical bundles, as an essential component of formulaic language, are fixed-form recurrent word combinations of multiple words that are frequently used in academic writing (Biber & Conrad, 1999; Hyland, 2008a). In a study conducted by Erman and Warren (2000), lexical bundles are reported to form around 53 % of the written texts. These frequent lexical combinations are very important in the development of academic writing skills as they are an indispensable part of the grammar materials; an indication of successful writing and the lexico-grammatical basis of a language (Coxhead & Byrd, 2007). Moreover, the frequent use of these lexical bundles enables the learners to use the language naturally and fluently (Ellis, Simpson-Vlach & Maynard, 2008).

However, their recurrent use cannot guarantee their easy acquisition by students (Biber & Barbieri, 2007; Cortes, 2006), which is the reason why they should be the focus of language pedagogy and instruction. Knowledge of lexical bundles, or clusters and chunks (Hyland, 2008b), is found to be of great help and importance for language learners (Chen & Baker, 2010; Haswell, 1991; Hyland, 2008a), as it can help tertiary-

level students develop fluent linguistic production and establish membership in their respective disciplines.

Thus, the comparison of native-English and nonnative-English writing can shed light on the structural and functional use of these lexical bundles in academic writing. A number of Iranian L2 writers have explored the use of lexical bundles in Iranian students' writing. For instance, some of the studies done on the structural patterns of Iranian learners' writing have indicated that certain structural categories were used more frequently by native scholars than Iranian experts and post graduate writers (Esfandiari & Barbary, 2017; Jalali, Eslami Rasekh & Tavangar Rizi, 2008) while some categories were found to have the lowest occurrence in the master's theses and doctoral dissertations of Iranian university students (Amirian et al., 2013; Jalali, 2017; Jalali et al., 2008). Moreover, Jalali et al. (2008) found similarity in the functions of lexical bundles used in Iranian and native English master theses and doctoral dissertations while Esfandiari and Barbary (2017) found differences in that regard.

1.2 Statement of the Problem

Several studies have found that learners experience difficulties while using bundles (Adel & Erman, 2012; Hyland, 2008b; Wei & Lei, 2011). That is, some studies have indicated that learners of English misuse, overuse or underuse these lexical bundles in their writing.

Moreover, although extensive research has been done on the use of lexical bundles in various fields, there is still paucity of research into the use of lexical bundles across various disciplines especially in the Iranian context. Therefore, this study aims to examine the use of four-word lexical bundles in the master theses, both nonnative and

native English writers across four disciplines, namely, mechanical engineering (ME) and civil engineering (CE), selected as hard sciences, and business (BS) and tourism (TR), as soft sciences, in terms of their frequencies, structures and functions.

1.3 Purpose of the Study

This study aims to explore the frequency, structure and functions of lexical bundles by comparing and contrasting two corpora of master theses in hard sciences, i.e., mechanical engineering (ME), and civil engineering (CE) and soft sciences, i.e., business (BS) and tourism (TR), one corpus representing Iranian second language writers and the other one representing English native writers. The reason why I intend to explore interdisciplinary variation in the use of lexical bundles is because it is still relatively under-researched and an interesting area for pedagogic reasons such as ESP studies, academic varieties, and genre characteristics (Biber & Barbieri, 2007).

The main research questions which this study intends to address are as follows:

- 1) What are the most frequently used four-word lexical bundles in the two corpora of native and nonnative (Iranian) writers of English?
- 2) What are the structural characteristics of these bundles in each corpus?
- 3) What are the functional characteristics of these bundles in each corpus?

1.4 Significance of the Study

As mentioned above, to the best of the researcher's knowledge, no study has compared the use of lexical bundles between native writers and Iranian L2 writers who study at an international university outside of Iran in writing master theses. Therefore, the findings of this study can help us gain a better insight into the use of lexical bundles across different disciplines. It can also reveal how Iranian master students use lexical bundles in writing their theses. Moreover, the findings can indicate how the use of

such bundles vary or is similar to that of the native-English writers. Such a perspective in turn can help language teachers and practitioners to devise better teaching materials for master students.

1.5 Definition of Key Terms

In this section, the operational definitions of the key terms of the study are provided.

1.5.1 Lexical Bundles

The term ‘lexical bundles’ was first coined by Biber et al. (1999) and is referred to recurring multiple-word combinations which are identified based on their frequent co-occurrence in texts. Biber and Barbieri (2007) enumerated three main characteristics of lexical bundles which are their commonality, non-idiomaticity as well as non-saliency as a complete structural unit.

1.5.2 Corpus Linguistics and Corpora

Corpus linguistics can be seen as either a methodology (Gray & Biber, 2013) or a theory (Baker, 2010) and is defined as the exploration of linguistic variation in the large data sets. According to Baker (2010), there are two types of corpora, namely ‘general corpus’ and ‘specialized corpus’. A general corpus usually consists of considerable word lengths which includes texts from various sources representing different linguistic contexts; while a specialized corpus, as the name suggests, is concerned with addressing certain research questions by applying specific restrictions to the data.

Some other researchers have grouped corpora based on the data collection processes. McEnery and Hardie (2012) divided corpora into two types: ‘monitor corpora’ and ‘sample corpora’ (or ‘balanced corpora’). The former ones are collected diachronically and pre-determined criteria are used to select a certain item; whereas sample corpora

are a collection of real language at a particular point in time. The two corpora collected in this study can be considered both specialized and sample corpora.

1.5.3 Discourse and Academic Discourse

Gray and Biber (2013) grouped various definitions on discourse into three main types as follows:

1. discourse as the language used in communication which traces variations occurring to various language forms or constructs;
2. “discourse as language structure above the sentence level,” which looks beyond the surface forms of the sentence; and
3. “discourse as social practices and ideologies” which are related to language and communication of the members of a certain discourse community (p.138).

Academic discourse, on the other hand, refers to how language is used and viewed in the academia (Hyland, 2009), that is, how language is used in academic texts.

1.6 Summary

In this chapter, first the background to the study was provided. Then, the statement and the purpose of the study were elaborated on. Later, the significance of the study and finally the definition of key terms were explained.

Chapter 2

REVIEW OF LITERATURE

In this chapter, first, the definitions of lexical bundle and other related concepts such as collocation and idiom are presented. Then, widely-used taxonomies of lexical bundles are discussed. Subsequently, studies on lexical bundle use from three perspectives, namely frequency, structure, and function, are reviewed. Finally, the most relevant studies on academic writing of both L1 English and L2 Iranian writers are presented.

2.1 Formulaic Language

Formulaic language is an important aspect of second language acquisition (Bannard & Lieven, 2012) which is viewed as “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated” and 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).

Such set phrases are reported to enable learners to have a better understanding of pragmatically appropriate language and can help them communicate even when they have not mastered the other aspects of language (Wood, 2015). These phrases are said to act as building blocks for other aspects of language (Ellis, 1996), and a growing body of research has investigated its role in academic writing. In this regard, Cowie (1992) found that mastery of formulaic sequences can improve second language writing skills. In other words, knowledge of formulaic sequences is necessary for

students entering English-medium programs as they are required to submit several writing tasks or projects every semester. There are different types of formulaic sequences such as fixed expressions, collocations and lexical bundles. The focus of this study will be lexical bundles, which are a defined type of formulaic language.

2.2 The Concept of Lexical Bundles

The pioneering work on lexical bundles is ascribed to Altenberg (1993, 1998), who systematically initiated research on these word combinations. However, the first appearance of this concept was in Biber et al.'s (1999) work. The scholars defined lexical bundles as recurring multiword combinations which are identifiable based on their frequency of occurrence across multiple texts. To identify lexical bundles, Biber and Barbieri (2007) proposed three key characteristics of them as follows: (1) "lexical bundles are by definition extremely common"; (2) "most lexical bundles are not idiomatic in meaning and not perceptually salient"; and (3) "lexical bundles usually do not represent a complete structural unit" (pp. 269-270).

Lexical bundles commonly occur in academic texts and act as both discourse markers which connect old information to new one (Biber & Barbieri, 2007) and as interactional devices between the writer and the reader (Hyland, 2005, 2008c). Nation (2013) emphasizes the learning value of bundles by considering their frequency and range. These multiword combinations are considered by many researchers as an important aspect of vocabulary knowledge and language production (Firth, 1957; Lewis, 2008; Sinclair, 1991).

In recent years, lexical bundles have received considerable attention in second language academic writing (e.g., Hyland, 2008a; Xu, 2012); lexical bundles are an

indication of fluency, accuracy, idiomaticity as well as belonging to a particular linguistic community (Li, 2016; Pawley & Syder, 1983). Also, lexical chunks are said to constitute a large proportion of language (Erman & Warren, 2000; Schmitt & Carter, 2004) of which lexical bundles form a major part of academic writing. In this regard, Biber et al. (1999) found that three-word and four-word lexical bundles are used more than 60,000 times and 5,000 times respectively in each million words. Some of the most frequent bundles such as *in order to*, *one of the*, *in the case of*, and *on the other hand* occur over 100 times per million words in academic texts making up about more than 20% of around 5.5 million words of the Longman Spoken and Written English Corpus (Biber et al., 1999). In addition, Hyland (2008b) found that the commonly used bundle, *on the other hand*, is repeated around 200 times per million words. Such numbers indicate the important value of these bundles for learning purposes (Nation, 2013).

Lexical bundles should be distinguished from other formulaic language terms such as idioms and collocations. There are two features of lexical bundles that distinguish them from idioms, namely frequency and transparency of meaning (Li, 2016). For example, lexical bundles occur 10 to 40 times in a million word while idioms such as *bury the hatchet* rarely occur in texts. Moreover, the meaning of lexical bundles is transparent throughout the texts while the meaning of idioms is not often transparent. However, the identification of lexical bundles in the sentence is not easy as they tend to occur everywhere in the sentence (Biber et al., 1999). Thus, identifying lexical bundles across various registers, disciplines, or genres is required to shed light on frequently occurring word combinations that can potentially serve as valuable language resources for second language learners.

Collocation, on the other hand, has been defined differently by various researchers in the field. For instance, collocations can be identified based on three approaches, namely, frequency, structure and collocability (Li, 2016), and more than one approach needs to be taken into account for the definition of collocation. According to a frequency-based approach, collocation is defined as words that frequently co-occur or occur in the proximity of one another (Nation, 2013; Sinclair, 1991). A structurally-based approach identifies collocations based on their structural formation or syntax (Nation, 2013; Sinclair, 1991). Collocability-based approach, on the other hand, looks at the mutual expectancy between words or the likelihood of their co-occurrence (Lewis, 2008; Nation, 2013). Li (2016) notes that collocations can either be fixed phrases such as *by the way* or combination of words that can be substituted with other words such as *have a meal*.

Overall, lexical bundles differ from collocations and formulaic sequences on two grounds: the degree of fixedness and arbitrary frequency-based identification criteria (Li, 2016). Lexical bundles are completely fixed in terms of their semantic and syntactic forms while collocations and formulaic sequences vary in terms of their fixedness. Further, an arbitrarily-set cut-off criteria is used to generate lexical bundles, which is the minimal number of occurrences of words and their distribution in texts. Moreover, lexical bundles are incomplete structural units such as *it is possible that*, *the importance of the*, which only represents part of a clause or sentence (Li, 2016).

2.3 Taxonomies of Lexical Bundles

After Altenberg's (1993, 1998) work, a large body of research has been conducted on lexical bundles in relation to various issues such as languages, registers, genres, disciplines, language proficiency, academic competence and moves. These studies

have almost followed three typical research foci namely, frequency, structure and function, which will be discussed further in the related studies section. In most of the studies, three popular taxonomies have been used: Biber et al.'s (1999) structural taxonomy, Biber and Barbieri's (2007) functional taxonomy, and Hyland's (2008a) functional taxonomy. These taxonomies will be discussed and described below and form the analytical framework used in this study.

2.3.1 Biber et al.'s (1999) Structural Taxonomy of Lexical Bundles

Biber et al. (1999) classified bundles in academic texts into 12 widely-used structural patterns which have been extensively used in previous studies. In this taxonomy, lexical bundles are categorized into three groups: phrasal, clausal and other expressions (see Table 2.1).

Phrasal bundles are further classified into three subcategories: Noun-Phrase (NP)-based, Preposition Phrase (PP)-based, and Verb Phrase (VP)-based bundles. NP-based bundles are also of two types: NP with of-phrase fragment (e.g., *the end of the*) and NP with other post modifier (e.g., *the extent to which*). PP-based bundles, on the other hand, include PP with embedded of-phrase fragment (e.g. *on the basis of*) and other Prepositional Phrase or fragment (e.g., *in accordance with, in this case*). Furthermore, VP-based bundles refer to any phrases with a verb element, and they are further subdivided into four categories: Anticipatory it + VP/adjective phrase (e.g., *it is necessary to*), Passive verb +PP fragment (e.g., *is shown in figure*), Copula be + noun phrase/adjective phrase (e.g., *is due to the*), and Pronoun/NP + be (e.g., *there are a number of*).

Table 2.1: Biber et al.'s (1999) Structural Classification of Lexical Bundles

Category	Example
A. Phrasal	
1. NP-based	
NP with <i>of</i> - phrase fragment	<i>the base of the</i>
NP with other post modifier fragment	<i>the extent to which</i>
2. PP-based	
PP with embedded <i>of</i> -phrase fragment	<i>as part of the</i>
Other prepositional phrase (fragment)	<i>at the same time</i>
3. VP-based	
Anticipatory it + VP/adjective phrase	<i>it can be seen</i>
Passive verb +PP fragment	<i>were carried out at</i>
Copula be + noun phrase/adjective phrase	<i>was similar to the</i>
Pronoun/NP + be	<i>there are a number of</i>
B. Clausal	
(V/Adjective +) to-clause fragment	<i>is not possible to</i>
(VP +) that-clause fragment	<i>has been proved that</i>
Adverbial clause fragment	<i>as we shall see</i>
C. Other Expressions	
	<i>than that of the</i>

Source: Biber et al.(1999, pp. 1014- 1015)

Furthermore, clausal lexical bundles include (verb/adjective +) to-clause fragment (e.g., *is likely to be*), (VP+) that-clause fragment (e.g., *has been shown that*), and adverbial clause fragment (e.g., *as shown in figure*). Finally, the last category is referred to as ‘Other Expressions’ which include expressions such as *as well as the*, *than that of the*, to name just a few.

2.3.2 Biber and Barbieri’s (2007) Functional Taxonomy of Lexical Bundles

Based on Biber, Conrad and Cortes’ (2003, 2004) earlier works, Biber and Barbieri (2007) proposed a functional taxonomy of lexical bundles by dividing the bundles into three functions namely, stance expressions, discourse organizers and referential expressions. The first function in this typology is stance bundles which are subdivided into two sub-functions of epistemic stance and attitudinal bundles. Epistemic stance bundles are also referred to as assessment bundles (e.g., *I believe that the, it is*

necessary for). On the other hand, attitudinal bundles, also known as attitude or modality stance bundles, are further divided into desire bundles (e.g., *I would like to*), obligation/directive bundles (e.g., *do you want me*), intention/prediction bundles (e.g., *are we going to*) and ability bundles (e.g., *it is possible to*).

Table 2.2: Biber and Barbieri's (2007) Functional Classification of Lexical Bundles

Stance bundles (SB)
Epistemic stance bundles: <i>I think that the, I believe that the</i>
Attitudinal/modality stance bundles
Desire bundles: <i>I don't want to</i>
Obligation/directive bundles: <i>you have to do</i>
Intention/prediction bundles: <i>what we're going to</i>
Ability bundles: <i>to be able to</i>
Discourse organizers (DO)
Topic introduction bundles: <i>What I want to do is</i>
Topic elaboration/clarification bundles: <i>has to do with the</i>
Identification/focus bundles: <i>those of you who</i>
Referential bundles (RB)
Imprecision bundles: <i>or somethings like that</i>
Bundles specifying attributes: <i>a little bit of</i>
Time/place/text-deixis bundles: <i>the beginning of the, in the United States</i>

Source: Biber and Barbieri (2007, pp. 270-272)

The next function is discourse organizers which are further divided into other functions such as topic introduction bundles (e.g., *do you know what*), topic elaboration/clarification bundles (e.g., *know what I mean*), and identification/focus bundles (e.g., *those of you who*). The third function is referential bundles which are categorized into four major functions namely, imprecision bundles (e.g., *or something like that*), bundles specifying attributes (e.g., *a little bit of*) and time/place/text deixis bundles (e.g., *the end of the, in the United States, as shown in figure*). It is worth

mentioning that due to its foundation on Halliday's (1994) functional linguistics, Biber and Barbieri's (2007) taxonomy is mostly biased towards spoken texts.

