ANADOLU ÜNİVERSİTESİ

LEXICAL BUNLDE USE OF TURKISH AND NATIVE ENGLISH WRITERS: A CORPUS-BASED STUDY

Yusuf ÖZTÜRK

(Yüksek Lisans Tezi)

Ocak 2014

LEXICAL BUNDLE USE OF TURKISH AND NATIVE ENGLISH WRITERS: A CORPUS-BASED STUDY

Yusuf ÖZTÜRK

MA THESIS

Department of Foreign Language Education

MA in English Language Teaching Program

Supervisor: Prof. Dr. Gül DURMUŞOĞLU KÖSE

Eskişehir

Anadolu University Graduate School of Educational Sciences

January, 2014

"This MA thesis has been funded by Scientific Research Unit of Anadolu University. Project No: 1305E093"

JÜRİ VE ENSTİTÜ ONAYI

Yusuf ÖZTÜRK'ün "Lexical Bundle Use Of Turkish and Native English Writers: A Corpus-Based Study" başlıklı tezi 23.01.2014 tarihinde, aşağıda belirtilen jüri üyeleri tarafından Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca Yabancı Diller Eğitimi Anabilim Dalı İngilizce Öğretmenliği programı yüksek lisans tezi olarak değerlendirilerek kabul edilmiştir.

	Adı-Soyadı	**İmza**
Üye (Tez Danışmanı)	: Prof.Dr.Gül DURMUŞOĞLU KÖSE	
Üye	: Prof.Dr. Ümit Deniz TURAN	
Üye	: Prof.Dr. Işıl AÇIKALIN	
Üye	: Doç.Dr.İlknur SAVAŞKAN	
Üye	: Yard.Doç.Dr.Gonca SUBAŞI	

Doç.Dr.Handan DEVECİ
Anadolu Üniversitesi
Eğitim Bilimleri Enstitüsü Müdür Vekili

ABSTRACT

LEXICAL BUNDLE USE OF TURKISH AND NATIVE ENGLISH WRITERS:

A CORPUS-BASED STUDY

Yusuf ÖZTÜRK

Anadolu University Graduate School of Educational Sciences

Department of Foreign Language Education – MA in English Language Teaching

January, 2014

Supervisor: Prof. Dr. Gül DURMUŞOĞLU KÖSE

In recent decades, English has been the lingua franca of the academia and students and scholars are expected to produce written works in this global language. However, this creates a pressure on non-native students and scholars since they need to have a native-like proficiency to be able to carry out their studies and publish their work, and they need to be familiar with the distinguishing features of academic discourse such as words or multi-word combinations. In this sense, many studies have been conducted to identify such type of combinations, one of which is the recurring multi-word expressions, mostly referred to as lexical bundles. Being composed of three or more words, lexical bundles such as *on the other hand* and *as a result of* are extremely common and important in shaping academic discourse. Moreover, lexical bundles vary

across different disciplines meaning that successful use of lexical bundles typical of a specific academic discipline is important for writers and the absence of such bundles may not sound fluent and native-like.

Recent studies (e.g. Adel & Erman, 2012; Chen & Baker, 2010) have revealed that non-native writers produce not only fewer types of lexical bundles, but also less varied ones. Furthermore, they also overuse a restricted number of bundles in their writing. However, they should have a successful and native-like control over the lexical bundles which are commonly used in their discipline. In this regard, there is a need for studies identifying the frequent lexical bundles in a particular discipline and examining non-native writers' texts to see to what extent they approximate native writers in terms of lexical bundle use.

The primary aim of this study was to investigate Turkish and native English postgraduate students' and native scholars' use of lexical bundles in a specific academic discipline, that is foreign language teaching, in terms of frequency, functions and structures of bundles. For this aim, a corpus of 150 texts was collected containing Turkish and native English students' MA and PhD theses along with native scholars' published research articles. Four-word combinations occurring 25 times per million words and appearing in at least 5 texts or more were identified as lexical bundles. WordSmith Tools 6 was used to retrieve the lexical bundles in the research corpus. The results revealed that Turkish postgraduate students used far more lexical bundles in their texts compared to both native students and scholars. However, there was a redundancy in Turkish students' texts when the token frequencies were examined, meaning that Turkish students overused most of the lexical bundles. On the other hand, statistical analysis of the bundle lists revealed that Turkish postgraduate students employed different bundles from their native peers and scholars. Finally, the structural and functional categories did not show any statistically significant differences.

Keywords: Lexical bundles, English academic discourse, texts created by Turkish postgraduate students, texts created by native English postgraduate student and native scholars

ÖZET

ANADİLİ TÜRKÇE VE İNGİLİZCE OLAN YAZARLARIN İNGİLİZCE AKADEMİK METİNLERDE SÖZCÜK ÖBEĞİ KULLANIMI:

BİR DERLEM ÇALIŞMASI

Yusuf ÖZTÜRK

Anadolu Üniversitesi Eğitim Bilimleri Enstitüsü

Yabancı Diller Eğitimi Anabilim Dalı - İngilizce Öğretmenliği Programı

Ocak, 2014

Danışman: Prof. Dr. Gül DURMUŞOĞLU KÖSE

Son yıllarda İngilizce akademik dünyada küresel bir dil haline gelmiştir. Öğrenciler ve akademisyenlerden de bu küresel dilde yazılı ürünler ortaya koymaları beklenmektedir. Ancak, bu durum anadili İngilizce olamayan öğrenci ve akademisyenler için baskı oluşturmaktadır. Bunun nedeni ise bu bireylerin çalışmalarını tamamlayabilmesi ve yayınlayabilmeleri için anadili İngilizce olanlara benzer bir dil yetisine sahip olmaları ve sözcükler ya da sözcük grupları gibi akademik söylemin ayırt edici bazı özelliklerine hakim olmaları gerekmektedir. Bu anlamda, birçok çalışma, çoğunlukla 'sözcük öbekleri' olarak adlandırılan, bu tip sözcük kombinasyonlarınıı incelemiştir. *Diğer yandan (on the other hand)* ya da *sonuç olarak (as a result of)* gibi üç ya da daha fazla sözcükten oluşan İngilizcedeki sözcük öbekleri son derece yaygın ve akademik

söylemin şekillenmesinde önemlidir. Ayrıca, akademik metinlerde kullanılan sözcük öbekleri disiplinler arasında farklılıklar gösterdiğinden belirli bir akademik disiplinde yaygın olan sözcük öbeklerinin doğru ve etkili kullanımı yazarlar için önem taşır ve bu öbeklerin yokluğu metnin akıcı ve doğal görünmemesine neden olur.

Güncel çalışmalar (örn. Adel & Erman, 2012; Chen & Baker, 2010) İngilizceyi yabancı dil olarak konuşan yazarların daha az sayıda farklı tür sözcük öbekleri kullandıklarını göstermektedir. Ayrıca, bu yazarların metinlerinde, anadili İngilizce olanların tersine, çok sınırlı sayıda sözcük öbeğinin aşırı kullanıldığı da belirtilmektedir. Ancak, İngilizceyi yabancı dil olarak konuşan akademik yazarlardan, disiplinlerinde yaygın olan sözcük öbeklerini etkili ve anadili İngilizce olanlara benzer biçimde kullanmaları beklenmektedir. Bu bağlamda, belli bir disiplinde sık kullanılan sözcük öbeklerini belirleyen ve İngilizceyi yabancı dil olarak konuşan yazarların metinlerini anadili İngilizce olan yazarların metinleri ile sözcük öbeği kullanımı açısından karşılaştırıp ne ölçüde benzerlikler ya da farklılıklar olduğunu inceleyen çalışmalara ihtiyaç vardır.

Bu çalışmanın temel amacı, belirli bir akademik disiplinde anadili Türkçe ve İngilizce olan lisansüstü öğrenciler ile anadili İngilizce olan akademisyenlerin İngilizce akademik metinlerindeki sözcük öbeği kullanımını incelemektir. Bu amaç doğrultusunda, yabancı dil öğretimi alanındaki anadili Türkçe ve İngilizce olan lisansüstü öğrencilerce yazılmış yüksek lisans ve doktora tezleriyle yine anadili İngilizce olan akademisyenlerin araştırma makalelerini içeren ve 150 metinden oluşan bir derlem oluşturulmuştur. Araştırma derleminde her bir milyon kelimede 25 kez tekrarlanan ve en az 5 farklı metinde kullanılan dört sözcüklü ifadeler bu çalışmada sözcük öbeklerinin belirlenmesinde kullanılmış ana ölçütlerdir. Bu ölçütler doğrultusunda araştırma derleminde sözcük öbeklerini tespit edebilmek için WordSmith Tools 6 yazılımı kullanılmıştır. Araştırma sonuçları, Türk lisansüstü öğrencilerin anadili İngilizce olan öğrenciler ve akademisyenlere göre çok daha fazla tip sözcük öbeği kullandıklarını göstermiştir. Ancak, öbeklerin sıklığı düşünüldüğünde Türk öğrencilerin metinlerin çok daha aşırı bir tekrar olduğu görülmüştür. Diğer yandan, öbek listelerinin istatistiksel analizi Türk öğrencilerin ve anadili İngilizce olan öğrenciler ile akademisyenlerin farklı sözcük öbekleri kullandıklarını göstermiştir. Son olarak,

öbeklerin yapısal ve işlevsel olarak analizinde istatistiksel olarak anlamlı bir fark çıkmamıştır.

Anahtar Kelimeler: Sözcük öbekleri, İngilizce akademik söylem, anadili Türkçe olan lisansüstü öğrencilerin metinleri, anadili İngilizce olan lisansüstü öğrencilerin ve akademisyenlerin metinleri

ACKNOWLEDGEMENTS

Writing this MA thesis would not have been possible without the guidance, help and support of many people around me.

Above all, I would like to express my sincere gratitude to my thesis supervisor, Prof. Dr. Gül DURMUŞOĞLU KÖSE, for her conscientious feedback, support and encouragement since the first moment I started this MA program. Without her caring, patience and being a research model, this thesis would not have been completed.

I would like to thank my thesis committee, Prof. Dr. Işıl AÇIKALIN, Prof. Dr. Ümit Deniz TURAN, Assoc. Prof. Dr. İlknur SAVAŞKAN and Assist. Prof. Dr. Gonca SUBAŞI for their feedback and contributions to this thesis.

I also want to thank all my teachers, colleagues and the head of the ELT department at Anadolu University, Prof. Dr. Zülal BALPINAR. It was a great chance for me to study and conduct research at this privileged department.

I would like to thank my colleague, Assist. Prof. Dr. İlknur YÜKSEL, for her guidance and support during my studies.

I would also like to thank Tuncay KARALIK and Sibel Söğüt their feedback, proofreading and valuable opinions on my work.

Perhaps the biggest thanks are for my family and particularly for my father, Ercüment ÖZTÜRK, who guided me in every phase of my life and especially pursuing an academic career. Thank you all for being with me no matter what.

I cannot forget my aunt, Prof. Dr. Füsun ÖZTÜRK KUTER, who also supported and encouraged me in my postgraduate studies. I am grateful to her for all the things she has done and being a model for me as an academic.

Finally, and most importantly, this thesis would have never been completed without the support, understanding and love of my wife, Esra ÖZTÜRK, and my son, Ali Bera ÖZTÜRK. You encouraged me to finish this thesis despite long hours away from you.

Yusuf ÖZTÜRK

Eskişehir, 2014

Curriculum Vitae

Yusuf ÖZTÜRK
English Language Teaching
Master of Arts (MA)

Education

B.A.	2010	Uludag University, ELT Department, Bursa.
H.S.	2006	Yunus Emre High School, Eskişehir.

Work Experience

2011-2014	Research Assistant, Anadolu University, Graduate School of Educational Sciences, Eskişehir.
2010-2011	Research Assistant, Muş Alparslan University, Education Faculty, Muş.
2006-2010	IT Assistant, KUTER Publishing and Presentation Services, Bursa.