2.3.3 Hyland's (2008a) Functional Taxonomy of Lexical Bundles

More recently, Hyland (2008a) proposed a functional framework of lexical bundles based on Halliday's (1994) functional linguistics and by drawing on Biber et al.'s (2003, 2004, 2007) taxonomies. The taxonomy was proposed based on three electronic corpora composing of a large body of academic texts collected from four different areas of study: applied linguistics, business studies, electrical engineering, and microbiology. Hyland's (2008a) taxonomy consists of three major functions of language: research-oriented, text-oriented and participant-oriented bundles. Research-oriented bundles are related to the ideational function of language such as location (e.g., *at the beginning of, in the present study*), procedure (e.g., *the use of the*), quantification (e.g., *a wide range of*), description (e.g., *the structure of the*) and topic (e.g., *the current board system*). Text-oriented bundles consist of textual functions or the text organization which include transition signals (e.g., *as well as the*), resultative signals (e.g., *due to the fact*), structuring signals (e.g., *the figure shows the*) and framing signals (e.g., *in the case of*). Participant-oriented bundles, on the other hand, are related to interpersonal functions which include stance features (e.g., *it is possible that*) and engagement features (e.g., *it was determined that*). Hyland (2008a) proposed a set of criteria for each of these sub-categories (see Table 2.3).

Table 2.3: Hyland's (2008a) Functional Taxonomy of Lexical Bundles
Research-oriented (RO)

- Location: *at the beginning of, the bottom of the*
 - Procedure: *the development of the, the implementation of the*
 - Quantification: *the magnitude of the, a wide range of*
 - Description: *in the form of, the behavior of the*
 - Topic: *in the hospitality industry*
-

Text-oriented (TO)

- Transition signals: *as well as the, the interaction between the*
 - Resultative signals: *due to the lack, in order to determine*
 - Structuring signals: *as seen in figure, is defined as a*
 - Framing signals: *to the fact that, as a matter of*
-

Participant-oriented (PO)

- Stance features: *it is possible to, is more likely to*
 - Engagement features: *it can be seen, it is clear that*
-

Source: Hyland (2008a, pp. 49-50)

2.4 Previous Studies on Lexical Bundles

As stated earlier in the previous sections, a large body of research has been conducted on lexical bundles since Altenberg's (1993) study. These studies have examined the use of lexical bundles from various perspectives: genre (Biber, 2006; Chen, 2010; Hyland, 2008a; Nesi & Basturkmen, 2006; Qin, 2014; Scott & Tribble, 2006), discipline (Durrant, 2017; Hyland, 2008b), proficiency levels (Ädel & Erman, 2012; Cortes, 2004; Hyland, 2008a; Pan, Reppen & Biber, 2016; Salazar, 2014), the behavior of native and nonnative writers (Pan et al., 2016; Römer & Arbor, 2009) and registers (Biber et al., 1999; Biber et al., 2004). However, almost all of these studies have adopted a similar methodology or framework by examining the bundles from three perspectives: frequency, structure and function. Some of the studies have looked at the lexical bundles from more than one perspective (frequency, structure and function). Since the focus of this study is on the use of lexical bundles in academic

writing (e.g., master theses) of native and nonnative writers of English, at first the most relevant studies on L2 academic writing will be reviewed in terms of frequency, structure and function.

2.4.1 Frequency-based Analysis

Biber et al. (2004) emphasized the key role of frequency in the identification of lexical bundles and stated that "... frequency data identifies patterns that must be explained" (p. 376). Biber and Barbiery (2007) also argued that since lexical bundles are used highly frequently, learners can acquire them naturally without receiving explicit instruction. However, they noted that discourse functions of bundles should be taught. Following this assumption, various corpus-based studies explored the differences if any in the frequency of lexical bundle use between native and nonnative writers of English (e.g., Ädel & Erman, 2012; Bychkovska & Lee 2017; Hsu, Chen, Yang & Liu 2017; Hyland, 2008a; Karabacak & Qin, 2013; Xu, 2012). It should be stated here that for the analysis of frequency of lexical bundles, two terms need to be defined, type and token. The former is concerned with unique bundles and the latter refers to the total occurrences of bundles (Chen & Baker, 2010).

In one study, Chen and Baker (2010) examined the number of lexical bundles used in Chinese student writing and their L1 English peers, studying in British universities. The scholars concluded that compared to native students, nonnative learners of English used much more limited sets of bundles (i.e., 90 compared to 120). Also, Ädel and Erman (2012) reported the use of fewer types of lexical bundles in Swedish students' writing than that of English students (i.e., 60 compared to 130). By the same token, Karabacak and Qin (2013) analyzed highly used lexical bundles of the argumentative papers written by Turkish, Chinese, and American students, and reported that

American students made more use of the bundles than their Turkish and Chinese counterparts (i.e., 94 compared to 92 and 54, respectively).

However, Pang (2009) and Bychkovska and Lee's (2017) results did not support the findings of the studies above. In Pang's (2009) research, Chinese university undergraduates used significantly more bundles (more than three times) in their essays than native British and American students. Likewise, Bychkovska and Lee (2017) found that L1 Chinese students applied more types of bundles in their argumentative essay writing than native English speaker students (i.e., 404 compared to 337). This mismatch is attributed to the language proficiency and study level, that is, as the proficiency level or study level increases, the use of lexical bundles decreases (Li, 2016). For example, Hyland (2008a) reported that Chinese master students incorporated more bundles in their theses than doctoral students in their dissertations and native authors in their articles (i.e., 149, 95, and 71, respectively). In the same vein, Wei and Lei (2011) identified 154 and 87 lexical bundles in doctoral dissertations of Chinese students and published articles of native English writers, respectively. Xu (2012) also found that Chinese master's students majoring in linguistics or applied linguistics employed more lexical bundles in their theses than PhD students in their dissertations and professional scholars in the same disciplines (i.e., 367, 168, and 169, respectively). Additionally, comparing lexical bundle usage by Spanish expert writers and native English authors, Pérez-Llantada (2014) reported that the Spanish writers used more bundles than their English counterparts.

In line with these findings, some similar results have also been reported on Iranian writers' use of lexical bundles (e.g., Alipour & Zarea, 2013; Amirian, Ketabi &

Eshaghi, 2013; Jalali, Eslami Rasekh & Tavangar Rizi, 2008). For instance, Jalali et al. (2008) compared 22 master theses and 12 doctoral dissertations (post-graduate writing) with 201 research articles written by native English speakers in applied linguistics and found that the number of different lexical bundles used by Iranian post graduate students (the study level was not differentiated in the study) was twice as much as the number of bundles used by English published authors (i.e., 261 compared to 125). Moreover, by comparing three corpora of research articles written by English authors in physics, computer engineering and applied linguistics and one nonnative corpus of Iranian writers in applied linguistics, Alipour and Zarea (2013) reported that Iranian authors employed more lexical bundles in their writing than their native counterparts. Similarly, Amirian et al. (2013) reported that the number of bundles used by Iranian master students were three times more than the number of bundles applied by their native English peers (i.e., 211 compared to 61).

However, through investigating the bundles used in the discussion sections of 60 articles (30 articles by each group) in the field of political sciences, Safarzadeh, Monfared and Sarfeju (2013) identified fewer types of lexical bundles used by Iranian published authors compared to native speaker professional writers (i.e., 132 to 178). Esfandiari and Barbary (2017) also reported similar results in their study by examining lexical bundles used in psychology research articles written by professional Iranian and English expert writers (i.e., 416 compared to 316). Overall, with regard to their study and proficiency level, research results on the frequency of lexical bundles used by Iranian L2 writers seem to indicate contrary findings.

2.4.2 Structural Analysis

Biber et al.'s (1999) model has been commonly used in majority of studies to identify the use of bundles by nonnative (L2) English speakers and native (L1) speakers and scholars in the field. Initially, Biber et al. (1999) found twelve frequently-used structural patterns in conversation and academic writing. Later, Biber et al. (2004) studied the use of lexical bundles in two registers of classroom teaching and textbooks and proposed three structural patterns namely, *verb phrase fragment*, *dependent clauses*, *noun and prepositional phrase fragments*.

2.4.2.1 Studies on Structural Analysis

Studies done by Biber et al. (1999, 2004) on spoken discourse revealed that the dominant structural pattern in conversation is verb phrase lexical bundles with about 90% occurrences which consist of 50% personal pronoun + verb phrase (e.g., *they want to*), 19% extended *verb phrase fragments* (e.g., *should be noted that*) and 17% question fragments (e.g., *do they ask to*). However, in academic prose 60% of the bundles were reported to be noun phrase (e.g., *the performance of the*) and prepositional phrase (e.g., *in a way that*) (Biber et al., 1999; Biber et al., 2003, 2004). Therefore, these studies concluded that frequently used structural categories of lexical bundles in academic writing are *noun phrase*, *prepositional phrase*, *passive verb phrase* and *anticipatory-it bundles* (Hyland, 2008a).

Some of the studies done on the structural patterns of Iranian learners' writing have indicated that *prepositional phrases with of* were used more frequently by native scholars than Iranian experts and post graduate writers (Esfandiari & Barbary, 2017; Jalali et al., 2008). Some others have also pointed that among all the structural patterns, *anticipatory it bundles* had the lowest occurrence in the master's theses and doctoral

dissertations of Iranian university students (Amirian et al., 2013; Jalali, 2017; Jalali et al., 2008).

Moreover, in Jalali et al.'s (2008) study, *noun phrases with of* were reported to be used much more than other phrases in the writing of the Iranian postgraduate students, while Amirian et al. (2013) found that in their master theses Iranian writers used *others* (those without complete and unified structure) patterns more than other structural categories. Additionally, Esfandiari and Barbary (2017) noted that like native authors, Iranian L2 writers also relied more on *prepositional phrases with of* in their writing. However, Jalali and Ghayoomi (2010) found no difference in the lexical bundle usage by Iranian postgraduate students and English professional writers although some differences were reported regarding the extent to which each group drew on some specific bundles.

By comparing the findings of these studies, I cannot find a consistent pattern which can be partly due to the fact that these authors looked at the structural bundles from various perspectives. For example, Jalali (2017) only researched the *anticipatory it* bundles in his study.

2.4.3 Functional Analysis

Functional analysis, as another research perspective on lexical bundles, looks at the intrinsic functions of these word combinations. As mentioned earlier in the taxonomy section, two existing and widely used functional taxonomies are Biber et al.'s (2004) and Hyland's (2008a) taxonomies. These taxonomies have been widely adopted and adapted in lexical bundles studies. Biber et al.'s (2004) taxonomy contains three key functions namely, *stance expressions, discourse organizers, and referential*

expressions, which are further divided into other smaller sub-categories. Hyland's (2008a) framework is a revised version of Biber et al.'s (2004) taxonomy which consists of three main functions: *research-oriented*, *text-oriented*, and *participant-oriented*. These categories are also further categorized into smaller sub-groups.

2.4.3.1 Studies on Functional Analysis

Biber et al. (2004) and Biber and Barbieri (2007) compared the use of lexical bundles in conversation, classroom teaching, textbooks and academic prose. The findings of their study indicated that *stance* and *discourse organizer bundles* were mostly used in oral language while *referential bundles* were prevalent in written university discourses.

A number of studies conducted on functional analysis of L2 writers' use of lexical bundles have adopted Biber et al.'s (2004) and Biber and Barbieri's (2007) functional models and reported similar results (Ädel & Erman, 2012; Chen & Baker, 2010; Pérez-Llantada, 2014; Uçar, 2017; Xu, 2012). According to the findings of these studies, there was a functional similarity in distribution of lexical bundles: English native student writers and scholars employed more referential and stance bundles, while L2 students and experts mostly used discourse organizer bundles.

Hyland (2008a) compared the use of bundles in master's theses and doctoral dissertations written by L1 Cantonese speakers with research articles written by native English experts. The results revealed that research-oriented bundles were extensively used by masters' students. On the contrary, research articles were dominated by text-oriented and participant-oriented bundles. In PhD dissertations, however, the

distribution of bundles was similar to that of published articles in using more text-oriented and less research-oriented bundles.

Studies done on functional analysis of Iranian writers' use of lexical bundles have mostly employed Hyland's (2008a) framework. In this regard, Jalali et al. (2008) analyzed master theses and doctoral dissertations as two post graduate genres written by Iranian writers and compared them with published research articles in the applied linguistics and found that both groups used research-oriented bundles the most and participant-oriented clusters the least in their writing. Somewhat similarly, Jalali and Ghayoomi (2010) analyzed the use of target bundles of two sets of postgraduate writing corpora by Iranian writers in Iran with one corpus of research articles by English writers in the same discipline and reported research-oriented clusters as the most frequently employed bundles by native experts. Further, by comparing Iranian students' and English students' master's theses, Amirian et al. (2013) indicated that Iranian students were more interested in research-oriented clusters while native English speaker students relied more on text-oriented bundles. However, Esfandiari and Barbary (2017) found that text-oriented bundles had the highest and participant-oriented clusters had the lowest frequency in the articles of both English and Iranian scholars in published articles of psychology.

2.5 Summary

In this chapter, the definitions of lexical bundle, other related concepts such as collocation and idiom as well as the widely used taxonomies of lexical bundles were presented and discussed. Then, the most relevant studies on lexical bundle use from three perspectives, namely frequency, structure, and function, were reviewed. In the next chapter, the methodology of the research will be discussed.

Chapter 3

METHODOLOGY

In this chapter the methodology adopted for the study is presented. First, research design and corpus building procedures are explained. Then, the collection of the two corpora created for this study (a corpus of master theses written by the Iranian L2 writers and a corpus of master theses written by the native English writers) is discussed. Next, bundle identification procedures in the two corpora are reviewed and the concordancing program is described in detail.

3.1 Research Design

This study adopted corpus analysis as a methodology which aims to explain the nature, structure and use of language and languages, especially issues such as language acquisition, language variation and language change (Kennedy, 2014). Corpus studies are often categorized into corpus-driven and corpus-based studies. Corpus-based studies draw on corpus data to explore, confirm or refute theories or hypotheses while corpus-driven approach uses the corpus itself as the source of hypothesis or theory-building (Tognini-Bonelli, 2001). However, to avoid binary distinction between corpus-based and corpus-driven linguistics (McEnery & Hardie, 2012), I use the term corpus-based which covers both concepts.

3.2 Corpus Building Procedure

The current study explored the use of lexical bundles in 120 theses from two corpora of master theses written by the Iranian L2 writers and L1 English writers. The first corpus contained 60 theses written by the Iranian graduate students who studied at

Eastern Mediterranean University (EMU), Famagusta, North Cyprus, and the other corpus included 60 theses written by native writers of English at three USA universities: California State University, Iowa State University and University of Nevada. The theses were selected from four disciplines, namely mechanical engineering (ME) and civil engineering (CE), selected as hard sciences, and business (BS) and tourism (TR), as soft sciences. The main reason for the selection of the four disciplines is that there is a paucity of research in the Iranian context on the comparison of lexical bundles use between the Iranian and the English writers and almost even no study on the use of lexical bundles across soft and hard disciplines. It is safe to say that the majority of the studies in the Iranian context has been conducted on the articles and theses written in the field of applied linguistics as the theses in other disciplines of study are written in Persian.

Since there are no fixed criteria in literature for the identification of the writers' first language (Esfandiari & Barbary, 2017), I adopted the method proposed by Wood (2001), who considered L1 English writer as any author whose first and last name is considered native by native English-speaking countries. Thus, the basis for the selection of English writers in native English theses was native English names. For the identification of the Persian writers also the same criterion was employed. Therefore, theses written by nonnative names were removed from the study on the basis of this criteria. The researcher is aware of the limitation of this selection; however, it was the most practical and convenient approach possible for the selection (Li, 2016).