Personal Information

Place and Date of Birth: Osmangazi, September 17, 1989.
Languages: English, French.
E-mail: yusufozturk@anadolu.edu.tr, yusufoz48@gmail.com

To my wife and son

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BAWE-EN: British Academic Written English Corpus-English Students

BAWE-CH: British Academic Written English Corpus-Chinese Students

BNC: British National Corpus

COCA: Contemporary Corpus of American English

EAP: English for Academic Purposes

EFL: English as a Foreign Language

ELT: English Language Teaching

ESL: English as a Second Language

FLOB-J: Freiburg-Lancaster-Oslo/Bergen Corpus

L1: First language

L2: Second language

LBs: Lexical bundles

LOCNESS: Louvain Corpus of Native English Essays

LSWE: Longman Spoken and Written English Corpus

NMA: Native English Postgraduate Students' MA Theses

NPhD: Native English Postgraduate Students' PhD Theses

NPMPT: Native English Postgraduate Students' MA/PhD Theses Corpus

NRA: Native English Scholars' Research Articles

NSRA: Native English Scholars' Research Articles Corpus

PM: per million words

T2K-SWAL: TOEFL 2000 Spoken and Written Academic Language Corpus

TMA: Turkish Postgraduate Students' MA Theses

TPhD: Turkish Postgraduate Students' PhD Theses

TPMPT: Turkish Postgraduate Students' MA/PhD Theses Corpus

WECCL: Written English Corpus of Chinese Learners

CHAPTER 1. INTRODUCTION

In recent years, the number of corpus-based studies have boosted with the advent of technology. Through the use of corpora, it is now easy to examine naturally occurring lengthy texts, either spoken or written, and reveal patterns in language use. Many studies have particularly focused on academic writing in this respect and revealed that 'language in use is characterized by repetition of fixed and semi-fixed multi-word combinations and by use of formulaic patterns' (Byrd & Coxhead, 2010, p. 32). If this is the case, then it would be of significance to study these multi-word expressions mostly referred to as lexical bundles in academic prose.

In this sense, this corpus-based study aimed to examine Turkish and native English writers' use of lexical bundles in terms of frequency, structure, and function. The research corpus was composed of Turkish and native English postgraduate students' MA and PhD theses along with native scholars' published research articles. The texts were chosen with a discipline-specific approach, and only texts that relates to foreign language teaching research were collected. Thus, the study compiled a list of frequent lexical bundles in this particular discipline as well as comparing the performance of Turkish and native English writers in terms of lexical bundle use.

This chapter provides a background for conducting such a study followed by the aims and research questions. Then, it explains what problems it addresses and what significance it would have for the literature. Finally, potential limitations of the study are discussed.

1.1. Background to the Study

For some time now, English has been the lingua franca of academia and is a global means of communication in the dissemination of knowledge and science (Björkman, 2013). However, this situation may create a pressure and disadvantage for students and scholars worldwide whose first language (L1) is not English. The reason could be that they need to have a native-like proficiency to be able to fulfill their studies and publish their work, and moreover they need to be familiar with 'the distinguishing features of

academic discourse such as vocabulary, norms, set of conventions, and modes of inquiry' (Zamel, 1998, p. 187). Therefore, there is an increasing number of studies conducted to examine and identify multi-word patterns argued to be an important component of academic writing in both non-native writing (e.g. Wei & Lei, 2011; Hyland, 2008a) and general academic writing (e.g. Liu, 2012; Byrd & Coxhead, 2010). One type of these expressions appears to be the conventionalized multi-word combinations mostly referred to as 'lexical bundles' (Biber, Johansson, Leech, Conrad & Finegan, 1999).

Firstly used by Biber et al. (1999), the term 'lexical bundles' can be briefly described as expressions of three or more words that show a statistical tendency to co-occur in a particular corpus and are identified based on a standardized frequency and distribution criteria. To give an example, common lexical bundles in conversation include *I don't know what* or *I said to him*, and in academic prose *as a result of* or *on the other hand*. What is remarkable about lexical bundles is that they are extremely common and constitute an important part of discourse.

Biber et. al. (1999) found that 21% of all the words in their academic prose corpus occurred in a recurrent lexical bundle. Beside their recurrent nature, lexical bundles also have particular characteristics distinguishing them from other types of multi-word expressions like collocations and idioms. Biber and Barbieri (2007) emphasizes that 'most lexical bundles are not idiomatic in meaning and not perceptually salient' (p. 269). In this sense, one can easily understand the meaning of a lexical bundle only by looking at its individual items unlike idiomatic expressions such as *kick the bucket* where more than the literal meaning of the items is needed. Moreover, lexical bundles are not usually complete structural units as in the examples of *in the case of* and *the base of the* (ibid). Rather, they are mostly part of longer structures. Finally, lexical bundles, as seen in the examples, include both function and content words.

Having mentioned what and how common they are, Coxhead and Byrd (2007) argues that these sets of words or bundles are important for writers and teachers for at least three reasons:

> (1) [such bundles] are often repeated and become a part of the structural material used by advanced writers, making the students' task easier because they work with ready-made sets of words rather than having to create each sentence word by word; (2) as a result of their frequent use, such sets become defining markers of fluent writing and are important for the development of writing that fits the expectations of readers in academia; (3) these [bundles] often lie at the boundary between grammar and vocabulary; they are the lexicogrammatical underpinnings of a language so often revealed in corpus studies but much harder to see through analysis of individual texts or from a linguistic point of view that does not study language-in-use. (pp. 134-135)

What makes these bundles perhaps more important for people writing for academic purposes is that lexical bundles vary across different disciplines (Hyland, 2012), which means that successful use of lexical bundles typical of a specific academic discipline is important for writers and the absence of such bundles may reveal 'the lack of fluency of a novice' (p. 165). There is no doubt that another dimension of difficulty is also added for the writers who are the non-native speakers of the language they are writing in (Wei & Lei, 2011; Adel & Erman, 2012) since the mature use of these expressions is 'a marker of proficient language use of a particular register, including academic writing' (Cortes, 2004, p. 398).

1.2. Statement of the Problem

With the developments in corpus linguistics, multi-word combinations were reported in a number of studies, although in different types or terms, adding weight to the importance of multi-word units in language (Chen & Baker, 2010). As mentioned above, a considerable number of studies focused on 'lexical bundles' in different genres, registers, and by different groups (native/non-native) or levels of writers (low/high proficient). In these studies, particularly for non-native writers, such multi-word expressions like lexical bundles are argued to be an important component of fluent linguistic production and a crucial part of native-like proficiency (Cowie, 1998; Hyland, 2012; Simpson-Vlach & Ellis, 2010). Furthermore, the lack of such bundles may result in the fluency of a novice since such units vary across different academic disciplines

(Hyland, 2008a). As a matter of fact, based on the previous research, it was hypothesized that non-native writers would produce fewer bundles overall (Erman, 2009; Howarth, 1998) and less varied ones (Granger, 1998; Lewis, 2009) than native writers.

In this regard, since it is important for non-native writers to sound native-like and academic when structuring their texts, there is a need for further studies investigating their use of lexical bundles and examining to what extent their level of use approximate native writers. On the one hand, although there have been studies conducted such as Hyland (2008a) focusing on the lexical bundle use of non-native postgraduate and expert writers, and Chen and Baker (2010) investigating native and non-native student writing along with native expert writing, the literature on Turkish writers' use of lexical bundles in English academic texts is extremely limited. Only a few studies (Bal, 2010; Karabacak & Qin, 2012) included corpora containing texts produced by Turkish writers. On the other hand, it is obvious that taking a discipline-specific approach in examining lexical bundles would yield more useful results since writers from different disciplines rely on different bundles to structure their texts (Hyland, 2008b). Moreover, Cortes (2002) argues that students' written production should be examined at different levels and in different disciplines.

Apart from some studies (e.g. Adel & Erman, 2012) using research corpora composed of texts within a single discipline, most of the research studied texts from a wide range of disciplines, e.g. from medicine to applied linguistics. Furthermore, the native expert writing (i.e. published research articles) was not included in the discipline-specific studies although it can also provide useful data when combined with student writing (Chen, 2009). If such a study takes student writing as a baseline for both native and non-native data, the actual use of expert writers with comparison to student writings would not be possible considering the fact that native peer writing does not always include ideal and standard usage.

1.3. Aim and Research Questions

Considering the aforementioned issues and problems, the present study aimed (a) to identify which lexical bundles are frequently used by Turkish and native English postgraduate students and native scholars in a specific academic discipline, that is foreign language teaching, and (b) to compare their performance in terms of frequency, functions and structures of the bundles. Thus, the following research questions guided the study:

(1) Which lexical bundles are frequently used by Turkish and native English postgraduate students and native scholars?

(2) To what extent do Turkish and native English postgraduate students and native scholars differ in terms of:
   (a) type and token frequency of the lexical bundles,
   (b) their structures,
   (c) and functions?

1.4. Significance

The study of lexical bundles may have great pedagogical value to teachers of English for academic purposes and their students (Hyland 2012). Moreover, identifying the bundles frequently used in a discipline may provide advantages for novice writers in their writing process (Hyland, 2008b). In other words, 'an awareness of lexical bundles can empower L2 writers' in presenting their ideas and giving the right impression (Pang, 2010, pp. 4-5).

As from the methodological perspective, to our knowledge, only two studies (Bal, 2010; Karabacak & Qin, 2012) have ever examined Turkish writers' use of lexical bundles. Bal (2010) investigated the use of lexical bundles in a corpus of published research articles produced by Turkish scholars in six different disciplines while Karabacak and Qin (2012) examined the frequency and types of lexical bundles in the argumentative essays of Turkish, Chinese and American university students with a reference corpus of articles from two American newspapers. Bal (2010) did not have a native reference corpus and focused on a wide range of disciplines. Karabacak and Qin

(2012), on the other hand, only identified the bundles in their corpora without an in-depth analysis into the structures and functions of these bundles although Biber et al (1998) argue that it should, and also studied a relatively small research corpus. Therefore, the present study is a preliminary attempt to make an in-depth investigation of the texts by Turkish writers (i.e. MA/PhD theses) in a specific discipline in terms of their lexical bundle use in comparison with native writers (i.e. MA/PhD theses) and native established scholars (i.e. published research articles).

1.5. Limitations

The current study has a limitation regarding the generalizability of the results. The study took a discipline-specific approach in the compilation of the research corpus and only texts that fell in the area of foreign language teaching research were included for analysis. Therefore, the results and the implications of this study would be limited to this discipline.

CHAPTER 2. REVIEW OF LITERATURE

2.1. Introduction

This section attempts to describe what a corpus and a corpus-based study is along with their characteristics. It, then, focuses on recurrent multi-word expressions in general. Following is the definition of lexical bundles, which is a key to this study. Lastly, major studies conducted in the literature as well as the ones in the Turkish context will be discussed.

2.2. Definition and Characteristics of Corpus and Corpus-Based Studies

Basically, a corpus refers to any body of texts (McEnery & Wilson, 1996). However, a more recent description would be as machine-readable collection of texts, either spoken or written, which were produced for real communicative purposes. This collection is compiled in a way to be representative and balanced in terms of a specific linguistic variety or register or genre, and used for linguistic analysis (Gries, 2009). Here, 'real communicative purpose' means that texts should be produced in a natural communicative setting rather than being created for the purposes of putting them into a corpus except the learner corpora that would be discussed below. The characteristics described in the above description were also pointed out by McEnery and Wilson (1996, pp. 29-32) according to whom a modern corpus had four major characteristics:

- sampling and representativeness;
- finite size;
- machine-readable form;
- a standard reference.

These characteristics distinguish corpus from only being a collection of texts. This means that a corpus should certainly contain texts, either spoken or written, but it should also represent a variety of a language under examination so that the researcher can present the tendencies of that variety as accurate as possible (McEnery & Wilson, 1996). Another point is the size of the corpus, which is rather a matter of debate although more is always perceived better (Sinclair, 2001). The reason for this is that a

corpus compiled ten years ago and accepted as large may be seen as a small one since today corpora get bigger and bigger with the advent of computer technology.

Actually, advantages of corpus-based approach come from computer technology itself, which allows the researcher to store and analyze a large amount of data while providing reliable and consistent analyses (Biber, Conrad and Reppen, 1998). Further, another term 'machine-readable' means that texts are stored in a plain text format on a computer so that they may be searched and manipulated for linguistic analysis. Additionally, machine-readable texts may include additional information referred to as annotation which will be touched upon further on. And lastly, some corpora, especially large ones like COCA and BNC, may constitute a standard reference for the language variety it represents meaning that it is widely available to other researchers (McEnery & Wilson, 1996).

Having discussed the description of a corpus, different scholars also emphasized the characteristics of corpus-based linguistic analysis. Biber et al. (1998, p. 4) argues that a corpus-based analysis:

- is empirical, analyzing the actual patterns of use in natural texts;
- utilizes a large and principled collection of natural texts, known as a 'corpus', as the basis for analysis;
- makes extensive use of computers for analysis, using both automatic and interactive techniques;
- depends on both quantitative and qualitative analytical techniques.

Here, what Biber and his colleagues (1998) put differently from what was already mentioned above is that analyzing a corpus includes more than merely counting certain linguistic features. The quantitative patterns should also be subject to qualitative, functional interpretations. In other words, the researcher makes use of computers both for deriving numerical/automatic data and for making linguistic judgments which are decisions made by the human analyst.

In the last two decades, compilation of corpora has widened in two senses in general. The first one is much larger, mega-corpora like COCA or BNC. These are constantly expanding corpora, new texts are continually being added and these corpora are used to inform dictionaries or grammar books as well as being used in research. The second one includes much smaller, specialized genre-based corpora, many of which contain texts of

written or sometimes spoken academic discourse whose findings are used to inform pedagogy in EAP (Flowerdew, 2002). A similar distinction is also made by Gries (2009) as general corpora intending to be representative and balanced for a language as a whole, and specific corpora that are restricted to a particular variety, register, genre. Although large corpora can make an invaluable contribution to ELT lexicography and language description, unlike small corpus resources, they appear to have less relevance to EAP writing instruction and other areas of ELT (Tribble, 2002). At this point, perhaps it would be useful to mention another distinction commonly referred in the area of corpus linguistics as raw corpora and annotated corpora. The difference between these two types of corpus is that annotated corpora contain additional information such as origin or genre of the corpus data, and various linguistic information like parts of speech or syntactic patterns (Gries, 2009). Other distinctions also exist regarding different characteristics including diachronic and synchronic corpora, monolingual-parallel corpora, and static-dynamic/monitor corpora. Diachronic focuses on how a language/variety changes over time while synchronic corpora is rather a snapshot at a particular point of time. As can be guessed from the names, monolingual corpora include texts from one particular language/variety whereas parallel corpora contain the same texts in several different languages. On the other hand, static corpora have a fixed size, however, dynamic corpora may be constantly extended with new texts (ibid).

Another important issue is the question of what to consider when building a corpus. As noted, a corpus should be representative and balanced in terms of a specific linguistic variety or register or genre. However, how to do this is what actually matters at this point. In her recent chapter, Reppen (2010) discusses the key considerations in building a corpus. Regarding the size, she argues that it is difficult to decide on how much is enough for a corpus and not a case of one size fits all. So, she describes two factors to resolve this question; representativeness and practicality. Representativeness addresses the question of whether the researcher has collected enough texts to be able to represent the type of language under investigation. However, practicality refers to the time constraints that should be considered while collecting texts. Reppen (2010) also asserts that it is sometimes possible to completely represent that variety or genre as in the case of examining the books of a particular author although it is not the case in most of the time. Therefore, in the cases where it is not possible to have a complete

representation, ‘corpus size is determined by capturing enough of the language for accurate representation’ (ibid, p.55). Although there is common belief that creating a written corpus would be much easier than that of a spoken corpus, there are still difficulties for the researcher. For example, Nelson (2010) notes some of these as choosing the texts, accessing the texts, making them machine-readable, storage and analysis.

2.3. Corpus Software Tools

Having mentioned the definition, characteristics, and different types, it should also be noted that a corpus cannot go beyond a collection of text unless it is not processed by a text retrieval tool, or a corpus software tool, which makes it possible to conduct ‘observations of various kinds’ (Hunston, 2002, p.3). One of these tools is WordSmith Tools (Scott, 2011) that is also used in the current study. It is one of the advanced and widely used software tools in corpus-based studies. Apart from its numerous features, it has three main functions, which are namely ‘Concord’, ‘KeyWords’, and ‘WordList’. Figure 1 shows the home screen of the software.

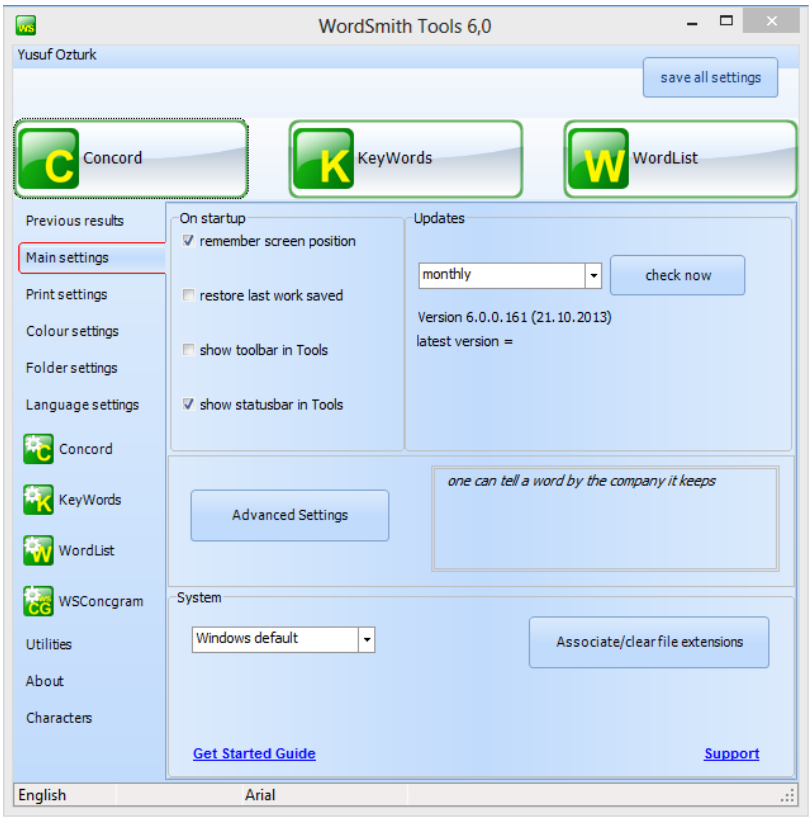


Figure 1. WordSmith Tools 6 Home Screen

As for the Concord function, it is the one most researchers refer to as concordance. Here, the user simply inputs a word or a word combination after selecting the text files and the tool brings the concordance lines containing the query along with statistics. The search screen of Concord can be seen in Figure 2.

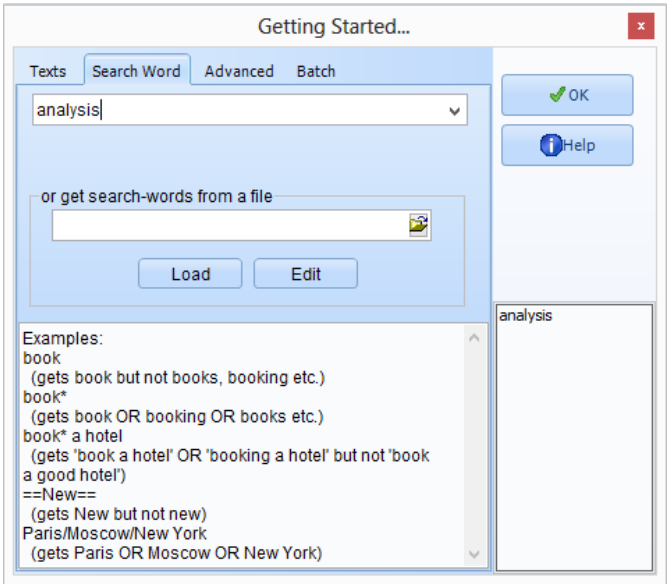


Figure 2. Searching a query in Concord

If the corpus has been annotated beforehand (e.g. parts-of-speech tagged), the search query does not have to include specific words; it may also contain word classes. For example, a researcher can easily search for adjective-noun collocations in an annotated corpus. As seen, the screenshot above represents a sample word, 'analysis', in a sub-corpus used in this thesis. Then, Concord brings up the concordance window as in Figure 3.

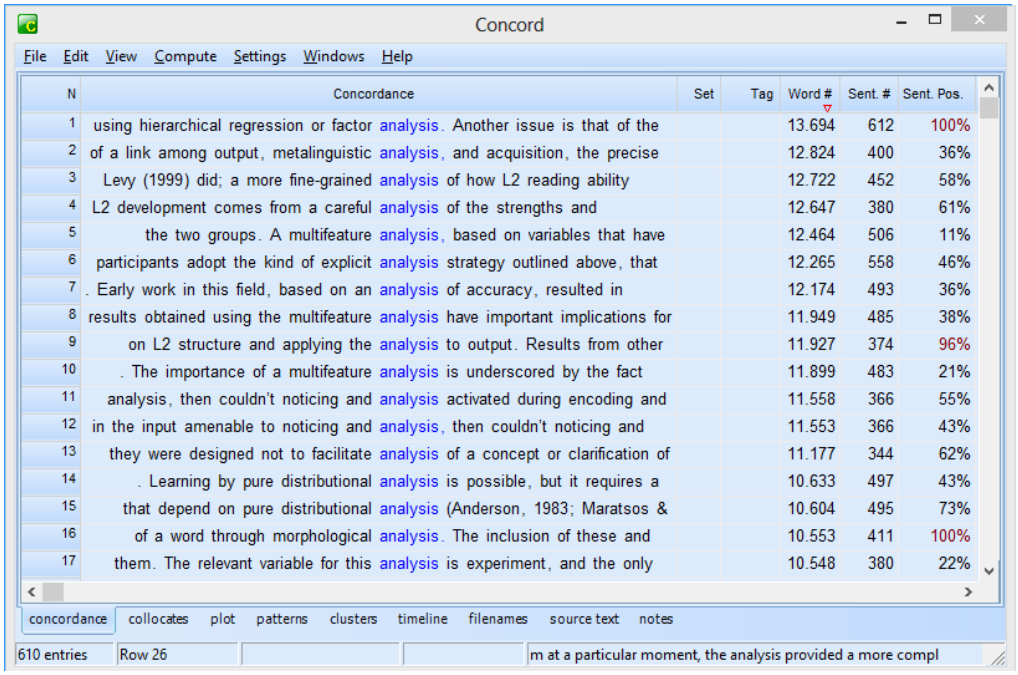


Figure 3. Concordance lines for the word 'analysis'

The lines can be sorted in various ways like alphabetically or by frequency. Along with the statistical information that Concord offers, what is as important is the context that the concordance lines provide. It is particularly key for qualitative analysis such as functional meanings, which would be discussed further on. Additionally, using the Concord window, a researcher can list which words that the query word collocates with or what clusters or bundles contain it. Speaking of bundles, another function of WordSmith is called WordList (See Figure 4), which helps the researcher retrieve continuous word strings, or bundles, with a specified length and frequency criteria.

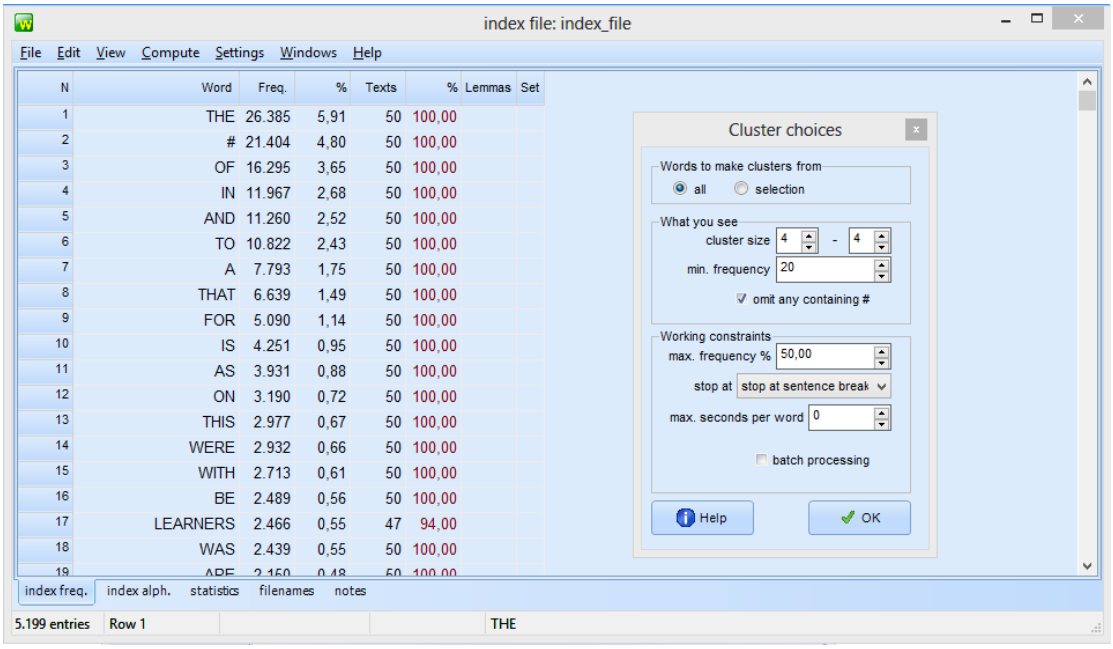
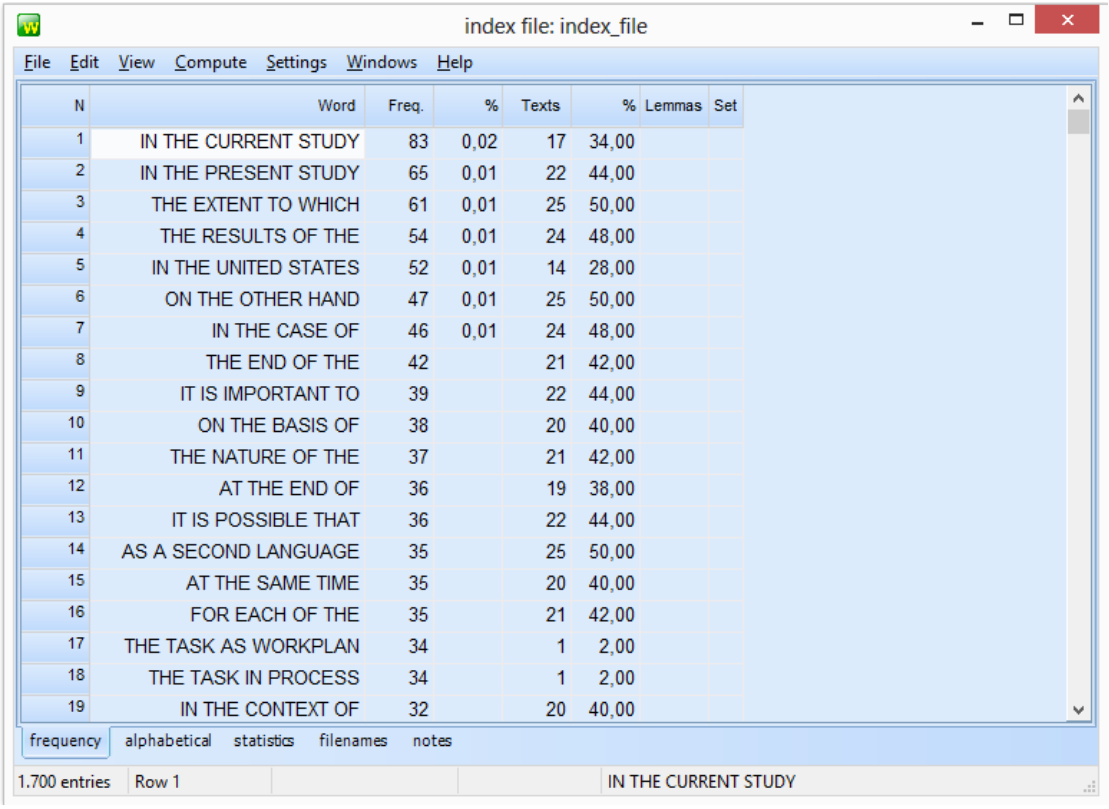