Moreover, since the two main concerns of the corpus building are corpus size and representativeness (Gray & Biber, 2013), efforts were made to select a large corpus

which can adequately represent the occurrence of the lexical bundles (Biber, 2006). The native speaker corpus (NSC hereafter) consisted of 712,728 and the nonnative speaker corpus (NNSC) or the Iranian L2 corpus contained 675,157 words which together totaled 1,387,885 words. The number of words in each discipline varied to some degree. For example, in the NSC, native speaker business (NSB) contained 150,295 with an average length of 10,019 for each thesis; native speaker civil engineering (NSCE), 239,825 words with an average word length of 15,988; native speaker mechanical engineering (NSME), 170,610 words with an average word length of 11,374; and, native speaker tourism (NST), 151,998 with an average word length of 10,133. In the nonnative speaker corpus (NNSC), nonnative speaker business (NNSB) consisted of 129,461 with an average word length of 8,630 for each thesis; nonnative speaker civil engineering (NNSCE), 170,388 with an average word length of 11,359; nonnative speaker mechanical engineering (NNSME), 132,128 with an average word length of 8,808; and nonnative speaker tourism (NNST), 243,180 with an average word length of 16,212 for each thesis. It should be mentioned here that word length in sub-corpora varied to some degree which the researcher had less control over. Table 3.1 displays information on the number of words and average word length in each discipline of the two corpora. As can be seen, the NSC sections comparatively contained more words than the NNSC sections.

Table 3.1: Words Counts and Average Word Length of the Sections of the Two Corpora

	NSB	NSCE	NSME	NST	NNSB	NNSCE	NNSME	NNST
Theses	15	15	15	15	15	15	15	15
Words	150,295	239,825	170,610	151,998	129,461	170,388	132,128	243,180
Length	10,019	15,988	11,374	10,133	8,630	11,359	8,808	16,212

3.2.1 The Nonnative Speaker Corpus (NNSC)

Since lexical bundles are register-bound (Conrad, 1996), I focused only on master theses. First, by using the online database at Eastern Mediterranean University, master theses written by the Iranian L2 writers in four disciplines of mechanical engineering, civil engineering, business and tourism were randomly selected and downloaded into four different files. The reasons for selecting theses written by the Iranian L2 writers from Eastern Mediterranean University are twofold. The first reason is to do with the lack of access to master theses written in Iran. The researcher contacted some universities in Iran and asked for permission to get access to master theses but unfortunately was not granted permission. The second reason is related to the language in which theses are written in Iran. The majority of theses written in fields other than applied linguistics are in Persian. Thus, the researcher selected theses written by the Iranian L2 writers in Northern Cyprus for the stated reasons which somehow limited the scope of our data for the nonnative theses. I also attempted to select disciplines of the same size in both corpora; however, it was almost impossible to find theses of the same size for each field in the two corpora. Moreover, I ensured the recency of the theses by selecting the thesis written between the years 2010 and 2017.

3.2.2 The Native Speaker Corpus (NSC)

As mentioned in the data collection section, the theses for the NSC were downloaded from the online library of three top universities in the USA. The mechanical and civil engineering theses were downloaded from Iowa State University online database (<https://lib.dr.iastate.edu/>). Tourism theses were downloaded from University of Nevada in Las Vegas (<https://digitalscholarship.unlv.edu/thesesdissertations/>) and business theses from California State University (<https://pqdtopen.proquest.com/search.html>). The selection criteria used to select

theses from these universities were random selection as well as open access to theses from their online repository.

3.3 Bundle Identification

Each corpus in this study contained 60 theses, 15 theses from four disciplines (business, civil engineering, mechanical engineering and tourism), each written between years 2010 and 2017. First, the electronic copies were downloaded from the digital repositories and online libraries of the respective universities mentioned earlier. Then, the electronic copies (PDF files all) were converted into word documents at <https://document.online-convert.com/convert-to-txt> which is a popular online website for conversion of textual and non-textual formats into multiple formats. Later, non-textual annotations such as titles, tables, graphs, formulas, and references were erased. Subsequently, the word documents were converted into text for the final analysis.

The current study aimed at finding the most common lexical bundles in the two corpora by examining both their structural and functional aspects. The criteria for bundle identification, such as the frequency of lexical bundles and range cut-off points (range limits), vary from study to study; however, in this study I followed Cortes' (2004) criteria and focused on four-word lexical bundles since most of the four-word bundles include the three-word bundles and their rate of occurrence is usually higher than five-word bundles. Moreover, four-word lexical bundles encompass a large number of structures and functions (Hyland, 2008b).

For the analysis of frequency different criteria have been adopted by different researchers as the frequency criterion is "somewhat arbitrary" (Biber & Barbieri, 2007, p. 267; Hyland, 2008, p. 8) and is always based on a normalized frequency which is

the number of occurrences of a certain lexical bundle per one million words (pmw). That is, some studies set the frequency rate at 10 times pmw (Biber et al., 1999) while some followed a higher or stricter cut-off point. In this study, I adopted Biber and Barbieri's (2007) frequency criterion which is a cut-off point of 40 times pmw.

Besides, for the analysis of range or measures of dispersion which is used to minimize the effect of one writer preference in skewing the results (Pan et al., 2016), again like the other two previously mentioned criteria (the number of words per bundle and frequency of occurrence), different researchers have used different cut-off point. Biber et al. (2004) set the cut-off point criterion at 2%, Biber and Barbieri (2007) considered the rate of occurrence around 5% of texts while Hyland (2008) set the cut-off point in around 10% of texts (Hyland, 2008). However, in this study I even followed a stricter selection criterion for the occurrence of lexical bundles in at least 20% of the theses which in our study was three theses.

It should be mentioned here that since the word length in each sub-corpus varied to a smaller degree, the frequency cut-off point also varied from each section of the corpus to the other. For example, by looking at Table 3.2, it can be seen that the frequency cut-off point for native civil-engineering theses is 9, while for nonnative business thesis it is 5, which has been set according to the number of words in each sub-section, 239825 and 129461 respectively.

Table 3.2: Frequency and Range of the Sub-sections of the Two Corpora

Discipline	NSB	NSC	NSM	NST	NNSB	NNSC	NNM	NNST
Frequency	6	9	7	6	5	7	5	9
Range	3	3	3	3	3	3	3	3
Word count	150295	239825	170610	151998	129461	170388	132128	243180

3.4 Concordancing Software: AntConc

The concordancing software used to analyze the data in this study was AntConc computer software version 3.5.2, which is a handy and user-friendly text analysis tool designed and developed by Laurence Anthony (2018). The reason for the selection of this software was the features that the program has, making it suitable for studies on lexical bundles. The most useful features of the software are word and keyword frequency generators, minimum range, maximum range, and tools for cluster or N-grams (groups of 'N' words which appear in sequence in the text) analysis (see Figure 3.1). Also, the software can efficiently identify word bundles or combinations, which makes it a good fit for the data analysis in this study.

The word documents were all converted into plain texts before being uploaded into AntConc. Then, using the "N-Grams" command in AntConc, I set the frequency counts at 4-grams for the analysis of the entire 4-word lexical bundles. This command performed a full extract of any 4-grams from each sub-corpora of the corpus. Moreover, by using the minimum and maximum 4-gram frequency as well as the minimum and maximum range, I ensured the occurrence of the expressions with the specified frequency and range. After running the program based on the previously mentioned settings, a list of four-word lexical bundles with the specified cut-off range and frequency was retrieved. Then, each expression was manually checked to see whether the expression was a four-lexical bundle or not. As a result, expressions that were not deemed to be four-word lexical bundles were excluded from the study. (of the world' s and x y and z)

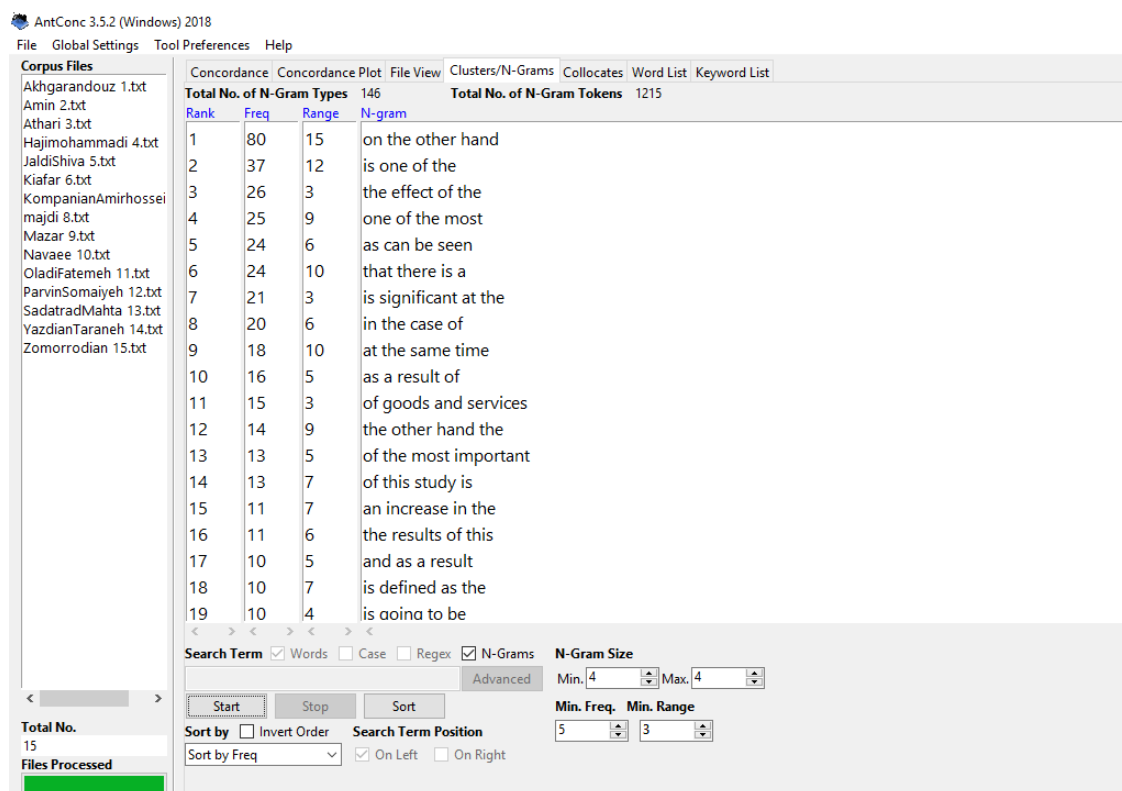


Figure 3.1: AntConc Program Displaying the Non-native English Business Students' Use of Bundles

Further, apart from quantitative analysis of the data, AntConc also helped with the qualitative interpretation of the results which this study aimed at: the description of the functional and structural analysis of the identified lexical bundles (see Figure 3.2). Additionally, due to variability in the (grammatical) structures and functions of lexical bundles (Biber et al., 1999, 2003, 2004), in this study I analyzed each bundle elaborately in its context to determine its function.

Furthermore, lexical bundles with similar structural forms and functions were categorized together based on the functional and structural taxonomies by considering their use and meaning in the context where they occurred. Finally, a second rater, a PhD graduate candidate in English language teaching with similar research background, went through the entire data by analyzing the functional and structural

categories separately to ensure the consistency of rater judgments which added to the reliability of the data analysis.

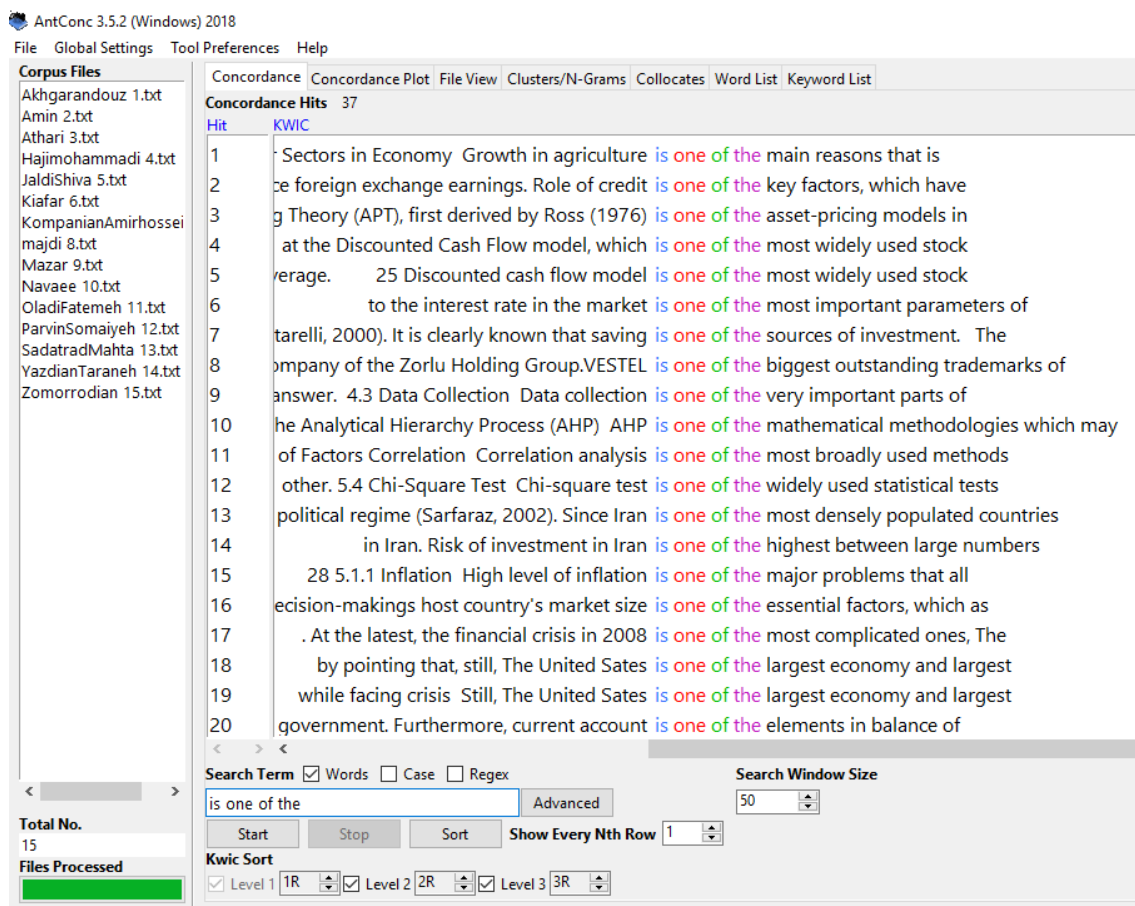


Figure 3.2: AntConc Screenshot Showing the Concordances

3.5 Summary

In this chapter, first I discussed the research design and the corpus-based approach adopted for the study. Then, the procedure for data collection, the collection of the two native and nonnative corpora created for this study, was described. Later, the process of data analysis was explained by describing the AntConc software used in this thesis to analyze the research data. In the next chapter, the findings of the study will be presented.

Chapter 4

RESULTS

In this chapter, the results of the quantitative and qualitative analyses are presented. First, the most frequently used lexical bundles in the two corpora are given. Then, the bundles are structurally categorized, tabulated and compared. Finally, the bundles are discussed and explained in terms of their functions in the text.

4.1 Bundle Frequency in the Corpora

The quantitative analysis of the frequency of lexical bundles across the two corpora revealed that graduate Iranian writers used more bundles than the native English thesis writers. In two soft sciences (namely, tourism and business), the Iranian L2 writers used more lexical bundles, however, in the hard sciences, only civil engineering native English writers used more lexical bundles both in terms of the number of different (type) lexical bundles and the number of total cases (token).

Table 4.1: The Frequency of Lexical Bundles Across the Four Sub-corpora

Corpus	Thesis	Words	Types	Token
NB	15	150,295	43	426
NT	15	151,998	68	810
NCE	15	239,825	141	2137
NME	15	170,610	92	988
NNB	15	129,461	130	1098
NNT	15	243,180	96	1582
NNCE	15	170,388	90	1030
NNME	15	132,128	140	1206

Note: NB: native business, NT: native tourism, NC: native civil engineering, NM: native mechanical engineering, NNB: nonnative business, NNT: nonnative tourism, NNC: nonnative civil engineering, NNM: nonnative mechanical engineering. Type: the frequency of each unique bundle, Token: the total occurrence of each bundle.