Figure 4. WordList function of WordSmith

When the texts are selected, WordList simply lists all the words by frequency. However, after moving on to computing clusters, a smaller window pops up with the length and frequency criteria according to which the word strings would be identified. Then, the outcome would be a list of the word combinations retrieved out of the texts selected. Figure 5 shows a sample analysis for four word combinations in the same sub-corpus.



N	Word	Freq.	%	Texts	%	Lemmas	Set
1	IN THE CURRENT STUDY	83	0,02	17	34,00		
2	IN THE PRESENT STUDY	65	0,01	22	44,00		
3	THE EXTENT TO WHICH	61	0,01	25	50,00		
4	THE RESULTS OF THE	54	0,01	24	48,00		
5	IN THE UNITED STATES	52	0,01	14	28,00		
6	ON THE OTHER HAND	47	0,01	25	50,00		
7	IN THE CASE OF	46	0,01	24	48,00		
8	THE END OF THE	42		21	42,00		
9	IT IS IMPORTANT TO	39		22	44,00		
10	ON THE BASIS OF	38		20	40,00		
11	THE NATURE OF THE	37		21	42,00		
12	AT THE END OF	36		19	38,00		
13	IT IS POSSIBLE THAT	36		22	44,00		
14	AS A SECOND LANGUAGE	35		25	50,00		
15	AT THE SAME TIME	35		20	40,00		
16	FOR EACH OF THE	35		21	42,00		
17	THE TASK AS WORKPLAN	34		1	2,00		
18	THE TASK IN PROCESS	34		1	2,00		
19	IN THE CONTEXT OF	32		20	40,00		

Figure 5. A sample list of bundles in WordList

Again, in this window, results can easily be sorted in a way that is more suitable for research purposes. Along with frequency information, the researcher can also find out how many texts a particular bundle occur in. This information is especially useful when setting a distribution, or dispersion criteria, which will be discussed in the methodology section. Although other functions of WordSmith will also be used, it is mainly the WordList that will be deployed in this thesis. Before moving on, there is also the KeyWord function of WordSmith which identifies words or word strings that are significantly more or less frequent in a corpus in comparison with a reference corpus. In the literature, this is usually called 'keyness analysis' rather than keyword since word strings are examined rather than single words. Here, WordSmith uses various statistical tests such as log-likelihood and chi-square (Scott, 2011).

2.4. Recurrent Multi-Word Expressions

Although mentioned above that there has been an increasing number of studies focusing on multi-word expressions, the interest in them is not new as it dates back even before

the use of machine-readable texts for linguistics analysis. Researchers had focused on how words co-occurred to form formulaic units in discourse. To give an example, Becker (1975) argued that people know more whole phrases than they know single words and he attempted to generate a taxonomy for lexical phrases consisting of six categories, i.e. poly-words, phrasal constraints, meta-messages, sentence builders, situational utterances, and verbatim texts. He also emphasized the recurring nature of language use arguing that people usually speak bringing together combinations of more than one word that they have heard before, and that 'productive processes have the secondary role of adapting the old phrases to the new situation' (p.1). Another researcher, Bolinger (1976) viewed language similarly by arguing that 'our language does not expect us to build everything with lumber, nails, and blueprint, but provides us with an incredibly large number of prefabs' (p.1). According to him, why certain words are produced together, and others not is because we have not heard it before, we have no memory of it. In other words, what he says is that the language use of people is shaped by the co-occurring words that they have heard before.

Although humans have a capacity of novelty in language production, they don't use this to their full extent; if they did, they wouldn't be accepted as 'exhibiting native-like control of language' (Pawley & Syder, 1983: p. 193). Therefore, novel creations constitute a minority of expressions in language use. Besides, recurrently used expressions are regarded as wholes. Pawley and Syder argue that 'some clauses are entirely familiar, memorized sequences. These are strings which the speaker or hearer is capable of consciously assembling or analyzing, but which on most occasions of use are recalled as wholes or as automatically chained strings' (p. 205). Since 'language in use is characterized by repetition of fixed and semi-fixed multi-word combinations and by use of formulaic patterns' (Byrd & Coxhead, 2010, p. 32), these combinations are important for L2 learners being an important component of fluent linguistic production and a crucial part of native-like proficiency (Cowie, 1998; Hyland, 2012; Simpson-Vlach & Ellis, 2010). Furthermore, Coulmas (1979) suggests that on the one hand, successful use of recurrent expressions ("routine formulae", in his terms) may enable learners to behave and use language appropriately in many situations in spite of their poor command of language; on the other hand, deficient knowledge of these

expressions may be a block even for learners having a better command of the target language.

One type of such strings which was the focus of a number of studies is lexical bundles, which are continuous strings of three or more words identified in a corpus based on a frequency and distribution criteria. Examining the literature for studies focusing on this type of combinations, it can be seen that different terminology is used. These include *clusters* (Hyland, 2008a-b; Schmitt, Grandage & Adolphs, 2004), *recurrent word combinations* (Altenberg, 1998; De Cock, 1998), *phrasicon* (De Cock et al., 1998), *n-grams* (Stubbs, 2007) and *lexical bundles* (e.g. Chen & Baker, 2010; Adel & Erman, 2012; Biber & Barbieri, 2007; Cortes, 2002). However, these terms usually refer to the same notion, i.e. continuous word sequences of different lengths identified taking a frequency-driven approach.

2.5. Definition and Characteristics of Lexical Bundles

Firstly used by Biber et al. (1999), the term 'lexical bundles' can be briefly described as expressions of three or more words that show a statistical tendency to co-occur in a particular corpus and identified based on a standardized frequency and distribution criteria. To give an example, common lexical bundles in conversation include *I don't know what* or *I said to him*, and in academic prose *as a result of* or *on the other hand*. What is remarkable about lexical bundles is that they are extremely common and constitute an important part of discourse. Biber et. al. (1999) found that 21% of all the words in their academic prose corpus occurred in a recurrent lexical bundle. Beside their recurrent nature, lexical bundles also have particular characteristics distinguishing them from other types of multi-word expressions like collocations and idioms. Biber and Barbieri (2007) emphasizes that 'most lexical bundles are not idiomatic in meaning and not perceptually salient' (p. 269). In this sense, the meaning of a lexical bundle only by looking at its individual items can easily be understood unlike idioms where more than the literal meaning of the items is needed. Moreover, lexical bundles are not usually complete structural units as in the examples of *in the case of* and *the base of the* (ibid). Rather, they are mostly part of longer structures. Finally, lexical bundles, as seen in the examples, include both function and content words.

Lexical bundles are extremely common in language use, but what makes these bundles perhaps more important for people writing for academic purposes is that lexical bundles vary across different disciplines (Hyland, 2012) which means that successful use of lexical bundles typical of a specific academic discipline is important for writers and the absence of such bundles may reveal 'the lack of fluency of a novice' (p. 165). There is no doubt that another dimension of difficulty is also added for the writers who are the non-native speakers of the language they are writing in (Adel & Erman, 2012). Different studies (e.g. Adel & Erman, 2012; Chen & Baker, 2010; De Cock, Granger, Leech & McEnergy, 1998) showed that non-native writers produce not only fewer types of lexical bundles, but also less varied ones, compared to native English writers. Similarly, some studies also found that non-native writers overuse a restricted number of bundles (DeCock et al., 1998; Wei & Lei, 2011).

2.6. Studies on Lexical Bundles

This section presents studies focusing on the notion of lexical bundles as well as the ones containing texts created by Turkish EFL learners. Table 1 lists some of these studies along with their foci, research corpus and corpus size. Then, the findings of these studies will be expanded to draw a clear background of the literature on lexical bundles.

Table 1. Major studies on lexical bundles

Authors	Year	Focus	Corpus	Corpus Size
DeCock, Granger, Leech & McEnery	1998	Formulaic competence of advanced adult EFL learners	25 informal interviews with both EFL learners and native speakers	Similar lengths in both corpora (around 62,975 words)
Biber, Johansson, Leech, Conrad & Finegan	1999	LBs in conversation and academic prose	LSWE Corpus	Over 40,000,000
Cortes	2002	LBs in native freshmen compositions	Compositions (311 papers)	360,704
Cortes	2004	LBs use of published authors and university students at three levels in two disciplines	Published writings and student writings	Published writings: 1,992,531; Student writings: 904,376
Biber, Conrad & Cortes	2004	LBs in classroom teaching and textbooks vs. conversation and textbooks	T2K-SWAL Corpus	2,009,400
Scott & Tribble	2006	LBs in student and expert writing in a specific discipline	MA dissertations and BNC World English Edition	POZ_LIT: 352,258 BNC: 1,500,000
Biber & Barbieri	2007	LBs in a wide range of spoken and written university registers	T2K-SWAL and LSWE	T2K-SWAL: 2,541,795 LSWE Academic:

				5,330,000
Hyland	2008b	LBs and disciplinary variation	Research articles, doctoral dissertations and master's theses	3,500,000
Hyland	2008a	LBs and writing expertise	Research articles, doctoral dissertations and master's theses	3,500,000
Ping	2009	LBs in native vs. non-native texts	LOCNESS and WECCL	1,300,000
Byrd & Coxhead	2010	Identifying the most pedagogically useful LBs	414 academic texts in four disciplines	3,600,000
Chen & Baker	2010	LBs in native vs. non-native peer texts along with expert texts	FLOB-J, BAWE-EN and BAWE-CN	470,000
Wei & Lei	2011	LBs and writing expertise	20 doctoral dissertations and 120 published articles	2,250,000
Adel & Erman	2012	LBs in native vs. non-native peer texts	325 student essays	1,110,000

LBs: Lexical bundles

Although it was previously mentioned that the term lexical bundles firstly appeared in Biber et al. (1999) and became the focus of many studies thenceforth, De Cock et al. (1998) conducted a similar study though they used the term formulaic expressions referring to automatically extracted combinations of two, three, four and five words. They simply aimed to reveal the formulaic competence of advanced adult EFL learners of French L1 and their research corpus included a comparable set of native speakers of English. Though they focused on the recurrent word combinations in informal speech, their results were important as being one of the first studies of its kind. They found that advanced EFL learners used multi-word combinations, and in some cases even more combinations than native speakers. However, they also added that the learners' use were 'not necessarily the same as those used by the native speakers' in terms of frequency, syntactic uses and pragmatic functions (p. 78).

Elaborating more on the notion of formulaic expressions or recurrent multi-word combinations, Biber et al. (1999) wrote a chapter in Longman Grammar of Spoken and Written English titled as "*Lexical expressions in speech and writing*" (pp. 987-1036). They termed the word combinations which 'are recurrent expressions, regardless of their idiomaticity, and regardless of their structural status' (p. 990) as "lexical bundles". They also limited their description of lexical bundles in terms of length of the expressions. To put in other words, lexical bundles are recurrent expressions of three or more words which can be either idiomatic or not, or can be a complete structural unit or not. In their analysis, they used the two registers in LSWE (Longman Spoken and Written English) Corpus (i.e. a large corpus containing more than 40,000,000 words of text). These registers were conversation and academic prose, which, they argued, would 'show the most striking differences in language use' (p. 990). Findings revealed that lexical bundles were 'extremely common both in conversation and academic prose' (p. 994). Some of the most frequent lexical bundles in conversation included *I don't know what, I don't want to, I was going to* and *what do you*. As for academic prose, some frequent bundles were *in the case of, on the other hand, in order to* and *one of the*. They also found that an important proportion of discourse in conversation and academic prose was 'made up of recurrent lexical bundles' (p. 995). To put this finding in numbers, 30% of the words in conversation and 21% of the words in academic prose occurred in a recurrent lexical bundle. With regard to their structural status, as also referred in the

definition, only 15% of the lexical bundles in conversation and less than 5% of the lexical bundles in academic prose represented complete structural units. Furthermore, the methodology adopted in Biber et al. (1999) as well as Biber et al. (2004) which uses a frequency and a distribution cut-off point along with a structural and functional taxonomy developed in these studies was employed in a number of studies investigating the use of multi-word expressions, mostly referred to as lexical bundles and/or clusters, including the present study.

Focusing more on university contexts, Biber et al. (2004) investigated the use of lexical bundles in university classroom teaching and textbooks and compared the findings to their previous research in conversation and academic prose (1999). They used texts from the T2K-SWAL Corpus (TOEFL 2000 Spoken and Written Academic Language Corpus) which contained around two million words. The analysis revealed that the bundles in classroom teaching and textbooks differed dramatically from conversation and academic prose, and lectures used twice as many bundles than textbooks and academic prose. Furthermore, VP-based bundles were commonly used in conversation and classroom teaching while rarely employed in textbooks and almost never preferred in academic prose. The researchers argued that these patterns containing VP fragments like *I mean you know* or *you don't have to* are associated with conversation used mostly for stance functions while academic prose uses mostly NP/PP-based bundles for referential functions like *at the end of* or *in terms of the*.

In a further study, Biber and Barbieri (2007) investigated lexical bundles in a wide range of spoken and written university registers, including both instructional registers and student advising/management registers. These registers included, for example, office hours, class management talk, written syllabi, etc. They used the same research corpus with Biber et al. (2004) along with LSWE, and found that although lexical bundles were relatively rare in the academic/instructional written registers compared to spoken university registers, they were more common in the written non-academic registers than in any other university register. So, different from the previous study (Biber et al., 2004) showing that lexical bundles were more common in speech than in writing, they were also common in instructional written discourse.

As listed in Table 1 above, apart from Biber and his colleagues, Cortes (e.g. 2002, 2004) was one of the researchers who conducted many research studies regarding

the use of multi-word expressions. In 2002, she analyzed native freshman compositions in terms of lexical bundle use and examined whether the students' use of lexical bundles would be more similar to those found in conversation rather than academic prose, as argued by composition instructors. In her research corpus, she included 311 student writings piling up a total of 360,704 words and used a specially-designed computer program. The results revealed 93 different types of bundles occurring 20 or more times in a million words and in 5 or more texts. At first glance, the bundles produced by the students were structurally similar to those in academic prose. However, a more detailed analysis of the functions and structures showed that the bundles used were mostly served as temporal or location markers, which are not exclusively used in academic prose, which points to the importance of analyzing lexical bundles in terms of both their structures and functions. For further studies, she suggested that these bundles should be analyzed in detail both structurally and functionally, and students' written production should be examined at different levels and in different disciplines.

In a further study, Cortes (2004) compared the use of lexical bundles by published authors in history and biology and by students at three different levels (i.e. undergraduate lower division, undergraduate upper division, and graduate level) in these disciplines. The bundles employed by the published authors were called the 'target bundles'. The researcher included these particular disciplines in her research corpus of about 2 million words because 'they represent different research and methodological traditions, showing some of the diversity present among university disciplines, and because each of them considers writing to be an important skill in the development of academic competence' (p. 402). The findings revealed that the students rarely used the lexical bundles identified in the corpus of published writing. In a similar study, Scott and Tribble (2006) also looked at student writings and published articles in literary criticism, but this time student writings were MA dissertations by Polish students. As for the findings, student writers used less varied and less sophisticated lexical bundles than expert writers. The researcher argued that the low use of anticipatory-it four-word bundles in student texts might be an indication of less evaluation, or point to the fact that evaluation was done in a different but equally appropriate way.

Concerning the issue of writing expertise and academic disciplines, two studies by Ken Hyland (2008a-b) should also be mentioned here. Actually, these studies made use of the same research corpus, however, set out to investigate different research questions. The 3.5-million-word corpus contained 120 published papers in four disciplines (30 papers in each), and 80 PhD and Master's theses (20 in each disciplines) of students at five Honk Kong universities. Arguing that control of lexical bundles is an important component of fluent linguistic production, one of the studies (2008b) aimed to investigate disciplinary variations in the use of these bundles in student writings and published articles. The discipline was chosen 'to represent a cross-section of academic practice' and included electrical engineering, microbiology, applied linguistics and business studies (p. 8). Student writings were produced by Cantonese L1 speakers at five Hong Kong universities. However, as for the research articles, L1 of the writers was not an issue for the researcher who took them as expert writers. The results revealed 240 different 4-word lexical bundles and the most frequent bundle being *on the other hand* occurring 200 times per million words, almost doubling the next placed bundles *at the same time* and *in the case of*. What is more is that there was a disciplinary variation in the distribution of the bundles. For instance, many bundles in engineering were not found in the other disciplines and there was a greater reliance on these structures than the other fields. Moreover, biology had the smallest range of different bundles and the fewest bundles overall.

In the other study (2008a), Hyland compared the use of lexical bundles in the texts by different levels of writers. He found that the frequency of forms, structures and functions varied considerably across student and expert writing and argued that 'the research articles […] contained far fewer clusters and far fewer different clusters overall; they included largely different clusters to the student genres, with less than half of the forms overlapping in the most common 50 items, and with far more noun phrase + of structures' (p. 59). Furthermore, the Master's theses showed an opposite pattern including a large number of bundles, which is, according to Hyland, is 'no accident' since there is also the issue of different genres. It does not necessarily reflect these ESL writers' deficiencies in English or 'their ability to control the conventions of academic writing in a foreign language' (p. 59).

Another study focusing on different levels of writers is that of Wei and Lei's (2011) that investigated the use of lexical bundles in a corpus of doctoral dissertations by Chinese L1 learners and published journal articles by professional writers. The whole corpus of 20 doctoral dissertation by Chinese L1 learners and 120 published articles by professional writers contained 2,250,000 words. The researchers firstly identified what bundles the Chinese advanced EFL learners used along with their structures and functions in context, and then they examined the nature of these bundles compared to those used by professional writers. Supporting Hyland (2008b), the findings showed that the advanced learner writers used much more bundles and different bundles than the professional writers did. As for the structural differences, the learners used more passive and less anticipatory-it structures.

Apart from the research on the use of lexical bundles in different registers, genres and levels of writers, Byrd & Coxhead's (2010) study set out to identify and examine the bundles in general academic writing. They used the corpus created for the development of Academic Word List (Coxhead, 2000) which included 414 academic texts (i.e. journal articles, book chapters, course workbooks, laboratory manuals, and course notes) in four disciplines (i.e. arts, commerce, law, and science) and contained 3.5 million words. This study set out to create a pedagogically-useful list of lexical bundles used in these four disciplines. They firstly sought for patterns of similarity and difference across the disciplines. The results revealed that 73 bundles were shared across four disciplines, however, these bundles did not occur in equal number in each discipline. So, the researchers reduced this list to the bundles reasonably well-distributed across the disciplines and the new list contained 35 bundles. Then, they compared it with the lists reported in Biber et al. (2004) and Hyland (2008a). The final list consisted of 21 lexical bundles, 'which can be viewed by teachers and materials writers as highly important and fairly stable across a variety of types of academic prose' (p. 39), as argued by the researchers.

Considering the importance of lexical bundles as discussed in the studies mentioned above, a group of studies (Ping, 2009; Chen & Baker, 2010; Adel & Erman, 2012) compared non-native English speakers' performance in using these bundles with that of native speakers. These studies mostly contained essays in their research corpora. Ping (2009) compared LOCNESS (Louvain Corpus of Native English Essays) and

WECCL (Written English Corpus of Chinese Learners). The study identified 361 lexical bundles in WECCL and only 54 in LOCNESS at a frequency cut-off point of 40 times per million words. Although there seems to be a huge difference between native and non-native essays, the researcher argues that a great number of the bundles employed by non-native learners were topic-related content bundles and there were a lot of functional bundles used by native speakers, but not by non-natives.

As for Chen and Baker's (2010) study, they also used two existing corpora; the Freiburg-Lancaster-Oslo/Bergen (FLOB) corpus, and two sub-corpora of the British Academic Written English (BAWE) Corpus. A different aspect of this study is the inclusion of native expert writing. The researchers supported the idea that while comparing native and non-native peer performance, having native expert data would also provide useful data combined with student writing. So, the comparison was conducted between three groups; native expert, native and non-native student writing. While native expert writing and native student writing contained similar amount of bundles (108 and 104 types, respectively), non-native student writing included 80 types of bundles. The findings also revealed that non-native writers had some control of these bundles, but do not 'demonstrate it as diversely and robustly as native writers do' (p. 43).

Similarly, Adel and Erman (2012) compiled a corpus of essays produced by non-native Swedish L1 students and native English (British) students at undergraduate level, though not quite similar amounts in each year. However, this study only focused on texts written in a specific discipline, that is linguistics. What is also different from other studies is that the researchers also examined whether the frequency of the bundles in both corpora differed significantly using log-likelihood statistic. Native students' texts contained a far wider range of bundles than those of non-native students, with a total of 130 as compared to 60. Moreover, frequency of the 70% of the bundles occurring in one corpus (43 types in non-native data and 89 in native data) differed statistically significantly from the other.

So far, this review has touched upon the studies focused on lexical bundles from a wide range of perspectives. Lastly, the studies using a corpus of texts produced by Turkish L1 writers and focusing on their use of lexical bundles will be examined. As

mentioned before, the literature on Turkish writers' use of lexical bundles is extremely limited. The only two studies are represented in Table 2 below.

Table 2. Research on Turkish Writers' Use of Lexical Bundles in English texts

Authors	Year	Focus	Corpus	Corpus Size
Bal	2010	LBs in Turkish scholars' research articles	200 articles in six disciplines	1,005,137
Karabacak & Qin	2012	LBs in novice argumentative essays and expert articles as reference	51 papers in total	43,700

LBs: Lexical bundles

To our knowledge, only two studies in the literature examined the issue in terms of Turkish L2 writers. The first one, Bal's study (2010), investigated the use of lexical bundles in research articles written in English by Turkish scholars, and collected a corpus of 200 articles in six different discipline although the texts in each discipline were different in both number and length. The most frequent lexical bundles found were *on the other hand, the end of the, as well as the, in the case of* and *one of the most*, out of the 99 bundles identified at 20 times per million words. This study merely identified the bundles in the research corpus and categorized them structurally and functionally. It didn't make any comparisons with a reference corpus or similar studies. Another issue was the homogeneity of the corpus. Since these bundles vary based on disciplines, having different number of texts in different lengths would ignore this variation.

On the other hand, Karabacak and Qin investigated the use of lexical bundles in argumentative papers written by three groups of university writers, Turkish, Chinese, and Americans. They also compiled a corpus of newspaper articles to identify target bundles to make a comparison with. The analysis revealed that 96 bundles were used by Turkish and Chinese students but never used by American students. And they concluded that some bundles are not acquired naturally, meaning that simple exposure does not transfer directly into students' production in writing. Therefore, they suggested that explicit teaching might be required to hasten their acquisition process.

Considering the studies discussed in this chapter, lexical bundles are reported to be common in academic prose and that non-native writers seem to differ from native English writers in various respects. Since it is an important component of fluent language use, to what extent non-native writers approximate native writers is a notable issue. Furthermore, the variation across different academic disciplines makes having a successful control over these bundles more significant to the academic writers in a particular discipline. On the other hand, it is obvious that there has not been any systematic and in-depth study on Turkish L1 writers' use of lexical bundles in English. From this perspective, the current thesis will be a preliminary attempt in investigating the use of lexical bundles in a corpus of academic texts by Turkish and native English postgraduate students, and scholars in a particular discipline, i.e. foreign language teaching research.

CHAPTER 3. METHOD

3.1. Introduction

This chapter describes the characteristics of the research corpus and how it was compiled. Then, it explains the procedure followed to identify the lexical bundles found in the research corpus. Lastly, taxonomies and descriptions for the structural and functional analysis of the bundles are provided.