However, it is the number of types that represent their frequency in various studies which has also been used in this study. Table 4.1 displays the analysis of lexical bundles across the two sub-corpora.

The thirty most frequent lexical bundles in each sub-corpus are presented in the following tables and in each sub-corpora the most frequent lexical bundles have been highlighted.

Table 4.2: Top 30 Frequent 4-word Bundles in Business Sub-corpus in Rank Order

Rank	Native Business	Freq	Rank	Non-native Business	Freq
1	in the united states	38	1	on the other hand	80
2	are more likely to	16	2	is one of the	37
3	at the end of	15	3	the effect of the	26
4	in the case of	15	4	one of the most	25
5	the end of the	14	5	as can be seen	24
6	the purpose of this	14	6	that there is a	24
7	in addition to the	13	7	is significant at the	21
8	in order to be	12	8	in the case of	20
9	it is important to	12	9	at the same time	18
10	as a result the	11	10	as a result of	16
11	as well as the	11	11	of goods and services	15
12	growth rate of the	11	12	of the most important	13
13	to be able to	11	13	of this study is	13
14	at the same time	10	14	an increase in the	11
15	on the other hand	10	15	the results of this	11
16	when it comes to	10	16	and as a result	10
17	one of the most	9	17	is defined as the	10
18	the growth rate of	9	18	is going to be	10
19	the value of the	9	19	the null hypothesis of	10
20	as a result of	8	20	there is a significant	10
21	as a way to	8	21	when it comes to	10
22	in the form of	8	22	are presented in table	9
23	is one of the	8	23	in order to get	9
24	it is necessary to	8	24	in the long run	9
25	the rest of the	8	25	in the short run	9
26	the success of the	8	26	it is important to	9
27	is similar to the	7	27	of the iranian economy	9
28	it is not surprising	7	28	that there is no	9
29	that make up the	7	29	the findings of the	9
30	will need to be	7	30	the value of a	9

By looking at Table 4.2, we can see that the most frequent bundles in business sub-corpus are *in the case of, it is important to, at the same time, on the other hand, when it comes to, one of the most, the value of the, as a result of, is one of the.*

Table 4.3: Top 30 Frequent 4-word Bundles in Tourism Sub-corpus in Rank Order

Rank	Native Tourism	Freq	Rank	Non-native Tourism	Freq
1	in the united states	76	1	one of the most	75
2	it is important to	50	2	is one of the	71
3	one of the most	26	3	on the other hand	58
4	as well as the	25	4	in the context of	51
5	the purpose of this	24	5	in the case of	49
6	studies have shown that	23	6	as a result of	34
7	in the hospitality industry	20	7	as one of the	33
8	the needs of the	20	8	as well as the	29
9	is one of the	18	9	in the process of	29
10	in addition to the	17	10	customer satisfaction and loyalty	27
11	in the Las Vegas	17	11	of this study is	27
12	of the united states	16	12	of the most important	26
13	of this paper is	15	13	for the purpose of	22
14	this paper is to	14	14	image of a destination	22
15	to the success of	14	15	of the study this	22
16	the purpose of THIS	12	16	the findings of this	21
17	can be used to	12	17	that there is a	20
18	purpose of this paper	12	18	at the same time	19
19	purpose of this study	12	19	an important role in	18
20	the success of the	12	20	between customer satisfaction and	18
21	the united states and	12	21	is located in the	18
22	an example of this	11	22	customer satisfaction and customer	17
23	are more likely to	11	23	of the study the	17
24	as part of the	11	24	can be considered as	16
25	of this study was	11	25	one of the main	16
26	the end of the	11	26	satisfaction and customer loyalty	16
27	the creation of the	10	27	all over the world	15
28	the success of a	10	28	North Cyprus as a	15
29	a part of the	9	29	in the form of	14
30	an important part of	9	30	in the united states	14

By looking at Table 4.3, we can see that the most frequent bundles in tourism corpus are: *in the United States, one of the most, as well as the, and is one of the.*

Table 4.4: Top 30 Frequent 4-word Bundles in Civil Engineering Sub-corpus in Rank Order

Rank	Native Civil Eng.	Freq	Rank	Non-native Civil Eng.	Freq
1	as shown in figure	83	1	on the other hand	63
2	as well as the	56	2	is one of the	35
3	is shown in figure	50	3	one of the most	30
4	shown in figure the	47	4	in the construction industry	24
5	the results of the	44	5	as shown in figure	21
6	was found to be	43	6	used in this study	19
7	can be seen in	35	7	as a result of	18
8	it can be seen	33	8	can be seen in	18
9	it was found that	32	9	it can be seen	18
10	can be seen that	28	10	one of the main	18
11	it is important to	28	11	is shown in figure	17
12	in addition to the	24	12	of the most important	17
13	in order to determine	24	13	are shown in figure	16
14	be seen in figure	23	14	as it can be	15
15	the bottom of the	23	15	in this chapter the	15
16	the top of the	23	16	the percentage of the	15
17	the center of the	22	17	in the case of	14
18	can be used to	21	18	in the following sections	14
19	it should be noted	21	19	management in construction industry	14
20	the length of the	21	20	to be used in	14
21	the total number of	21	21	at the end of	13
22	in the number of	19	22	in the field of	13
23	were found to be	19	23	in this study the	13
24	should be noted that	18	24	shown in figure the	13
25	the accuracy of the	18	25	the compressive strength of	13
26	the distance to the	18	26	the other hand the	13
27	the use of the	18	27	in comparison with the	12
28	in the United States	17	28	is based on the	12
29	of the number of	16	29	the results of this	12
30	the behavior of the	16	30	in order to find	11

By looking at Table 4.4, we can see that the most frequent bundles in tourism corpus are: *as shown in figure, is shown in figure, shown in figure the, can be seen in, and it can be seen.*

Table 4.5: Top 30 Frequent 4-word Bundles in Mechanical Engineering Sub-corpus in Rank Order

Rank	Native Mechanic Eng.	Freq	Rank	Non-native Mechanic Eng.	Freq
1	can be used to	38	1	on the other hand	43
2	as shown in figure	35	2	is one of the	35
3	as well as the	23	3	as can be seen	33
4	is shown in figure	21	4	can be seen in	26
5	the top of the	21	5	is shown in figure	26
6	can be seen in	20	6	one of the most	25
7	the size of the	19	7	presented in table and	20
8	it is important to	18	8	the performance of the	20
9	as a function of	17	9	be seen in figure	19
10	the results of the	16	10	is shown in fig	16
11	as a result of	15	11	which is shown in	14
12	et al developed a	15	12	as shown in figure	13
13	shown in figure the	15	13	in this study the	13
14	is shown in fig	14	14	the aim of this	13
15	an example of the	13	15	the efficiency of the	13
16	are shown in figure	13	16	can be seen from	12
17	as a result the	13	17	in most of the	12
18	as seen in figure	13	18	as illustrated in figure	11
19	as well as a	13	19	at the end of	11
20	it was found that	13	20	is shown in fig	16
21	on the order of	13	21	which is shown in	14
22	figure provides an example	12	22	it is possible to	11
23	provides an example of	12	23	as a result the	10
24	the performance of the	12	24	is due to the	10
25	used in this study	12	25	is to investigate the	10
26	is dependent on the	11	26	it can be seen	10
27	is the number of	11	27	it is necessary to	10
28	it can be seen	11	28	shown in figure the	10
29	the accuracy of the	11	29	the effect of the	10
30	the bottom of the	11	30	the end of the	10

By looking at Table 4.5, we can see that the most frequent bundles in mechanical engineering corpus are: *as shown in figure, is shown in figure, can be seen in, shown in figure the, is shown in fig, as a result the, the performance of the, and it can be seen.*

4.2 Structural Categorization of Lexical Bundles

As mentioned earlier in previous chapters, Biber et al.'s taxonomy (1999) is used to structurally analyze the lexical bundles used in the two corpora. However, I made slight modifications to the other group (shown in Table 4.6). In Biber et al.'s taxonomy, lexical bundles which do not fit very well in the other groups are categorized as *other* expressions; in this analysis, two separate categories *other noun phrases* and *other verb fragments* are added to account for the noun phrases and verb phrases that did not fit in noun or verb phrase categories. The reason for this modification in this study was that in the especially the Iranian corpus, there were a number of lexical bundles that did not fit neatly in categories proposed by Biber et al. (1999) (refer to Table 2.1).

4.2.1 Structural Comparison of Lexical Bundles

In this section, the structural analysis of the lexical bundles used by both the Iranian L2 and the native English thesis writers will be presented across each sub-corpus separately (refer to Appendix A). Table 4.6 shows the structural analysis of the bundles across two sub-corpora of business theses.

Table 4.6 indicates that in total the Iranian students of business used three times more lexical bundles than the native English students of business studies. Moreover, the percentage of the lexical bundle use shows that the Iranian business master students used noun phrase-base (36.14%) and preposition phrase-based (24.61%) bundles more than other categories. The native English writers also used more noun phrase-based

(34.8%) and prepositional phrase-based bundles (27.91%) than the other bundles in their theses. Figure 4.1 shows this difference better.

Table 4.6: Structural Comparison of Bundles in Business Master Theses of Native English and Iranian Students

<i>Category</i>	<i>Pattern</i>	<i>Type</i>	<i>Token</i>	<i>Percentage (%)</i>	
<i>NP-based</i>	noun phrase with post-	of	14/27	118/218	32.48/20.76
	modifier fragment	other	0/12	0/77	0/9.23
	other noun phrase		1/8	20/51	2.32/6.15
<i>PP-based</i>	preposition + noun	of	5/6	52/59	11.67/4.61
	phrase fragment	other	7/26	96/275	16.24/20
<i>VP-based</i>	anticipatory it + V/Adj		3/2	27/14	6.96/1.53
	passive verb+ PP fragment		0/12	0/83	0/9.23
	copula be + N/Adj phrase		2/6	15/83	4.64/4.61
	pronoun/NP + be		0/4	0/29	0/3.07
	other verb fragments		0/6	0/35	0/4.61
<i>Clause-based</i>	V/Adj+ to-clause fragment		4/6	40/36	9.28/4.61
	VP+ that-clause fragment		2/5	13/51	4.64/3.84
	adverbial clause fragment		0/2	0/29	0/1.53
<i>Others</i>	other expressions		5/8	45/58	11.6/6.15
<i>Total</i>			43/130	426/1098	100/100

Note: the numbers display the percentage of each bundle category, the first number before each slash represents the percentage of lexical bundle use by the native English writers and the second number after each slash shows the percentage of bundle use by the Iranian L2 writers. NP: noun phrase, PP: preposition phrase, VP: verb phrase.

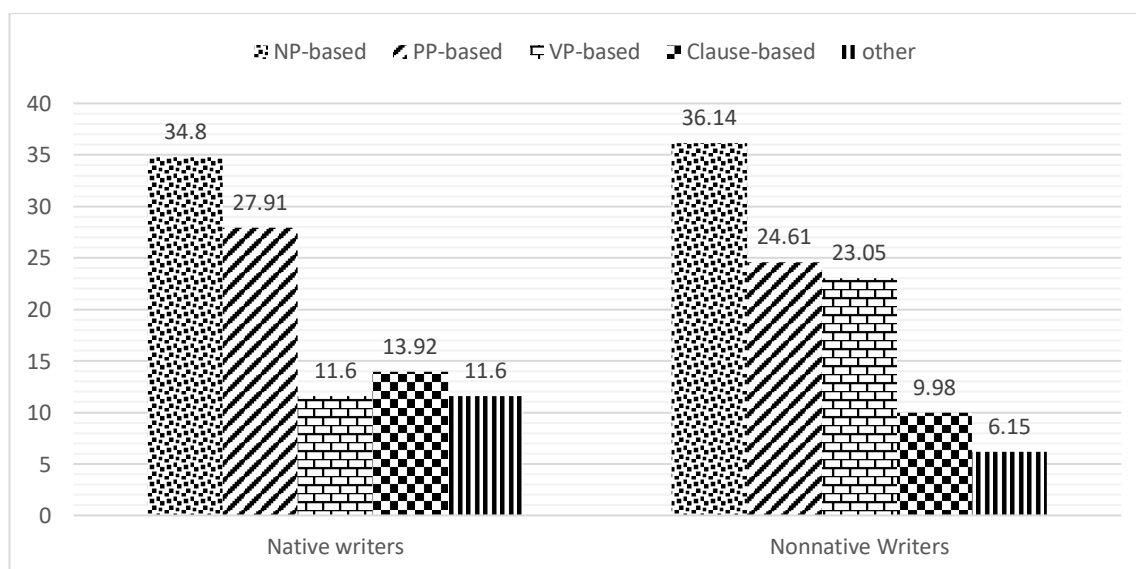


Figure 4.1: Structural Analysis of 4-word Lexical Bundles in Business Sub-corpus

Table 4.7: Structural Comparison of Bundles in Tourism Master Theses of Native English and Iranian Students

<i>Category</i>	<i>Pattern</i>	<i>Type</i>	<i>Token</i>	<i>Percentage (%)</i>
<i>NP-based</i>	noun phrase with post- of	20/28	223/391	29.41/29.16
	modifier fragment other	0/6	0/75	0 / 6.25
	other noun phrase	5/4	39/70	7.35 / 4.16
<i>PP-based</i>	<i>preposition + noun of</i>	4/12	39/288	5.88 / 12.50
	<i>phrase fragment other</i>	9/18	175/294	13.23 /18.75
<i>VP-based</i>	anticipatory it + V/Adj	4/1	72/11	5.88/1.04
	passive verb+ PP fragment	1/4	6/50	1.47/4.16
	copula be + N/Adj phrase	1/2	18/89	1.47/2.08
	pronoun/NP + be	3/3	26/35	4.41/3.125
	other verb fragments	1/4	7/63	1.47/ 4.16
<i>Clause-based</i>	V/Adj+ to-clause fragment	8/2	80/20	11.76/2.08
	VP+ that-clause fragment	3/3	37/42	4.41/3.12
	adverbial clause fragment	0/2	0/20	0/2.08
<i>Others</i>	other expressions	9/7	88/134	13.23/7.29
<i>Total</i>		68/96	810/1582	100/100

Likewise, the structural analysis of the lexical bundles in the Iranian and the native English students' tourism theses indicated that the Iranian L2 students used more bundles than the native English students. As Table 4.7 indicates, the Iranian students used noun phrase-based and prepositional phrase-based lexical bundles more extensively (39.57% and 31.25 %, respectively) than the other lexical bundles. Similarly, the structural analysis of native English master theses showed that noun phrase-based and prepositional phrase-based categories of bundles were the most frequently used ones (36.76 % and 19.11 %, respectively). Figure 4.2 shows the difference.

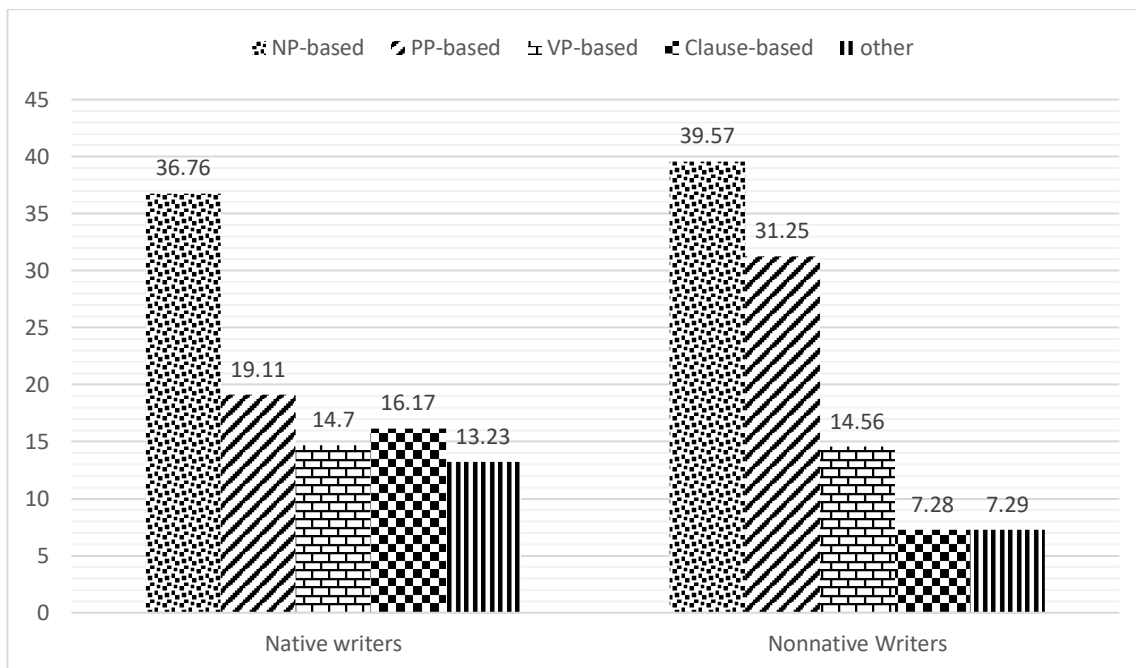


Figure 4.2: Structural Analysis of 4-word Lexical Bundles in Tourism Sub-corpus

In the next sections, the lexical bundles used in the theses of hard sciences (mechanical and civil engineering) will be compared and analyzed structurally.