3.2. Research Corpus

As discussed above, corpora are in nature two-fold; mega/big corpora, or small specialized corpora. The corpus that was used in this study is a small and specialized one in line with the aims of the research. Although Sinclair (2004) asserts that 'small is not beautiful' (p. 189) when it comes to building a corpus, small corpora better suits the teaching contexts with specific needs such as ESP or EAP (Flowerdew 2002, 2004; Tribble, 2002). Furthermore, while large corpora provide insights into the patterns in the language as a whole, small and specialized corpora 'give insights into patterns of language use in particular settings' (Koester, 2010, p. 67).

For the purposes of the study, a research corpus with three main sub-corpora was compiled. It included Turkish and native English postgraduate students' MA/PhD theses, and native scholars' published research articles as baseline. These genres were chosen as they 'represent the key research genres of the academy' (Hyland, 2008a, p. 47). Figure 6 represents the research corpus and its sub-corpora.

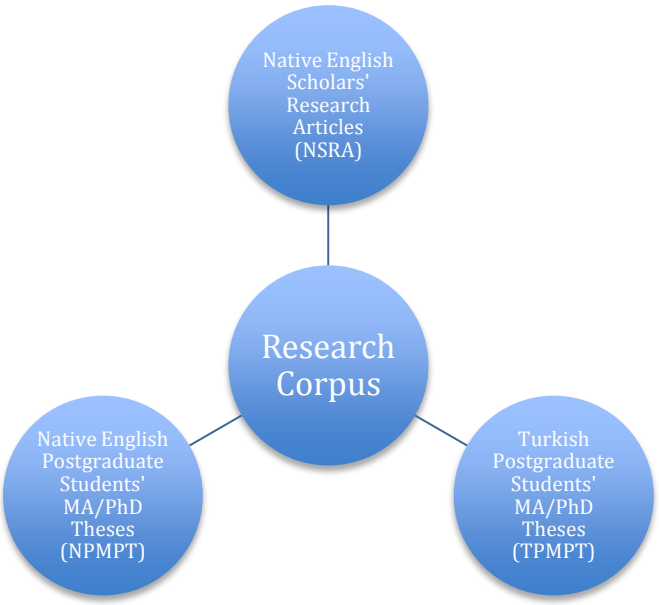


Figure 6. Research corpus with three sub-corpora

The texts were chosen from the area of foreign language teaching within a certain time interval of the last 10 years (i.e. between 2003-2013). Although the method of this study included a standardized cut-off point to be able to compare corpora with different lengths, the number of texts in the research corpus varied since three genres of academic writing (i.e. MA/PhD theses and research articles) usually differ in length. So, the Turkish Postgraduate Students MA/PhD Theses sub-corpus included 20 PhD and 30 MA theses while the Native Writers Corpus also contained 20 PhD and 30 MA theses as well, and the Native English Postgraduate Students MA/PhD Theses sub-corpus was composed of 50 published research articles of native scholars in the area. To sum up, the whole corpus was composed of 150 texts (Appendix I) in total, and yielded around 3 million words. Figure 7 shows the distribution of the texts across research corpora.

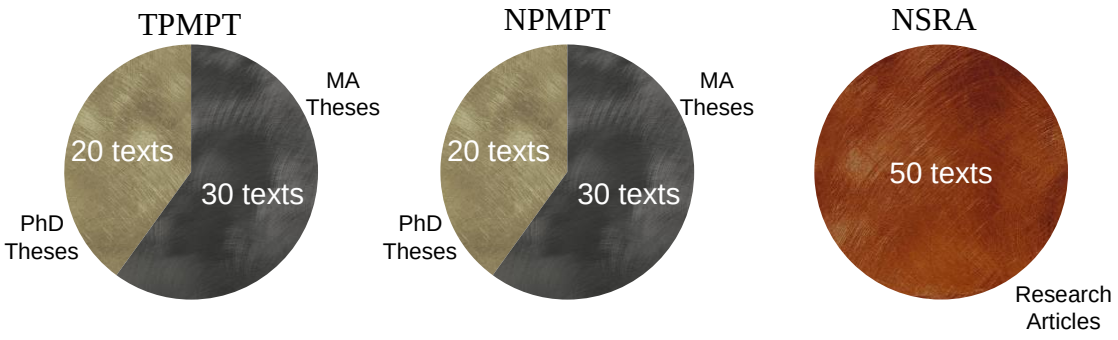


Figure 7. Distribution of texts across research corpora

The theses of the Turkish students were taken from the Higher Education Council theses network. As for the native peer theses, ProQuest's online theses and dissertations library was used. Native scholars' articles, on the other hand, were obtained from the last 10 volumes (i.e. 53-62) of *Language Learning* journal. Selection of the texts for all the three sub-corpora mentioned above was conducted through expert opinion and based on a topic-wise selection. This means that in the research corpus, different sub-corpora would not have very different topics. The texts were chosen to have similar or comparable topics across the research sub-corpora.

As can be seen in Appendix I, each text was coded so that the reader would know which quotation included in the results section was used in which text. The Turkish students' texts were coded and listed as TMA-1 and TPhD-1, native English students' texts as NMA-1 and NPhD-1, and scholars' articles as NRA-1.

At this point, it may be necessary to clarify the terms 'thesis' and 'dissertation'. They are used differently in different countries. As in the case of most UK, Hong Kong, and Australian universities, a thesis is written for a PhD or an M.Phil., while a much shorter dissertation is for a taught Master's degree. However, in many American universities, it is the opposite, theses are written at Master's level and a doctoral dissertation at PhD level (Bunton, 2002). In Turkey, an L1 equal for the word thesis (i.e. *tez*) is used at both MA and PhD level. Thus, in this study, the term 'thesis' was preferred for both MA and PhD level completion work to avoid any confusion.

3.3. Corpus Statistics

As mentioned above, the whole research corpus contained a total of nearly 3 million words. Around 2.5 million words of this constituted the MA and PhD theses. Table 3 provides a detailed picture of the three sub-corpora in terms of the total number of words that they contained after the direct quotations, tables/figures and other similar elements were excluded from all the texts manually. In particular, the quotations were excluded so that the final version of the corpus texts would only contain the writers' own use.

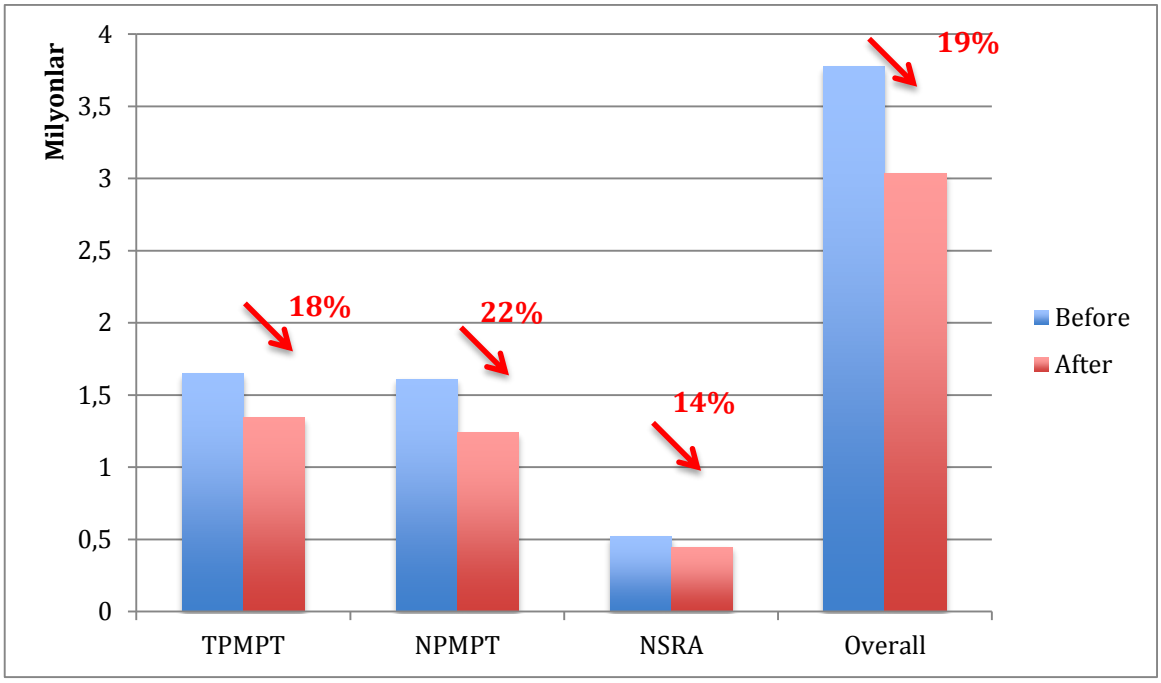
Table 3. Distribution of the total number of words in MA and PhD theses by Turkish and native English students, and research articles by native scholars

		No. of Words	Total
TPMPT	MA	612.379	1.346.396
	PhD	734.017	
NPMPT	MA	457.594	1.239.392
	PhD	781.798	
NSRA		446.009	446.009
		Total	3.031.797

TPMPT: Turkish Postgraduate Students' MA/PhD Theses
NPMPT: Native Postgraduate Students' MA/PhD Theses
NSRA: Native Scholars' Research Articles

As for the overall numbers, there is a slight difference between the theses written by Turkish and native English postgraduates in terms of the total number of words they contained. It was not really possible to make all numbers equal in both sides because the topics and the availability of the theses, especially the ones by native English students, were also considered. However, as it will be discussed while presenting how the lexical bundles were identified in the corpus, a standardized cut-off point frequency was preferred considering the variations in length, or total number of words, across sub-corpora.

The research corpus also included native established scholars' published research articles. These articles were taken as baseline data while comparing Turkish and native English students' MA and PhD theses. 50 research articles obtained from the last 10 volumes of *Langauge Learning* yielded a total of 446.009 words, which is again what is left after exclusion**.** Figure 8 below shows to what extent the total number of words each sub-corpus contained after all the direct quotations and other elements like tables and titles were excluded.



TPMPT: Turkish Postgraduate Students' MA/PhD Theses
NPMPT: Native Postgraduate Students' MA/PhD Theses
NSRA: Native Scholars' Research Articles

Figure 8. Difference in the corpus size after exclusion

After all the texts were revised manually to exclude all the direct quotations and other elements, there was a decrease in the number of words they included as shown in Figure 8. However, the amount of this change seemed to be at similar proportions across the research corpus. Namely, the decrease was 18% in Turkish Postgraduate Students' Theses sub-corpus, 22% in Native English Postgraduate Students' Theses sub-corpus, 14% in Native Scholars' Articles Corpus, and finally 19% in the whole corpus.

3.4. Identifying Lexical Bundles

The present study focused on four-word lexical bundles for two reasons. Firstly, four-word bundles are the most studied length in such studies and considered to be manageable in size for further analysis (Chen & Baker, 2010). Secondly, they are 'over 10 times more frequent than five-word sequences and offer a wider variety of structures and functions to analyze' (Hyland, 2012, p. 151).

Deciding on the length, the next step was setting a frequency cut-off point and a distribution criterion to identify the lexical bundles in the corpus. Cut-off points are not raw frequencies of word strings, but rather standardized ones described in 'per million words'. In the literature, these cut-off points vary from 10 times (Biber et. al., 1999) to 40 times (Biber & Barbieri, 2007) per million words. The cut-off points used in similar studies are presented in Table 4.

Table 4. Frequency cut-off points used in the literature

Study	Corpus	Corpus Size	Frequency Cut-off (per million words)
Cortes (2004)	Published and student writings	2,897,000	20 times
Hyland (2008a)	Research articles, doctoral dissertations and master's theses	3,500,000	20 times
Ping (2009)	Native and non-native peer essays	1,300,000	20 times
Chen & Baker (2010)	Native expert, native and non-native student peer texts	164,742 155,782 146,872	25 times
Adel & Erman (2012)	Native and non-native student essays	247,435 863,207	25 times

As Biber and Barbieri (2007) states, these cut-off points are 'somewhat arbitrary' (p. 267.), usually decided based on the size of the corpus. Considering the estimate size of the corpus which will be used for the purposes of this study, 40 times per million words was set to be the frequency cut-off point. Along with their frequency, the distribution of the lexical bundles throughout the corpus is another criterion for identification in order to avoid individual idiosynracies (Biber, Conrad & Cortes, 2004). This criterion was 5 texts in most studies (e.g. Biber et. al., 1999; Cortes, 2004) while

Hyland (2008a & 2008b) preferred 10% of all the texts in his corpus. Since each sub-corpus of this study includes a total of 16 texts and thus, taking 10% of the texts could be misleading, it would be more suitable to set the distribution criteria at 5 or more texts. To sum up, as represented in Table 5, four-word combinations which occurred 40 times per million words and appeared in at least 5 texts in each sub-corpus were identified as lexical bundles.

Table 5. Frequency cut-off points used in the current study

Corpus	Corpus Size	Cut-off Point (per million words)	Cut-off Point (raw frequency)
TPMPT	1,346,396	25	34
NPMPT	1,239,392	25	31
NSRA	446,009	25	11

TPMPT: Turkish Postgraduate Students' MA/PhD Theses
NPMPT: Native Postgraduate Students' MA/PhD Theses
NSRA: Native Scholars' Research Articles

To retrieve four-word lexical bundles in the corpus based on these criteria, WordSmith Tools 6 (Scott, 2011) was used. Before computing the texts, all the direct quotations were deleted since the writers' own use of lexical bundles was the focus. In addition, all the tables/figures, end/foot notes, and references/appendices were excluded, leaving back only plain text produced by the writers. Right after the initial analysis, two issues needed to be addressed before moving on to further analysis. One is the exclusion of content/context dependent bundles such as *second language acquisition process* or *in the Turkish context* since they need 'to be removed as they are not the "building blocks" which carry a distinct discourse function' (Chen, 2009, p. 58). The other was combining the overlapping bundles such as *it has been suggested* and *has been suggested that*, and combining them into a five-word bundle as in *it has been suggested that*. Such bundles are combined in order to avoid inflated results (Chen & Baker, 2010).

Firstly, lexical bundles in each sub-corpus were identified using the 'clusters' function of WordSmith. As for the first research question, the bundles frequently used in the three sub-corpora were analyzed. Then, for the second research question, lexical

bundles in these sub-corpora were compared in terms of type/token frequency including statistically significant differences using the KeyWords, structures and functions of the bundles. Using the KeyWords function of Wordsmith is useful in especially comparing the relative frequency of combinations across corpora with different sizes and for determining whether the frequency of a combination is statistically higher in one corpus or sub-corpus than another. As shown in Figure 9, the KeyWord function asks for two wordlist files created by the WordList function to compare.

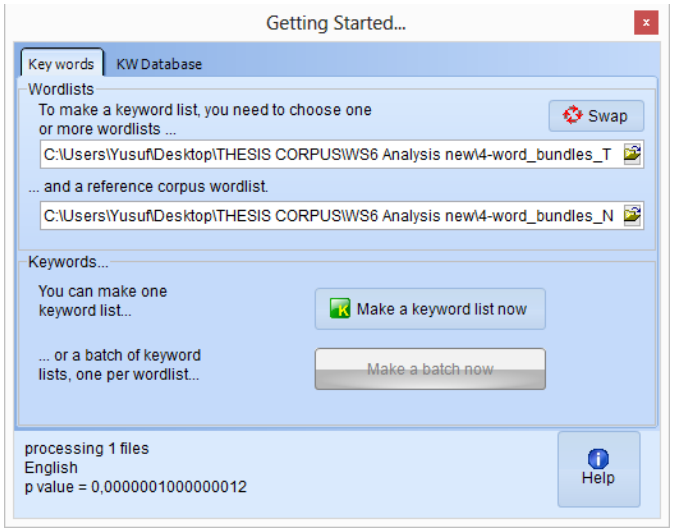
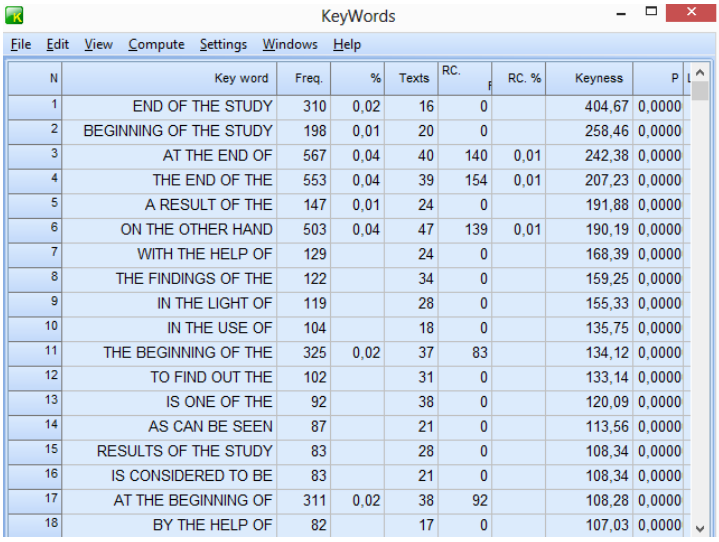


Figure 9. KeyWord function of WordSmith

What this function does here is simply comparing the frequency of bundles in the first wordlist with reference to the reference wordlist by taking the differences in corpus size into account. The outcome is a list of bundles, as seen in Figure 10, which are statistically significantly overused or underused when compared to the reference wordlist.



N	Key word	Freq.	%	Texts	RC.	RC. %	Keyness	P
1	END OF THE STUDY	310	0,02	16	0		404,67	0,0000
2	BEGINNING OF THE STUDY	198	0,01	20	0		258,46	0,0000
3	AT THE END OF	567	0,04	40	140	0,01	242,38	0,0000
4	THE END OF THE	553	0,04	39	154	0,01	207,23	0,0000
5	A RESULT OF THE	147	0,01	24	0		191,88	0,0000
6	ON THE OTHER HAND	503	0,04	47	139	0,01	190,19	0,0000
7	WITH THE HELP OF	129		24	0		168,39	0,0000
8	THE FINDINGS OF THE	122		34	0		159,25	0,0000
9	IN THE LIGHT OF	119		28	0		155,33	0,0000
10	IN THE USE OF	104		18	0		135,75	0,0000
11	THE BEGINNING OF THE	325	0,02	37	83		134,12	0,0000
12	TO FIND OUT THE	102		31	0		133,14	0,0000
13	IS ONE OF THE	92		38	0		120,09	0,0000
14	AS CAN BE SEEN	87		21	0		113,56	0,0000
15	RESULTS OF THE STUDY	83		28	0		108,34	0,0000
16	IS CONSIDERED TO BE	83		21	0		108,34	0,0000
17	AT THE BEGINNING OF	311	0,02	38	92		108,28	0,0000
18	BY THE HELP OF	82		17	0		107,03	0,0000

Figure 10. KeyWord function of WordSmith

A bundle which has a positive keyness score occurs more often (overuse) than would be expected by chance in comparison with the reference corpus. A bundle which has a negative keyness score occurs less often (underuse) than would be expected by chance in comparison with the reference corpus. The higher the keyness score, the more statistically significant the key lexical bundle.

3.5. Structural Categorization

At this step, Biber et. al.'s (1999) taxonomy on categorizing lexical bundles in the Longman Grammar of Spoken and Written English was used as it is the only taxonomy encountered in the literature with slight adaptations. It included twelve structural categories and is described in Table 6.

Table 6. Structural categories of lexical bundles (Biber et. al., 1999, pp. 1014-1024)

Category	Example
Noun phrase with *of*-phrase fragment	*the end of the, the purpose of the*
Noun phrase with other post-modifier fragments	*the extent to which, the relationship between the*
Prepositional phrase with	*as a result of, on the basis of*

embedded *of*-phrase fragment	
Other preposition phrase (fragment)	*in the present study, on the other hand*
Anticipatory *it* + verb phrase/adjective phrase	*it is possible to, it can be seen*
Passive verb + prepositional phrase fragment	*are shown in table, is related to the*
Copula *be* + noun phrase/adjective phrase	*is one of the, was no significant difference*
(Verb phrase +) *that*-clause fragment	*should be noted that, that there is a*
(Verb/Adjective +) *to*-clause fragment	*are likely to be, has been shown to, to be able to*
Adverbial clause fragment	*as shown in figure, if there is a*
Pronoun/noun + *be* + (...)	*this is not the, there was a significant*
Other expressions	*as well as the, than that of the*

After categorizing lexical bundles based on their structures manually, a further statistical analysis was conducted. Chi-square test was done to see whether there are significant differences between Turkish and native English postgraduate students and scholar in terms of structures of the lexical bundles. Differences were further analyzed referring to related studies in the literature.

3.6.Functional Categorization

Final step of the analysis included the functional categorization of the lexical bundles identified in the corpus. With regard to this, the widely used taxonomy initially designed by Cortes (2002), and later improved in Biber et al. (2003, 2004 & 2007) was used in this study.

As described in Figure 11, the taxonomy includes three primary discourse functions which are (1) stance expressions, (2) discourse organizers, and (3) referential expressions (Biber and Barbieri, 2007, pp. 270). Stance bundles such as *are more likely to* and *it is important to* are used to express attitudes or assessments in terms of certainty

or uncertainty that frame some other proposition. Discourse organizers such as *on the other hand* and *in contrast to the* express the connections between prior and coming discourse. On the other hand, referential bundles including *at the beginning of* and *in the current study* make direct reference to physical or abstract entities, or to the textual context itself, either to identify the entity or to single out some particular attribute of the entity as especially important.

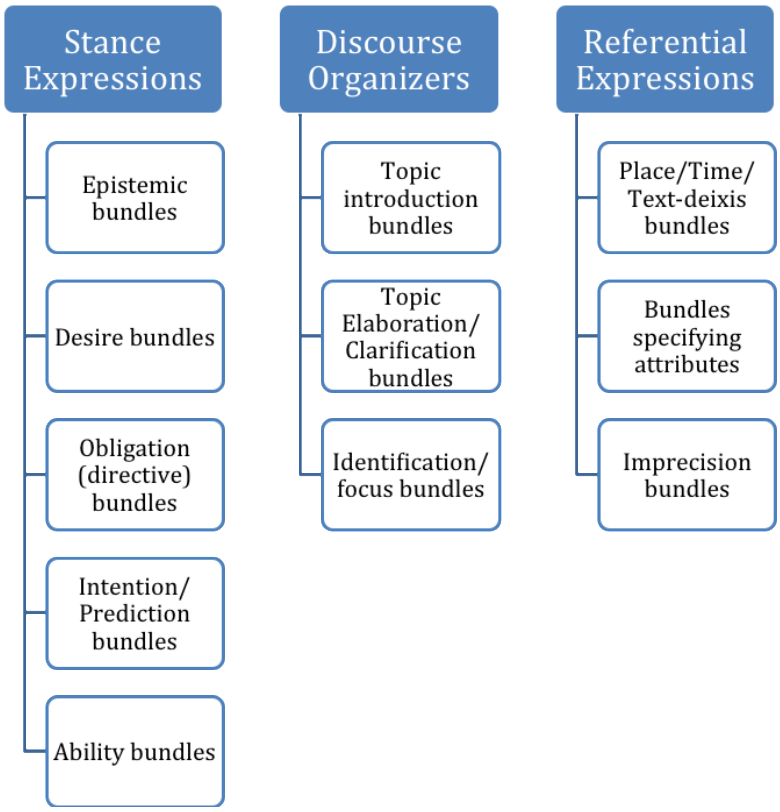


Figure 11. Functional taxonomy of lexical bundles (Biber & Barbieri, 2007, pp. 270-272)

The lexical bundles identified in the three sub-corpora were categorized functionally. To decide the function of bundles, concordance lines were checked and when necessary, the parts of source texts including the bundles were read. On the other hand, the bundles of one sub-corpus (i.e. NSRA) were categorized with a second researcher although using an inter-rater was not a practice preferred in such studies. Since the categorizations mostly overlapped, the rest of the analysis was done only by the researcher of the current study.

Finally, the same as the structural analysis, chi-square test was conducted right after the lexical bundles identified in the corpus were categorized functionally. It was intended to reveal whether any significant differences exist between Turkish and native texts.

CHAPTER 4. RESULTS AND DISCUSSION

4.1. Introduction

This chapter presents the results of the analysis in three main steps. Firstly, overall results with regard to the lexical bundles identified in each sub-corpus are presented while making comparisons with similar studies. Secondly, the difference between the three sub-corpora is discussed in terms of statistical significance of the frequency of the bundles. Finally, comparisons are given to see to what extent Turkish and native English writers differ in structures and functions of the bundles they used.

4.2. Overall Results

As mentioned in the methodology section, following the automatic retrieval of the bundles in each sub-corpus there was a process of manually excluding the context- and content-dependent bundles so that the bundles used for the analysis would be more meaningful. The following presents the change in the number of bundles after context/content-dependent ones were excluded.

Table 7. The number of lexical bundles before and after the manual exclusion

Corpus	Before	After
TPMPT	286	125
NPMPT	124	69
NSRA	133	77
Total	544	271

TPMPT: Turkish Postgraduate Students' MA/PhD Theses
NPMPT: Native Postgraduate Students' MA/PhD Theses
NSRA: Native Scholars' Research Articles

As seen in Table 7, the number of bundles identified in the Turkish writers MA and PhD theses (n=125) almost as twice as that of native writers' theses (n=69) and established scholars' research articles (n=77). In the initial retrieval from WordSmith, the difference is even bigger. Based on mere frequency and distribution criteria before the exclusion, Turkish writers' theses contained nearly two and a half times more

bundles (n=286) than native writers' thesis (n=124) and native established scholars research articles (n=133). In other words, 161 bundles in Turkish writers' theses, 55 bundles in native writers' theses, and 56 bundles in native established scholars' articles were excluded from the final reporting of the findings.