Besides, the percentage of lexical bundle use showed that verb phrase-based, preposition phrase-based and noun phrase-based categories (42.12 %, 19.97 %, 18.56 %, respectively) were the most commonly used ones by the Iranian master students. However, in the native English students' theses, verb phrase-based, noun phrase-based, preposition phrase-based categories were the most commonly used categories of lexical bundles (33.67 %, 31.51 %, and 20.64 %, respectively). Figure 4.3 displays the difference between native and the Iranian L2 students better.

Table 4.8: Structural Comparison of Bundles in Mechanical Engineering Master Theses of Native English and Iranian Students

<i>Category</i>	<i>Pattern</i>	<i>Type</i>	<i>Token</i>	<i>Percentage (%)</i>	
<i>NP-based</i>	noun phrase with post-modifier	of	27/22	29.34/15.71	
	fragment	other	2/3	2.17/2.14	
	other noun phrase		0/1	0/0.71	
<i>PP-based</i>	preposition + noun phrase	of	11/12	11.95/8.57	
	fragment	other	8/16	8.69/11.42	
<i>VP-based</i>	anticipatory it + V/Adj		7/10	7.60/7.14	
	passive verb+ PP fragment		16/32	17.39/22.85	
	copula be + N/Adj phrase		4/8	4.34/5.71	
	pronoun/NP + be		0/3	0/2.14	
	other verb fragments		4/6	4.34/4.28	
<i>Clause-based</i>	V/Adj+ to-clause fragment		3 / 4	3.26/2.85	
	VP+ that-clause fragment		2/0	2.17/0	
	adverbial clause fragment		3/4	3.26/2.85	
<i>Others</i>	other expressions		5/19	5.43/13.57	
<i>Total</i>			92/140	988/1206	100/100

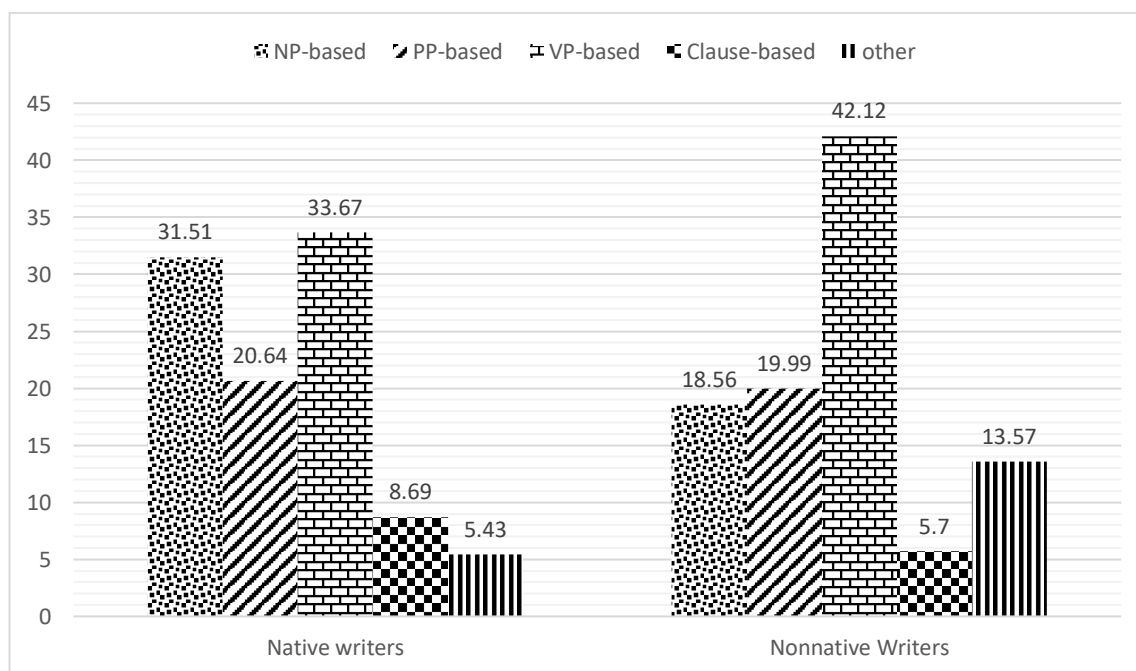


Figure 4.3: Structural Analysis of 4-word Lexical Bundles in Mechanical Engineering Sub-corpus

Table 4.9: Structural Comparison of Bundles in Civil Engineering Master Theses of Native English and Iranian Students

<i>Category</i>	<i>Pattern</i>	<i>Type</i>	<i>Token</i>	<i>Percentage (%)</i>	
<i>NP-based</i>	noun phrase with post- modifier	of	37/16	538/176	26.24/17.77
	fragment	other	3/2	45/23	2.12/2.22
	other noun phrase		6/1	59/7	4.25/1.11
<i>PP-based</i>	preposition + noun	of	16/9	207/97	11.34/10
	phrase fragment	other	15/23	179/294	10.63/25.55
<i>VP-based</i>	anticipatory it + V/Adj		8/6	159/57	5.67/6.66
	passive verb+ PP fragment		16/14	294/159	11.34/15.55
	copula be + N/Adj phrase		2/1	21/35	1.41/1.11
	pronoun/NP + be		0/0	0/0	0/0
	other verb fragments		0/0	0/0	0/0
<i>Clause-based</i>	V/Adj+ to-clause fragment		17/3	249/30	12.05/3.33
	VP+ that-clause fragment		11/2	150/16	7.80/2.22
	adverbial clause fragment		4/2	114/36	2.83/2.22
<i>Others</i>	other expressions		6/11	122/100	4.25/12.22
<i>Total</i>			141/90	2137/1030	100/100

The structural analysis of the lexical bundles in the civil engineering theses showed that unlike the other sub-corpora, the native English students used more lexical bundles than the Iranian master students. Moreover, except in the prepositional phrase-based category, the native English master students used more lexical bundles in all the other categories. However, the word counts in the two corpora might have had some influence in this regard which requires further research. In addition, unlike the other sub-corpora, the percentage of lexical bundles use revealed that the Iranian master students used prepositional phrase-based, verb phrase-based, and noun phrase-based categories extensively (35.55 %, 23.32 %, and 21.10 %, respectively). However, the native English students used noun phrase-based, clausal-based and prepositional phrase-based categories of lexical bundles the most (32.61 %, 22.68% and 21.97 %, respectively) in their theses (see Figure 4.4).

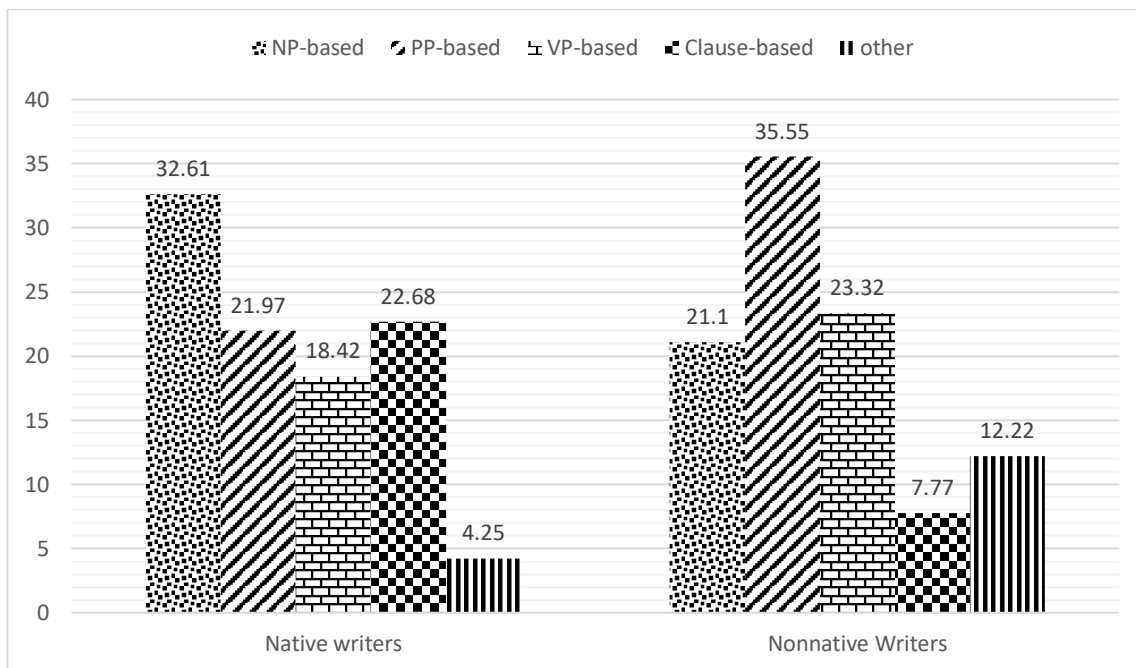


Figure 4.4: Structural Analysis of 4-word Lexical Bundles in Civil Engineering Sub-corpus

Overall, the structural analysis of lexical bundle use in the two corpora showed that the Iranian master students used more lexical bundles than the native English students except in the civil engineering sub-corpora. Moreover, some of the categories of lexical bundles such as noun phrase-based and prepositional phrase-based categories were the mostly used lexical bundles by the Iranian L2 and the native English writers.

4.3 Functional Categorization of Lexical Bundles

To functionally analyze the four-word lexical bundles in the two corpora, Hyland's (2008a) functional taxonomy was used. The reason for the adoption of this model was its suitability for academic writing. (refer to Table 2.3).

4.3.1 Functional Analysis of Lexical Bundles

In this section, the four-word lexical bundles used in the two corpora will be functionally analyzed (refer to Appendix B). First, the lexical bundles used in the two soft sub-corpora are analyzed. Table 4.10 displays the functional analysis of four-word lexical bundles in business sub-corpus.

Table 4.10: Functional Analysis of 4-word Bundles in Business Master Theses of Native English and Iranian Students

Category	Function	Type	Token	Percentage (%)
Research-oriented	<i>Location</i>	3/8	39/59	6.81/6.01
	<i>Procedure</i>	6/7	45/37	13.63/5.26
	<i>Quantification</i>	7/11	60/121	15.90/8.27
	<i>Description</i>	4/12	33/77	9.09/9.02
	<i>Topic</i>	4/7	70/59	9.09/5.26
Text-oriented	<i>Transition signals</i>	5/14	47/158	11.36/10.52
	<i>Resultative signals</i>	3/24	31/190	6.81/18.04
	<i>Structuring signals</i>	1/13	6/90	2.27/9.77
	<i>Framing signals</i>	4/18	39/163	9.09/13.53
Participant-oriented	<i>Stance features</i>	6/10	56/88	13.63/7.51
	<i>Engagement features</i>	1/9	6/79	2.27/6.76
	<i>Total</i>	44/133	432/1121	100/100

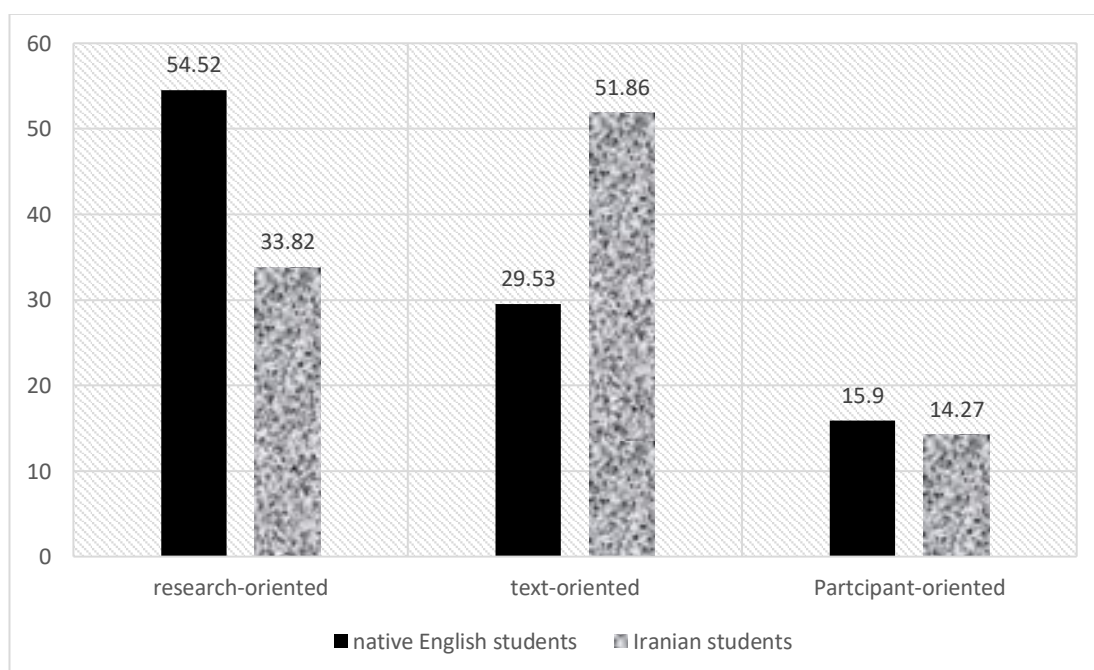


Figure 4.5: Functional Analysis of 4-word Lexical Bundles in Business Sub-corpus

Table 4.10 shows that the Iranian writers used more lexical bundles than the native English writers in the business corpus. The functional analysis, however, indicates that the native English writers used research-oriented bundles (54.52 %) the most while the Iranian students used text-oriented bundles the most in their theses (51.86 %).

In the following section, soft science corpus of tourism theses will be functionally analyzed. Table 4.11 displays the frequency and percentage of each functional category.

Table 4.11: Functional Analysis of 4-word Bundles in Tourism Master Theses of Native English and Iranian Students

Category	Function	Type	Token	Percentage (%)
Research-oriented	<i>Location</i>	4/7	38/137	5.79/7.07
	<i>Procedure</i>	9/8	94/116	13.04/8.08
	<i>Quantification</i>	6/13	82/289	8.69/13.13
	<i>Description</i>	11/10	108/110	15.94/10.10
	<i>Topic</i>	12/17	194/232	17.39/17.17
Text-oriented	<i>Transition signals</i>	3/5	49/127	4.34/5.05
	<i>Resultative signals</i>	3/13	37/165	4.34/13.13
	<i>Structuring signals</i>	8/7	65/108	11.59/7.07
	<i>Framing signals</i>	2/10	12/160	2.89/10.10
Participant-oriented	<i>Stance features</i>	5/7	41/148	7.24/7.07
	<i>Engagement features</i>	6/2	86/22	8.69/2.02
	<i>Total</i>	69/99	806/1614	100/100

Similar to the previous sub-corpus findings, Table 4.11 indicates that the Iranian writers used more bundles than the native English writers in the tourism corpus. The functional analysis, further, revealed that both the native English writers and Iranian writers used research-oriented bundles the most (60.85 % and 55.55 % respectively). Moreover, the Iranian writers used more text-oriented bundles than the native English writers (35.35 % and 23.16 %, respectively). Figure 4.6 shows the difference in each category.

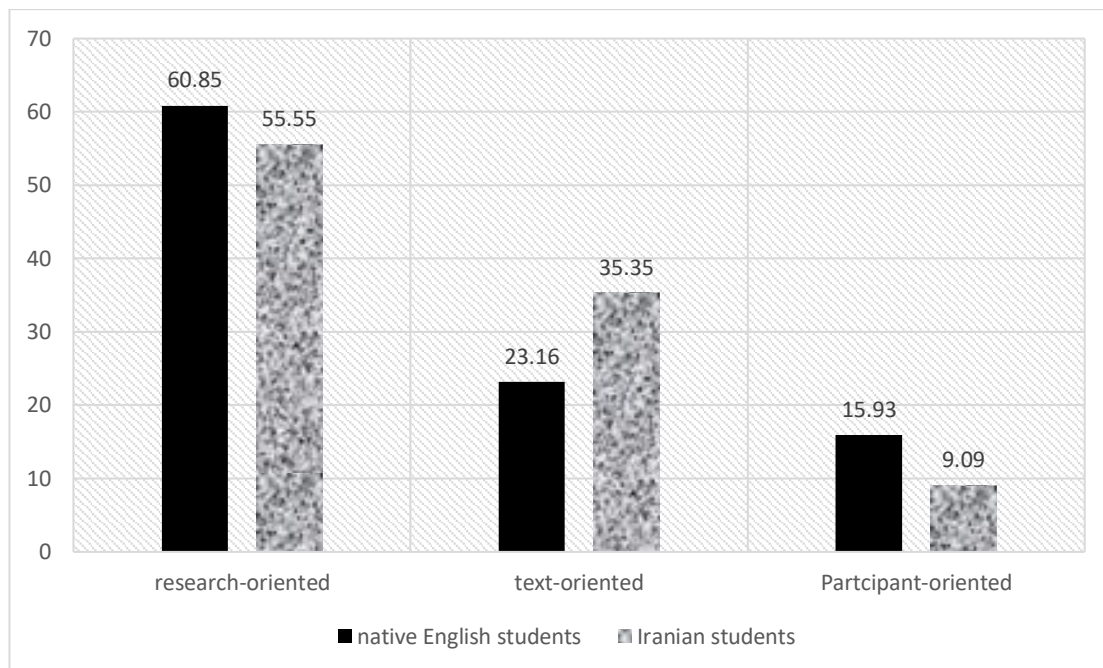


Figure 4.6: Functional Analysis of 4-word Lexical Bundles in Tourism Sub-corpus

Next, the two hard-science corpora will be functionally analyzed. First, the functional analysis of civil engineering sub-corpus will be presented. Table 4.12 shows the percentage and frequency of use of each functional category of lexical bundle in the civil engineering sub-corpus.

Table 4.12: Functional Analysis of 4-word Bundles in Civil Engineering Master Theses of Native and Iranian Students

Category	Function	Type	Token	Percentage (%)
Research-oriented	<i>Location</i>	11/7	167/69	7.58/7.77
	<i>Procedure</i>	15/5	199/48	10.34/5.55
	<i>Quantification</i>	19/6	231/104	13.10/6.66
	<i>Description</i>	14/2	173/22	9.65/2.22
	<i>Topic</i>	11/9	123/100	7.58/10
Text-oriented	<i>Transition signals</i>	6/5	119/105	4.13/5.55
	<i>Resultative signals</i>	23/13	375/119	15.86/14.44
	<i>Structuring signals</i>	18/17	358/203	12.41/18.88
	<i>Framing signals</i>	9/12	110/112	6.20/13.33
Participant-oriented	<i>Stance features</i>	1/2	11/26	0.68/2.22
	<i>Engagement features</i>	18/12	296/122	12.41/13.33
	<i>Total</i>	145/90	2162/1030	100/100

Table 4.12 shows that unlike the other sub-corpora, the native English writers used more lexical bundles than the Iranian L2 writers in the civil engineering corpus. The functional analysis also indicated that the native English writers used research-oriented bundles the most (48.25 %) while the Iranian L2 writers used text-oriented bundles the most (52.2 %). Figure 4.7 shows the difference in each category.

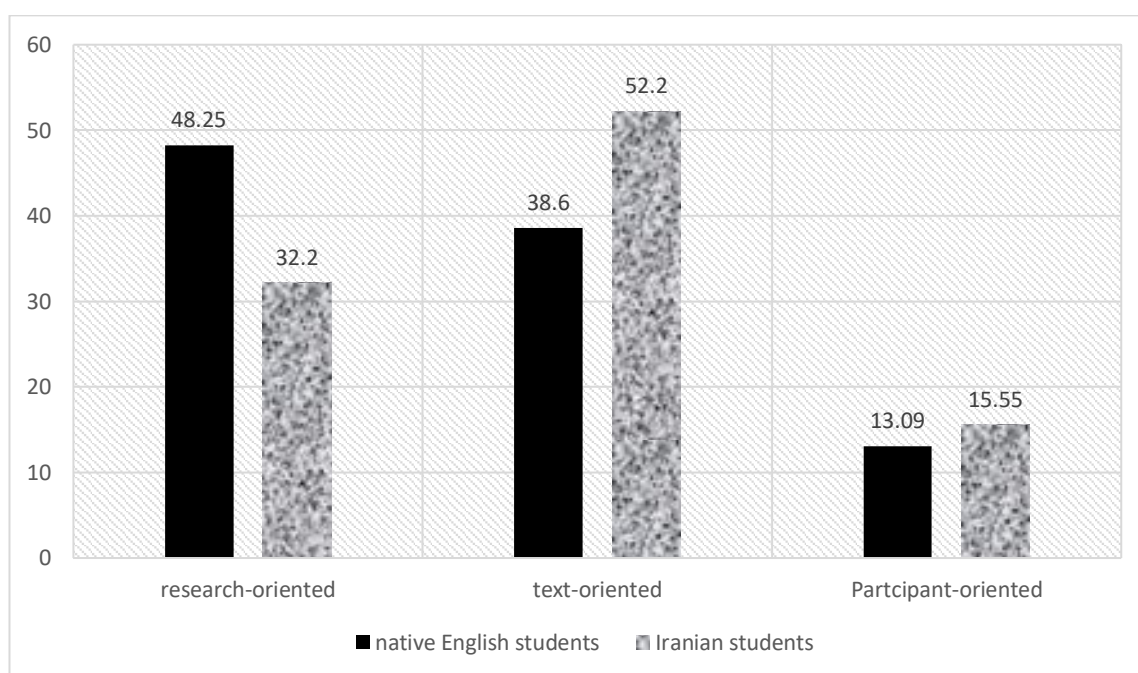


Figure 4.7: Functional Analysis of 4-word Lexical Bundles in Civil Engineering Sub-corpus

Finally, Table 4.13 displays the frequency and percentage of functional category use. As the table demonstrates, similar to the findings of the soft sciences, the Iranian L2 writers used more functional categories of lexical bundles. The percentage of functional category use shows that both native and the Iranian L2 writers used text-oriented bundles the most (47.8 % and 53.93 %, respectively). Figure 4.8 shows the difference between the three categories.

Table 4.13: Functional Analysis of 4-word Bundles in Mechanical Engineering Master Theses of Native and Iranian Students

Category	Function	Type	Token	Percentage (%)
Research-oriented	<i>Location</i>	7/6	79/49	7.60/4.31
	<i>Procedure</i>	13/11	147/75	14.13/7.91
	<i>Quantification</i>	11/16	87/155	11.95/11.51
	<i>Description</i>	8/6	89/56	8.69/4.31
	<i>Topic</i>	1/7	7/44	1.08/5.03
Text-oriented	<i>Transition signals</i>	4/7	53/89	4.34/5.03
	<i>Resultative signals</i>	17/22	161/148	18.47/15.82
	<i>Structuring signals</i>	16/40	199/352	17.39/28.77
	<i>Framing signals</i>	7/6	74/45	7.60/4.31
Participant-oriented	<i>Stance features</i>	1/6	8/42	1.08/4.31
	<i>Engagement features</i>	7/12	84/141	7.60/8.63
	<i>Total</i>	92/139	988/1196	100/100

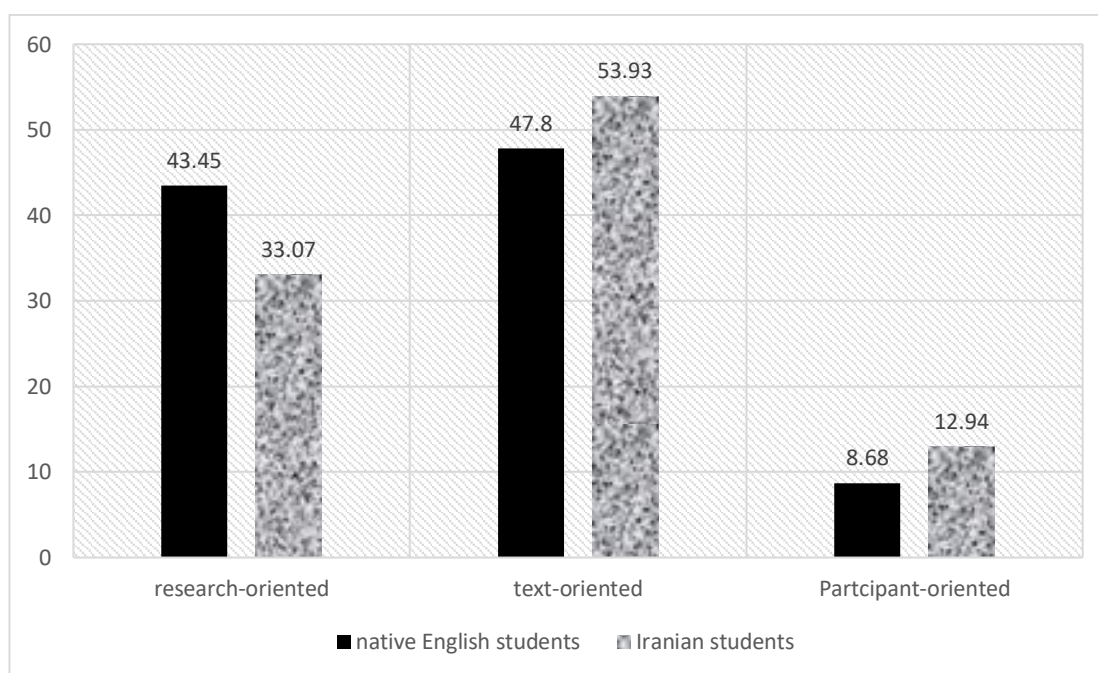


Figure 4.8: Functional analysis of 4-word Lexical Bundles in Mechanical Engineering Sub-corpus

4.4 Summary

In this chapter, the frequency, structural and functional analysis of lexical bundles used in two corpora, namely native English and the Iranian L2 writers' theses corpora, were

presented. First, the frequency of the most commonly used four-word lexical bundles used in each corpus was analyzed and 30 most commonly used ones were presented. Then, the four-word lexical bundles identified in the two corpora were structurally analyzed on the basis of Biber et al.'s (1999) structural taxonomy. Finally, the lexical bundles were functionally analyzed according to Hyland's (2008a) functional model. In the next chapter, the findings presented in this chapter will be discussed in detail.

Chapter 5

DISCUSSION AND CONCLUSION

In this chapter, the findings of the study consisting of three sections of frequency, structural and functional analyses of four-word lexical bundles are discussed and compared with the results of the previous studies. Later, the conclusion is presented and then the limitations of the study are described. Finally, the implications and the suggestions for future studies are discussed.

5.1 Discussion

The study investigated the frequency, structure and function of lexical bundles by comparing and contrasting two corpora of master theses in hard sciences, i.e., mechanical engineering (ME), and civil engineering (CE) and soft sciences, i.e., business (BS) and tourism (TR), one corpus representing the Iranian L2 English writers and the other one representing English native writers.

To this end, the three research questions addressed in the present study will be answered one by one. First the frequency of the bundles will be reviewed and compared with relevant studies in the literature. In the second section, the results of structural analysis of lexical bundles will be discussed. Finally, the results of functional analysis of bundles will be elaborated.

5.1.1 Frequency of Bundles

In this section, I address the first research question of our study which is related to the frequency of most common bundles and the similarities and differences in the two corpora. The following findings were obtained as a result of the analysis:

In comparison to their native counter parts, the Iranian master students relied more on four-word lexical bundles in their writings in terms of token and type. In this regard, the results confirmed the findings of Alipour and Zarea (2013), Amirian et al. (2013), Bychkovska and Lee (2017), Hsu et al. (2017), Jalali et al. (2008), Pang (2009) and Pérez-Llantada (2014), who found that nonnative English speakers used more lexical bundles than native English speakers.

In soft sciences (business and tourism), the Iranian L2 students used more four-word lexical bundles than the native English students. However, in hard sciences (civil and mechanical engineering), the case was somehow different, as in the civil engineering sub-corpus, English writers used more bundles. This heavy reliance on lexical bundles use nonnative English students can be rooted in the instruction that these students usually receive in academic writing courses. That is, formulaic language can be overused, underused or misused by nonnative English students due to usually impoverished language input (Schmitt & Carter, 2004).

The results of the study on civil engineering master students' corpus are also in line with the findings of some other studies which found that English native students used more lexical bundles than nonnative English students (e.g., Ädel & Erman, 2012; Karabacak & Qin, 2013). Our results also confirmed the findings of Esfandiari and

Barbary (2017) and Safarzadeh et al. (2012), who found that the Iranian writers used less lexical bundles than the native English writers.

Moreover, by looking at the 30 most frequent four-word lexical bundles in each sub-corpus, I found that only a few bundles were commonly used between the two groups' writing. This can show the discrepancy between the native and the Iranian students' writing.

5.1.2 Structural Analysis of Lexical Bundles

Before I start this section, it is worth mentioning that the model used to analyze the structural characteristics of the two corpora is Biber et al.'s (1999) structural taxonomy of lexical bundles. In this taxonomy, lexical bundles are classified into five structural categories: NP-based (including *NP with of-phrase fragment* and *NP with other postmodifier*), PP-based (including *PP with embedded of-phrase fragment* and *PP with other post modifier*), VP-based (including *Anticipatory it + VP/Adj phrase*, *Passive verb + PP fragment*, *Copula be + N/Adj phrase*, and *Pronoun/NP + be*), clause-based (including *V/Adj + to-clause fragment*, *VP+ that-clause fragment*, and *Adverbial clause fragment*), and other expressions (including expressions such as *as well as the* and *than that of the*). It should be mentioned that with regard to the four-word bundles in this study that do not fit very well in other categories above, two new structural subcategories (*NP* and *VP*) were created. In what follows, the findings of our second research question will be discussed.

In soft science fields (business and tourism), both the native English and the Iranian L2 writers preferred to employ NP-based and PP-based categories respectively in their theses more than the other structural categories. This finding somewhat confirmed the

findings of Hyland's (2008a) and Jalali et al.'s (2008) studies which reported that NP-based was the most frequently used lexical bundles in the writing of the L1 Persian postgraduate students.

In the hard science corpora (civil and mechanical engineering), Iranian mechanical engineering writers relied more on VP-based, PP-based and NP-based categories respectively. However, the native English students did not follow the same trend and they preferred to use VP-based, NP-based and PP-based categories respectively the most in their writings.

In addition, in the civil engineering sub-corpus, the Iranian L2 writers used PP-based, VP-based and NP-based categories respectively more than the other categories of bundles. This affirms Esfandiari and Barbary's (2017) results that the Iranian writers relied more on prepositional phrase-based lexical bundles in their writing. On the other hand, the native English students extensively applied NP-based, clausal-based and PP-based categories respectively in their writings.

5.1.3 Functional Analysis of Lexical Bundles

This section answers the third question of the present study which is concerned with the functional features of the bundles in each corpus and their differences and similarities in the two corpora of the native English and the Iranian L2 writers. It should be stated here that the functional analysis draws on Hyland's (2008a) functional taxonomy of lexical bundles which consists of three major functions of language: research-oriented bundles which are related to the ideational function of language such as *location*, *procedure*, *quantification*, *description*, and *topic*, text-oriented bundles that are concerned with textual functions or the text organization which include

transition signals, resultative signals, structuring signals, and framing signals, and participant-oriented bundles which are related to interpersonal functions which include stance features, and engagement features.

The findings to the third question can be summarized as follows: the Iranian L2 students showed more interest in the use of text-oriented bundles particularly structuring and resultative signals in hard sciences (civil and mechanical engineering) and business sub-corpus, which affirms Esfandiari and Barbary's (2017) results that the Iranian writers used text-oriented bundles the most and participant-oriented clusters the least in their published articles of psychology. However, in the tourism sub-corpus, it was found that the Iranian L2 writers preferred to use research-oriented bundles more than other categories, which is in line with the findings of Amirian et al. (2013) and Jalaliet al. (2008), who reported that L1 Iranian students used research-oriented clusters more extensively than the other clusters. This also supports the results from Hyland (2008a) and Wei and Lei (2011), who found that Chinese master students' theses contained the most research-oriented, and the least text-oriented and participant-oriented bundles.

The native English writers, on the other hand, used research-oriented bundles the most in three sub-corpora, however, in the mechanical engineering sub-corpus, they relied more on text-oriented bundles.

5.2 Conclusion

This study aimed to explore the frequency, structure and functions of four-word lexical bundles by comparing and contrasting two corpora of master theses across four different fields of study (business and tourism as soft sciences and mechanical

engineering and civil engineering as hard sciences), one corpus representing Iranian L2 English writers and the other one representing English native writers. In this study there were disciplinary variations across these fields in terms of frequency, structure and functions of lexical bundle use.

The findings of the study also revealed that in soft sciences (business and tourism) and in one of the hard sciences (mechanical engineering), the Iranian L2 writers used more lexical bundles than their native English counterparts. The difference was statistically significant in business sub-corpus as the Iranian writers used three times more lexical bundles than English writers (130 to 43 in terms of types). However, the highest number of lexical bundles were reported to be used in civil engineering sub-corpus by the native English writers.

Moreover, structural differences were reported in each sub-corpus, that is, in soft science fields (business and tourism), the Iranian L2 writers employed NP-based, PP-based, and VP-based categories respectively the most while native writers were found to be more interested in using NP-based, PP-based and clause-based bundles. However, in the hard sciences (civil and mechanical engineering), it was found that the Iranian mechanical engineering writers preferred to use VP-based, PP-based and NP-based categories respectively the most, while native writers relied on VP-based, NP-based and PP-based categories the most. Further, in the civil engineering sub-corpus, the Iranian L2 writers applied PP-based, VP-based and NP-based categories respectively more than the other categories of bundles while in native theses NP-based, clausal-based and PP-based categories were respectively employed the most. Such variations in each sub-corpus is an indication of writers' reliance on different linguistic

devices to persuade their readers, develop their arguments, and establish credibility and by relying on different structural patterns.

Finally, the results indicated functional differences across the four different sub-corpora. That is, in both hard-science (civil and mechanical engineering) corpora and in one of the soft science fields (i.e., business), the Iranian writers extensively relied on text-oriented bundles to express themselves while in tourism sub-corpus, they tended to use research-oriented bundles the most. However, the native English writers, used research-oriented bundles the most in three sub-corpora, however, in the mechanical engineering sub-corpus, the native English writers employed more text-oriented bundles. Such a difference in functional use of bundles can indicate how the Iranian L2 writers resorted to various textual devices to develop their argument, to express their voice and stances as well as their identity as master students. Such functional variation can also show their different proficiency levels and textual experiences as well. A number of other issues such as interlanguage transfer, instructional differences, lack of rhetorical confidence, conventions and norms, limited vocabulary knowledge and learners' strategic differences have also been reported as reasons for discrepancies between the native English and nonnative English graduate writers (Li, 2016), which the researcher thinks could have influenced the discrepancy between the Iranian L2 and the native English writers' use of lexical bundles in this study as well.

5.3 Implications

This study can have several implications for writers of different disciplines to become familiar with different conventions and norms governing different fields of study. That is, bundles are used differently in different fields of study which makes its

identification worthwhile in English for academic purposes. Thus, materials developers and course designers can take this point into account by developing and designing appropriate learning tasks and lessons for second language speakers of different disciplines. Moreover, the students of the selected disciplines can also benefit from the results of this study as mastery of these expressions can lead to their academic success (Biber & Barbieri, 2007).

Furthermore, the findings of the present study provide some implications for genre-based pedagogies in ESP in which ESP teachers and students are presented with a different view of writing and a distinct set of teaching practices (Hyland, 2003). According to genre-based pedagogies, students are explicitly taught how to write by focusing on the structural and lexico-grammatical patterns of the target genre. Thus, ESP teachers should raise the students' awareness over the target genres by teaching the students explicitly how target genres are structured linguistically (Hyland, 2003). In this regard, lexical bundles or formulaic language can act as linguistic devices which can facilitate writing skills of nonnative writers. Thus, by adopting an ESP genre-based approach in their graduate level academic writing course (Işık-Taş, 2018), ESP teachers can raise the awareness of nonnative students over these rhetorical conventions and help expand their word knowledge of multiword combinations.

5.4 Limitations

Although the study has mostly achieved its aims, there were some limitations based on the methodological concerns. One of the limitations was the number of theses used from each discipline. The lack of access to the Iranian theses written in English inside of Iran was the real motivation to search for theses written by the Iranian students elsewhere. Thus, I focused on the Iranian students' theses written at Eastern

Mediterranean University in Northern Cyprus which considerably limited our data to theses written by Iranian master students at the respective university.

Another limitation was the study level from which the theses were selected, that is, the data were all selected from master theses which is a different genre than bachelor or PhD study level. This can be a limiting factor as master students usually possess the accumulated knowledge of bachelor years but lack the theoretical and practical expertise or knowledge of PhD students.

The number of words from each discipline of study was another limitation as access to both the Iranian and the native English theses from the selected universities was somehow limited. Thus, cautions should be taken not to generalize the findings of this study to a bigger context.

The last but not the least, the cut-off frequency set for this study was relatively high, which limited the number of bundles selected for the study and hence excluded a number of low frequency valuable lexical bundles.

5.5 Suggestions for Further Studies

Parallel to the study's limitations, some suggestions for further research are made here. In future studies, other disciplines can be selected, and for more reliable data, the researcher suggests focusing on one study discipline at a time. In our study making generalization and interpretations across discipline was somehow difficult as in some of the areas conflicting findings were reported. Furthermore, other sciences should also be investigated to see variations in lexical bundles used by the Iranian L2 students.

Moreover, later studies can work on bachelor and PhD students' writings such as theses, articles and projects by gathering data from other sources such as interviews and ethnographic studies to trace the development and acquisition of such valuable bundles over time. Such studies can also trace the difficulties faced by the Iranian L2 writers and the possible sources of errors that the Iranian students make in that regard.

In addition, the differences in word-length should also be taken into account as some of the thesis were either overlength or under-length. For later research, corpora of similar size should be selected, although control over the word length is not entirely possible.

Finally, three-word and five-word lexical bundles also deserve special attention as in this study only four-word lexical bundles were investigated. It is worth mentioning that identification of multiple-word bundles can provide valuable language resources for nonnative English writers and also facilitate their language production. Furthermore, the cut-off frequency can be lowered to include less-frequently occurring lexical bundles in later research.

5.6 Summary

In this chapter, the findings of the study were discussed. First, the discussion section was presented by focusing on frequency, structure and function of the bundles. Later, the conclusion was made, and then the study implications were described. Finally, the limitations and suggestions for further research were presented.

REFERENCES

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and nonnative speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92.
- Alipour, M., & Zarea, M. (2013). A disciplinary study of lexical bundles: The case of native versus nonnative corpora. *Taiwan International ESP Journal*, 5(2), 1-20.
- Altenberg, B. (1993). Recurrent word combinations in spoken English. In J. M. D'Arcy (Ed.), *Proceedings of the Fifth Nordic Association for English Studies Conference* (pp. 17-27). Reykjavik, Iceland: University of Iceland.
- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 101-122). Oxford, United Kingdom: Oxford University Press.
- Amirian, Z., Ketabi, S., & Eshaghi, H. (2013). The use of lexical bundles in native and nonnative post-graduate writing: The case of applied linguistics MA theses. *Journal of English Language Teaching and Learning*, 5(11), 1-29.
- Anthony, L. (2018). Antconc 3.5.2: A free text analysis software. Available from <http://www.antlab.sci.waseda.ac.jp/>

- Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh, United Kingdom: Edinburgh University Press.
- Bannard, C., & Lieven, E. (2012). Formulaic language in L1 acquisition. *Annual Review of Applied Linguistics*, 32(1), 3-16.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Philadelphia, PA: John Benjamins Publishing Company.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.
doi:10.1016/j.esp.2006.08.003
- Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: an initial taxonomy. In A. Wilson, P. Rayson & T. McEnery (Eds.), *Corpus linguistics by the Lune: a festschrift for Geoffrey Leech* (pp. 71–93). Frankfurt: Peter Lang.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
doi:10.1093/applin/25.3.371
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.

- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London, United Kingdom: Longman.
- Bychkovska, T., & Lee, J. J. (2017). At the same time: lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes, 30*, 38-52.
- Chen, L. (2010). An investigation of lexical bundles in ESP textbooks and electrical engineering introductory textbooks. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 107-125). London, United Kingdom: Continuum International Publishing Group.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30-49.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423. doi:10.1016/j.esp.2003.12.001
- Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education, 17*, 391-406.

- Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12(1), 33-43. doi:10.1016/j.jeap.2012.11.002
- Cowie, A. P. (1992). Multiword lexical units and communicative language teaching. In P. J. L. Arnaud, & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 1-12). Basingstoke: Macmillan.
- Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. *International Journal of Corpus Linguistics*, 19(4), 443-477.
- Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 38(2), 165-193.
- Ellis, N.C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 24, 143-188.
- Ellis, N. C., Simpson-Vlach, R. I. T. A., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375-396.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *TEXT*, 20(1), 29-62.

- Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes*, 29, 21-42.
- Firth, J. R. (1957). Modes of meaning. In J. R. Firth (Ed.), *Papers in linguistics 1934-1951* (pp. 190-215). London, United Kingdom: Oxford University Press.
- Gray, B., & Biber, D. (2013). Corpus approaches to the study of discourse. In K. Hyland & B. Paltridge (Eds.), *The Bloomsbury companion to discourse analysis* (pp. 138-152). London, United Kingdom: Bloomsbury Academic.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London, United Kingdom: Edward Arnold.
- Haswell, R. (1991). *Gaining ground in college writing: Tales of development and interpretation*. Dallas: Southern Methodist University Press.
- Hsu, J. W. K., Chen, H. J. H., Yang, T. Y. C., & Liu, C. Y. (2017). A study on the use of lexical bundles in native and nonnative English academic writing. *English Teaching & Learning*, 41(2), 55-98.
- Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, 12(1), 17-29.

- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192.
- Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62. doi:10.1111/j.1473-4192.2008.00178.x
- Hyland, K. (2008b). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21. doi:10.1016/j.esp.2007.06.001
- Hyland, K. (2008c). Persuasion, interaction and the construction of knowledge: Representing self and others in research writing. *International Journal of English Studies*, 8(2), 1-23.
- Hyland, K. (2009). *Academic discourse: English in a global context*. London, United Kingdom: Continuum.
- Işık-Taş, E. E. (2018). Authorial identity in Turkish language and English language research articles in Sociology: The role of publication context in academic writers' discourse choices. *English for Specific Purposes*, 49, 26-38.
- Jalali, H. (2017). Reflection of stance through it bundles in applied linguistics. *Ampersand*, 4, 30-39.

- Jalali, H., Eslami Rasekh, A., & Tavangar Rizi, M. (2008). Lexical bundles and intradisciplinary variation: The case of applied linguistics. *Iranian Journal of language Studies*, 2(4), 447-484.
- Jalali, H., & Ghayoomi, S. (2010). A comparative qualitative study of lexical bundles in three academic genres of applied linguistics. *Modern Journal of Applied Linguistics*, 2(4), 323-333.
- Kennedy, G. (2014). *An introduction to corpus linguistics*. London, United Kingdom: Routledge.
- Lewis, M. (2008). *Implementing the lexical approach: Putting theory into practice*. London, United Kingdom: Heinle Cengage Learning.
- Li, L. (2016). *Sentence initial bundles in L2 thesis writing: A comparative study of Chinese L2 and New Zealand L1 postgraduates' writing* (Doctoral dissertation). University of Waikato, New Zealand.
- Karabacak, E., & Qin, J. (2013). Comparison of lexical bundles used by Turkish, Chinese, and American university students. *Procedia-Social and Behavioral Sciences*, 70(0), 622-628. doi:10.1016/j.sbspro.2013.01.101
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, United Kingdom: Cambridge University Press.
- Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, 11(3), 283-283. doi:10.1075/ijcl.11.3.04nes
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71.
- Pang, P. (2009). A study on the use of four-word lexical bundles in argumentative essays by Chinese English-majors: A comparative study based on WECCL and LOCNESS. *CELEA Journal*, 32(3), 25-45.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). London, United Kingdom: Longman.
- Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84-94. doi:10.1016/j.jeap.2014.01.002

- Qin, J. (2014). Use of formulaic bundles by nonnative English graduate writers and published authors in applied linguistics. *System*, 42(1), 220-231.
doi:10.1016/j.system.2013.12.003
- Safarzadeh, M. M., Monfared, A., & Sarfeju, M. (2013). Native and nonnative use of lexical bundles in discussion section of political science articles. *Iranian Journal of Applied Language Studies*, 5(2), 137-166.
- Salazar, D. (2014). *Lexical bundles in native and nonnative scientific writing*. Amsterdam, Netherlands: John Benjamins Publishing Company.
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1-22). Philadelphia, PA: John Benjamins Publishing Company.
- Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education* (Vol. 22). John Benjamins Publishing.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, United Kingdom: Oxford University Press.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Uçar, S. (2017). A corpus-based study on the use of three-word lexical bundles in the academic writing by native English and Turkish nonnative writers. *English Language Teaching*, 10(12), 28.

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

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

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

Xu, F. (2012). The use and developmental features of lexical bundles in Chinese learners' English academic writing. *Foreign Languages and Their Teaching*, 4, 51-56. doi:10.13564/j.cnki.issn.1672-9382.2012.04.013

APPENDICES

Appendix A: Structural Classification of Lexical Bundles

Noun Structure

<p>Noun phrase + <i>of</i>-phrase fragment</p>	<p>the end of the, the purpose of this, growth rate of the, one of the most, the growth rate of, the value of the, the rest of the, the success of the, a better understanding of, an analysis of the, the creation of a, the economies of scale, the implementation of the, the majority of the, the effect of the, the results of this, , the null hypothesis of, the findings of the, the value of a, one of the biggest, the size of the, frequency and percentage of, the case of the, the first part of, the reliability of the, the quality of the, the second part of, one of the major, the aim of this, the findings of this, the first half of, the first stage of, findings of this study, purpose of this study, reliability of the questionnaire, results of this study, results of this thesis, the needs of the, the creation of the, the success of a, the different types of , the top of the, the use of the, purpose of this paper, an example of this, a part of the , an important part of, university of Nevada las, an example of a, a better understanding of , word of mouth and, one of the main, impacts of mass tourism, significance of the study, part of the island, the result of this, a large number of, a wide range of, the positive effects of, the quality of life, the importance of this, the performance of the, findings of the study, organization of the study, purpose of the study, republic of northern Cyprus, the development of tourism, the number of tourists, image of a destination, positive word of mouth, most of the respondents, Turkish republic of northern, the results of the, the bottom of the, the center of the, the length of the, the total number of, the accuracy of the, the behavior of the, the sensitivity of the, the capacity of the, the design of the, the location of the, the response of the, the effectiveness of the, the iowa department of, the remainder of the, the calculation of the, the objective of this, the use of a, a large amount of, the base of the, the ability of the, the magnitude of the, the number of crashes, the relative importance of, the average number of, the cost of the, the determination of the, the importance of the, Iowa department of transportation, the percentage of the, the compressive strength of, the middle of the, the field of construction, the main objective of, the main objectives of, and the other one, compressive strength of concrete, the complexity of the, the purposes of this, the same number of, the temperature of the, a function of the, a result of the, an example of the, the results from the, each of the three, the efficiency of the, the mechanical properties of, the thermal efficiency of, accuracy of the</p>
---	---

	results, mass flow rate of, the highest amount of, the thickness of the, the ratio of the, results of von mises.
Noun phrase with other post-modifier fragment	an increase in the, the fact that the, the relationship between the, the total variation in, a negative relationship between, a significant role in a positive impact on, a positive relationship between, positive impact on the, total variation in the, variables in the regression, variation in the dependent, an important role in, relationship between customer satisfaction, the extent to which, impact on the environment, and positive effect on, north cyprus as a, the distance to the, recommendations for future research, farm to market road, management in construction industry, the interaction between the, the distance between the, the working fluid in.
Other noun phrase	California state university Sacramento, long and short run, long run and short, run and short run, the most important factor, the short run period, the frequency and percentage, long run economic growth, the second one is, the Las Vegas strip, the literature review will , the united states the, the united states and, the best way to, customer satisfaction and loyalty, significant and positive effect, customer satisfaction and customer, satisfaction and customer loyalty, annual average daily traffic, life cycle cost analysis, the federal highway administration, the longitudinal and transverse, the top and bottom, the federal highway administration, time cost quality and, the mass flow rate.

Prepositional-phrase Fragments

Prepositional phrase + of	at the end of, in the case of, as a result of, in the form of, for the purpose of, in the light of, on the concept of, for the sake of, to the success of, as part of the, at the university of, in the context of, as one of the, in the process of, in the field of, to the lack of, on the importance of, in the number of, of the number of, at the time of, for each of the, through the use of, on top of the, at the top of, as a part of, on the basis of, from the perspective of, in terms of the, in the middle of, as a matter of, in terms of time, as a function of, on the order of, for the purposes of, from each of the, with the addition of, for a variety of, in most of the, with the help of, as the ratio of, in the direction of, on the performance of, on the surface of, as the number of, to the results of.
Other prepositional phrase (fragment)	in the united states, in addition to the, as a result the, at the same time, on the other hand, as a way to, on the other side, of the most important, in order to get, in the long run, in the short run, to the fact that, of the relevant

	<p>variables, for the first time, of the total variation, in a way that, in the dependent variable, in the first part, in addition to this, of goods and services, of the Iranian economy, according to the results, and on the other, as a result this, due to the fact, in the economy and, in the world and, from strongly disagree to, to an increase in, in the hospitality industry, in the las Vegas, of the united states, of Nevada Las Vegas , in the u s, within the united states, in the tourism industry, in the north Cyprus, due to the lack, between customer satisfaction and, of the study this, of the study the, of the study and, in this chapter the, with respect to the, in the following sections, in an effort to, both urban and rural, by the Iowa dot, for both urban and, for this reason the, of the change in, in figure and figure, of this research is, in the construction industry, in this study the, in comparison with the, of probability and impact, of fiber reinforced concrete, in the next chapter, in this case the, in the case study, according to astm c, in order to reduce, of the project and, in order to evaluate, in order to prevent, of time cost and, at this point the, in an attempt to, in contact with the, in order to improve, in order to decrease, in order to increase, in this section the, of maximum von mises.</p>
--	--

Verb Structure

<p>Anticipatory <i>it</i> + verb/adjective phrase</p>	<p>it is important to, it is necessary to, it is not surprising, it is clear that, it is important for, it is essential to, it is important that, it should be noted, it can be seen, it was found that, it can be concluded, it was decided to, it is difficult to, it was determined that, it is obvious that, it should be mentioned, it is possible to, it is seen that, it was observed that, it is clear from, it is observed that.</p>
<p>Passive verb + propositional phrase fragment</p>	<p>can be explained by, is defined as the, are presented in table, are shown in the, can be seen in, used in this study, used in this thesis, shown in the table, are related to the, be seen in table, is also known as, employed in this study, referred to as a, can be considered as, is based on the, based on the results, based on this information, is shown in figure, shown in figure the, be seen in figure, be used as a, are shown in figure, are shown in table, can be found in, referred to as the, was obtained from the, be included in the, accounted for in the, shown in figure and, shown in table the, seen in figure the, are given in table, can be defined as, is illustrated in figure, has been used in, will be discussed in, was used as a, was used as the, is referred to as, be attributed to the, can be seen from, be seen from the,</p>

	are presented in chapter, can be used in, can be expressed as is fixed to the, are explained in the, are used in the, is considered as the, is defined as a, is used in this, used in this project, given in table and, have been used in, defined as the ratio, shown in fig(NUMBER) the, made of titanium alloy, presented in table and.
Copula be + noun /adjective phrase	is one of the, is similar to the, is significant at the, are the most important, is a function of, is a relationship between, is no significant difference, is located in the, is due to the, be due to the, is dependent on the, is the number of, is to investigate the, is because of the, are the most common, is clear from the, is equal to the, is obvious that the.
Pronoun/noun phrase + be	there is a significant, there is a positive, there is a relationship, there is no significant, there has been a, this paper is to, example of this is, there is not any, the tourism industry is, tourism is one of, there is no need, and there is no, this is due to.
Other verb phrase (fragment)	have an influence on, have been carried out, have significant effect on, shows the frequency and, disagree to strongly agree, table shows the frequency, meet the needs of, has a significant and, increase the number of, play an important role, become one of the, provides an example of, et al developed a, has the potential to, that can be used, which is shown in, shows the effect of, mentioned in chapter the, explained in the previous.

Clausal Structure

Verb/adjective + to- clause fragment	are more likely to, will need to be, will be able to, to be able to, can be used to, is going to be, used to test the, is important to mention would be able to, to find out the, is more likely to, to ensure that the is to develop a, is important to understand, used to determine the, used to calculate the, has been shown to, was found to be, were found to be, was determined to be, was used to determine, was assumed to be, was calculated to be, is important to note, to account for the, to be more important, to determine if the, to be used in, can be applied to, investigate the effects of, to investigate the effects.
Verb phrase + that- clause fragment	that there is a, that make up the, that there is no, that there are two show that there are, the results show that, studies have shown that, research has shown that , the study found that, should be noted that, be seen that the, be noted that the, be concluded that the, that can be used, was found that the, can be concluded that, can be seen that, this

	indicates that the, should be mentioned that, the results showed that.
Adverbial fragment	clause as can be seen, as it has been, as shown in figure, as shown in table, as seen in figure, as it can be, as illustrated in figure.

Others

Other expressions	as well as the, when it comes to, of this project is, in order to be, economies of scale are, of this study is, not important at all, that the amount of, this study is to, important to mention that, when it comes to, in order to find, of this paper is, of this study was, as well as other, important to note that, paper is to develop, they do not have, a significant and positive, of North Cyprus is, and positively related to, both positive and negative, significantly and positively related, as well as a, when compared to the, in order to determine, order to determine the, the results obtained from, the other hand the, method is based on, of this chapter is, actions in order to, what are the main, in order to achieve, in order to identify, and as a result, and n is the, figure provides an example, the most commonly used, in table and illustrated, table and illustrated in, of the cycle is, of this thesis is, in figure it is, order to find the, order to improve the, the figure shows the, figure it can be, that by increasing the, the other hand in, and also it is, figure as can be, which is one of, where y is the.
--------------------------	---

Appendix B: Functional Classification of Lexical Bundles

Research-oriented (RO)

Location	at the end of, the end of the, at the same time, the bottom of the, the top of the, the center of the, at the time of, the location of the, on top of the, at the top of, along the length of, the top and bottom, used in this study, in contact with the, at this point the, this paper is to, all over the world, the middle of the, in the middle of, the first part of, in the case study, in the field of, is fixed to the, of the cycle is, in the direction of, for the first time, in the first part, the second part of, the first half of, the first stage of, in the world and, in the context of, at the same time, is located in the.
Procedure	the purpose of this, for the purpose of, an analysis of the, the implementation of the, that make up the, the creation of a, can be used to, the use of the, through the use of, was determined to be, the objective of this, the use of a, that can be used, was obtained from the, was used to determine, the design of the, be used as a, used to calculate the, the determination of the, was used as a, was used as the, for the purposes of, with the addition of, the purposes of this, to be used in, et al developed a, purpose of this paper, purpose of this study, the creation of the, is to develop a, paper is to develop, has been used in, the main objective of, the main objectives of, the aim of this, can be used in, as a function of, have been used in, that by increasing the, are explained in the, are used in the, is used in this, have been carried out, used to test the, variation in the dependent, in the process of, in the form of, purpose of the study, the development of tourism.
Quantification	growth rate of the, one of the most, the growth rate of, the value of the, the majority of the, the rest of the, is one of the, the total number of, in the number of, of the number of, the calculation of the, a large amount of, the magnitude of the, the number of crashes, the average number of, the cost of the, the capacity of the, the remainder of the, as a part of, the distance to the, is the number of, a large number of, a wide range of, the same number of, the distance between the, the temperature of the, each of the three, for a variety of, as part of the, a part of the, an important part of, the different types of, one of the main, as one of the, what are the main, in most of the, as the ratio of, accuracy of the results, as the number of, are the most common, the most commonly used, the highest amount of, the thickness of the, the ratio of the, which is one of, one of the biggest, one of the major, an increase in the, the second one is, the frequency and percentage, reliability of the questionnaire, to an increase in, the reliability of the, that the amount of, become one of the, the number of tourists, increase the number of, most of the respondents, tourism is one of.
Description	to be able to, in the form of, the success of the, a better understanding of, the length of the, the accuracy of the, the behavior of the, the sensitivity of the, the quality of the, the base of the, the performance of the, the ability of the, the relative

	importance of, to the lack of, of the change in, the importance of the, the size of the, a function of the, the complexity of the, has the potential to, an example of the, the value of the, the needs of the, to the success of, an example of this, the success of a, an example of a, a better understanding of , example of this is, word of mouth and, the percentage of the, on the performance of, on the surface of, the efficiency of the, the value of a, in the light of, is a function of, on the concept of, in the economy and, the total variation in, total variation in the, variables in the regression, the quality of life, positive word of mouth, significance of the study, the importance of this, on the importance of, organization of the study.
Topic	in the united states, California state university Sacramento, economies of scale are, the economies of scale, farm to market road, life cycle cost analysis, the Iowa department of, Iowa department of transportation, annual average daily traffic, both urban and rural, by the Iowa dot, for both urban and, the federal highway administration, the longitudinal and transverse, in the hospitality industry, in the Las Vegas, of the united states, the united states and, of Nevada las Vegas , university of Nevada Las, at the university of, the united states the, the Las Vegas strip, within the united states, in the us, in the construction industry, management in construction industry, the compressive strength of, of probability and impact, of fiber reinforced concrete, compressive strength of concrete, the field of construction, of time cost and, time cost quality and, the mechanical properties of, the thermal efficiency of, mass flow rate of, the mass flow rate, made of titanium alloy, of maximum von mises, the working fluid in, the null hypothesis of, of the Iranian economy, frequency and percentage of, of the total variation, long run economic growth, of goods and services, in the economy and, customer satisfaction and loyalty, north Cyprus as a, image of a destination, customer satisfaction and customer, satisfaction and customer loyalty, in the tourism industry, of north Cyprus is, Turkish republic of northern, in the north Cyprus, republic of northern Cyprus, the tourism industry is, hotels in north Cyprus, satisfaction customer loyalty and, part of the island, impacts of mass tourism.

Text-oriented (TO)

Transition signals	in addition to the, on the other hand, on the other side, as well as the, is similar to the, as well as a, be included in the, when compared to the, the interaction between the, as well as other, in comparison with the, the other hand the, and the other one, the other hand in, and also it is, and there is no, is equal to the, a positive relationship between, and on the other, in addition to this, a negative relationship between, disagree to strongly agree, from strongly disagree to, is a relationship between, is no significant difference, the relationship between the, there is a relationship, are related to the, between
---------------------------	---

	customer satisfaction and, relationship between customer satisfaction, the extent to which.
Resultative signals	as a result the, as a result of, in order to be, the results of the, was found to be, it was found that, were found to be, was found that the, the response of the, is due to the, the results of this, be due to the, due to the lack, can be found in, this indicates that the, has been shown to, in order to determine, used to determine the, to determine if the, recommendations for future research, in an effort to, order to determine the, accounted for in the, to account for the, the effectiveness of the, a result of the, it was observed that, and as a result, the results from the, the results showed that , in an attempt to, to an increase in, be attributed to the, the study found that, studies have shown that, research has shown that, in order to find, in order to achieve, in order to identify, in order to reduce, in order to evaluate, in order to prevent, to find out the, the results obtained from, actions in order to, to the results of, is to investigate the, investigate the effects of, to investigate the effects, in order to improve, the effect of the, is because of the, in order to decrease, in order to increase, order to find the, order to improve the, results of von mises, results of this study, this is due to, due to the fact, it is observed that, shows the effect of, as a result this, results of this thesis, according to the results, the findings of the, a significant role in, a positive impact on, have an influence on, findings of this study, in order to get, the findings of this, the results show that, is going to be, have significant effect on, positive impact on the, the result of this, findings of the study, the positive effects of, significant and positive effect, and positive effect on, impact on the environment, and positively related to, play an important role, significantly and positively related.
Structuring signals	of this project is, in figure and figure, as shown in figure, is shown in figure, shown in figure the, be seen in figure, be seen that the, in the following sections, are shown in figure, are shown in table, as shown in the, referred to as the, shown in figure and, shown in table the, as seen in figure, as shown in table, seen in figure the, research has shown that, of this research is, is shown in fig, as shown in fig, shown in figure a, from each of the, is referred to as, figure provides an example, provides an example of, and n is the, of this paper is, of this study was, of this study is, to ensure that the, meet the needs of, referred to as a, the literature review will, they do not have, in this study the, in the next chapter, of this chapter is, in the following section, will be discussed in, are given in table, is illustrated in figure, used in this study, in this chapter the, of the project and, as illustrated in figure, are presented in chapter, the figure shows the, are presented in table, shown in fig ...(NUMBER) the, presented in table and, which is shown in, used in this project, be seen from the, of this thesis is, in this section the, given in table and, explained in the previous, defined as the ratio, is defined as a, in table and illustrated, table and illustrated in, figure it can be, in figure it is, mentioned in chapter the, figure as can be, is considered as the, where y is the, be seen in table, are shown in the, used in this thesis,

	shown in the table, table shows the frequency, this study is to, show that there are, employed in this study, is also known as, shows the frequency and, of the study this, of the study the, of the study and.
Framing signals	in the case of, when it comes to, as a way to, that there is a, with respect to the, is based on the, to the fact that, on the basis of, from the perspective of, for each of the, for this reason the, in terms of the, as a function of, is dependent on the, on the order of, there has been a, in terms of time, according to the results, the fact that the, according to astm c, in this case the, method is based on, as a matter of, with the help of, the case of the, that there is no, that there are two, in the dependent variable, of the relevant variables, in the long run, in the short run, long and short run, long run and short, run and short run, in a way that, the short run period, for the sake of, based on the results, based on this information, there is not any.

Participant-oriented (PO)

Stance features	are more likely to, it is important to, it is necessary to, it is not surprising, will be able to, will need to be, to be more important, it is possible to, of the most important, is more likely to, the best way to, due to the fact, an important role in, can be applied to, is obvious that the, there is no need, are the most important, the most important factor, is significant at the, there is a significant, not important at all, there is a positive, can be explained by, there is no significant, would be able to, significant and positive, has a significant and, can be considered as, both positive and negative, do not have any.
Engagement features	this project is to, it can be seen, can be seen in, can be seen that, it should be noted, should be noted that, be noted that the, is important to note, be concluded that the, it can be concluded, it was decided to, important to note that, it is difficult to, to ensure that the, can be concluded that, it was determined that, was assumed to be, was calculated to be, is defined as the, it is clear that, it is seen that, it is important for, is important to understand, it is essential to, it is important that, as it can be, it is obvious that, it should be mentioned, should be mentioned that, can be defined as, it is necessary to, it is clear from, it is important to, as can be seen, can be seen from, can be expressed as, is clear from the, is important to mention, important to mention that, as it has been.