Comparing the three sub-corpora looking at the number of bundles identified, it can be said that native writers' theses and research articles showed similar patterns, at least in the number of bundle types. Conversely, Turkish writers' theses seemed to differ from native writers to a large extent and be quite repetitive in nature. To see whether this difference in the number of bundle types is also evident in different studies, Table 8 presents the findings of similar studies below.

Table 8. The number of bundle types in similar studies across different L1's

Study	Sub-Corpora	No. of Bundle Types
Hyland (2008a)	Research articles	71
	PhD theses (Cantonese L1)	95
	MA theses (Cantonese L1)	149
Ping (2009)	Essays (English L1)	54
	Essays (Chinese L1)	361
Chen & Baker (2010)	Journal/Book Section (English L1)	108
	Essays (English L1)	104
	Essays (Chinese L1)	80
Wei & Lei (2011)	Research articles	87
	PhD theses (Chinese L1)	154
Adel & Erman (2012)	Essays (English L1)	130
	Essays (Swedish L1)	60

The findings of the current study with regard to the number of lexical bundles in the research corpora are consistent with those of Hyland (2008b), Ping (2009), and Wei and Lei (2011). As seen above, not all the sub-corpora had an L1 criterion in the collection of texts. Research articles in Hyland's (2008b) and Wei and Lei's (2011) studies did not consider the nativeness of the writers, but rather took them as expert writers.

On the one hand, Chen and Baker's (2010) and Adel and Erman's (2012) studies revealed different results in spite of the common finding that non-native writers, to a large extent, differed from native English writers in their use of lexical bundles. The low number of bundles in non-native texts in these studies could be due to two reasons. Firstly, it could be due to the fact that they both used the same genre, i.e. argumentative essays. Secondly, there was an imbalance in the number of native and non-native texts in both studies, in the opposite ways though. Adel and Erman's (2012) corpus contained almost three times more non-native texts than native ones while Chen and Baker's (2010) corpus included 50% more native texts than non-native texts. These two setbacks actually exist in Ping (2009) as well although it revealed a different finding, which is higher number of bundles in non-native texts. The imbalance in the number of texts in two sub-corpora in Ping's study was incredibly huge, almost 7 times more texts by Chinese L1 students than native students. A common aspect of these three studies is that they used existing corpora for their research instead of compiling a new corpus for the purposes of their studies, which could explain the incompatible features of the corpora they used.

On the other hand, Wei and Lei (2011) and Hyland (2008b) revealed similar findings with the current study and these studies seem to be more meaningful since they included postgraduate theses. Chinese and Cantonese L1 students' theses employed much more lexical bundles than research articles. What is also common among these studies is the repetitive nature of the non-native texts, which was also revealed in the present study. As a result, since these studies also focused on advanced academic writing (i.e. theses/articles) as the current study, it can be inferred that when it comes to advanced academic writing, non-native writers including Turkish L1 writers tend to

employ considerably higher number of bundle types in a much more repetitive way, differing from native English writers.

Presenting the overall statistics regarding the types of bundles, the actual lexical bundles identified in Turkish and native English students' MA/PhD theses and native scholars' research articles are presented below. The whole lists are provided in Appendix II, but to see the most frequent lexical bundles, Table 9 lists the 50 most frequent bundles based on token frequency, and bundles shared by three groups are shaded in gray while those bundles of Turkish writers' shared by one of the groups emphasized in brown.

Table 9. List of the 50 most frequent lexical bundles identified in the research corpus

NSRA	#	NPMPT	#	TPMPT	#
in the current study	83	(at) + the end of the	154	at the end of + (the)	567
in the present study	65	it is important to	153	on the other hand	503
the extent to which	61	at the same time	151	the results of the	357
the results of the	54	as well as the	150	(at) + the beginning of the	325
on the other hand	47	on the other hand	139	as a result of + (the)	324
in the case of	46	the results of the	125	end of the study	310
(at) + the end of the	42	as a result of	105	beginning of the study	198
it is important to	39	at the beginning of + (the)	92	the analysis of the	182
on the basis of	38	in the present study	82	of the present study	177
the nature of the	37	in the form of	77	in terms of the	166
it is possible that	36	the results of this + (study)	74	in the present study	155
for each of the	35	the use of the	72	with the help of	129
at the same time	35	the total number of	68	at the same time	129
in the context of	32	to be able to	66	the findings of the + (study)	122
the results of this + (study)	30	the purpose of this + (study)	62	in the light of	119
in the form of	28	through the use of	61	to be able to	110
of the current study	28	to the fact that	59	one of the most	109
as well as the	27	in addition to the	58	in the use of	104
it is clear that	25	used in this study	57	to find out the	102
as a function of	25	in terms of the	57	that there is a	101
of the present study	24	in a variety of	54	is one of the	92
the total number of	24	the rest of the	54	as can be seen + (in)	87
with respect to the	24	in the current study	54	as well as the	86
the fact that the	22	in other words the	53	results of the study	83
were more likely to	22	in the case of	53	is considered to be	83
over the course of	21	for the purpose of	50	in addition to the	83
as a result of	20	is important to note + (that)	50	on the use of	82
in addition to the	20	in the following example	49	by the help of	82
with the exception of	20	at the time of	48	in order to find	80
the effect of the	20	the fact that the	48	in order to see	79
to ensure that the	19	a great deal of	48	in the field of	78
are presented in table	19	of the present study	47	the fact that the	77
in a way that	18	in the next section	47	the aim of the	77
the degree to which	18	the majority of the	45	to find out whether	76
in contrast to the	17	the role of the	45	in the form of	74
in the same way	17	in the context of	44	it can be concluded + (that)	73
at the time of	17	on the part of + (the)	44	the results of this	69
used in this study	17	the way in which	44	it was found that	69
a number of studies	17	can be found in	43	in other words the	69
in relation to the	17	in an attempt to	42	that most of the	68
there was also a	17	in a way that	41	the purpose of the	66
at the beginning of + (the)	16	for the purposes of	41	it can be said + (that)	66
that there is a	16	as well as a	40	that there was a	63
it should be noted + (that)	16	one of the most	40	in line with the	63
in terms of the	16	as a result the	40	that the use of + (the)	61
(as) + can be seen in	16	for each of the	39	of the fact that	61
the purpose of this	16	I was able to	39	in addition to this	59
the purpose of the	15	in an effort to	38	according to the results	59
in the field of	15	has been shown to	38	to the fact that	58
to the fact that	15	due to the fact	38	the findings of this	56

NSRA: Native Scholars' Articles NPMPT: Native Postgraduate Students' MA/PhD Theses

TPMPT: Turkish Postgraduate Students' MA/PhD Theses

The most frequently used lexical bundle in Turkish writers' theses was *at the end of + (the)*, which was used 567 times and also the most frequent bundle in native writers' theses with a frequency of 154 times. As for the native established scholars' published research articles, the most frequent bundle was *in the current study* with a frequency of 83, although it was not among the 50 most frequent bundles in the theses. Examining the table above, it can be easily seen that almost half of the 50 most frequent bundles in Turkish writers' theses were also used in native writers' theses and/or published research articles. Furthermore, many of the other bundles are actually variants of the shared bundles. To give an example, *in addition to the* was shared by the three groups of writers, but Turkish writers also used *in addition to this* which was not preferred by native English writers. Similarly, *end of the study* also appeared in Turkish writers theses in addition to *(at) + the end of*, but not in those of native writers.

Despite the huge difference in the number of bundle types discussed at the beginning of this chapter, based on the most frequently used 50 bundles, Turkish writers seem to employ similar lexical bundles with those of their native peers and native scholars. However, there were some bundles employed by Turkish writers, but never or very rarely occurred in native English writers' texts, and vice-versa. For instance, *with the help of* and *by the help of* are among those bundles. *By the help* of never occurred in native writers' theses and research articles while *with the help of* had a frequency of 12 times in total in opposed to 129 times in Turkish writers' theses:

> *"**With the help of** stories, it is much easier to create real life like atmosphere that students will be interested in and in which they will have fun." (*TMA-30)

> *"**By the help of** convenience sampling, we chose 30 of them as our participants."* (TPhD-19)

On the other hand, native English writers preferred *through the use of*, probably to denote a similar notion with *with the help of* and *by the help of*:

> *"Triangulation was achieved **through the use of** the interview process, field notes, and a second interview to discuss the results."* (NPhD-17)

"The present study included quantitative and qualitative data collection ***through the use of*** *a survey research methodology."* (NMA-3)

Another example of this can be *is considered to be* which occurred 83 times in Turkish writers' theses, but only 14 times in native writers' theses and research articles together.

"Therefore, the power of motivation ***is considered to be*** *a specific criterion for effectiveness of language education."* (TMA-11)

Instead of *is considered to be*, native English postgraduate students and scholars preferred different and usually more powerful stance bundles such as *it is important to*, *it is possible that* and *were more likely to* which Turkish students very rarely used:

"Finally, **it is important to** note that, consistent with the quantitative analysis, the differences in use of …" (NPhD-7)

"**It is possible that** these students' increased word consciousness was a reflection of their active learning style." (NMA-30)

"Another related possibility is that learners **were more likely to** notice linguistic items in code-related FFEs because..." (NRA-38)

Although *aim* and *purpose* are synonyms and interchangeably used in academic writing, the native writers seem to have a tendency to use *the purpose of the* or *the purpose of this*, but not *the aim of the* or *the aim of this* as the Turkish writers who used both variants:

*"**The purpose of this** study was to investigate and offer a descriptive evaluation of the interactions that take place between non-native English speaking students in university-level, academic writing classrooms."* (NMA-24)

*"**The purpose of the** second analysis is to provide supporting and illustrative evidence for the quantitative analysis."* (NRA-43)

*"**The aim of the** study is to find whether being an experienced and a novice supervisor has an effect on the supervisory styles…"* (TPhD-1)

To see whether this difference was due to the variations in American and British English, a comparative analysis was conducted in the Contemporary Corpus of American English (COCA). COCA containing around 450 million words has a function that can compare the frequency of words/word combinations in itself with reference to another corpus, e.g. British National Corpus (BNC). Table 10 below presents the raw and normalized frequency information regarding the bundles *the purpose of* and *the aim of*.

Table 10. The frequency of *the purpose of* and *the aim of* in COCA (450 million words) with reference to BNC (100 million words)

Word/Phrase	1: COCA	2: BNC	PM 1	PM 2	Ratio
the aim of	1300	1330	2.89	13.30	0.22
the purpose of	8702	2369	19.34	23.69	0.82

PM: per million words

Apart from the raw frequencies, the comparison tool of COCA also presents normalized frequencies (i.e. per million words) considering the differences in the size of two corpora, which is evident, i.e. COCA is almost five times larger than the BNC. In addition, taking the normalized frequencies, the tool calculates a relative percentage (ratio) and shades the words or word combinations in green if it is much more common than in the second corpus, and in red if it occurs considerably less than in the second corpus.

The bundle *the aim of* occurs 1300 times in COCA and 1330 times in BNC. Although the raw frequencies seem to be closer, there is a difference in normalized frequencies, being 2.89 and 13.30 times per million words, respectively. As it was shaded in red, *the aim of* occurs considerably less in COCA than in BNC. However, as for *the purpose of*, the normalized frequencies and ratios are close to each other,

meaning that this bundle is not overused or underused in either corpus compared to each other. Table 11 shows the frequency results this time for BNC with reference to COCA.

Table 11. The frequency of *the aim of* and *the purpose of* in BNC (100 million words) with reference to COCA (450 million words)

Word/Phrase	1: BNC	2: COCA	PM 1	PM 2	Ratio
the aim of	1330	1300	13.30	2.89	4.60
the purpose of	2369	8702	23.69	19.34	1.23

PM: per million words

As can be seen, *the aim of* is overused in BNC when compared to COCA and there seems to be no significant difference for *the purpose of*. Therefore, it can be argued that Turkish postgraduate students use both American and British English in their academic writing while the native English postgraduate students (i.e. American) seem to stick to American English, as expected. Furthermore, although the native English scholars producing the research articles used in the corpus was not necessarily American or British, they also underused *the aim of* and tended to use *the purpose of*.

Similarly, there are some lexical bundles frequently employed by native writers and native established scholars, but very rarely or never used by Turkish writers. Such bundles will be discussed in the statistical significance section below. However, one inference that can be made from Table 9 above is that although Turkish writers seem to employ similar bundles with native English writers especially when it comes to frequently used bundles, they make use of these bundles redundantly. As an example, *on the other hand* was used 47 times by native established scholars and 109 times by native writers. However, Turkish writers employed *on the other hand* 503 times, almost 10 times more than native established scholars and 5 times more than native writers.

*"**On the other hand**, some of the participants underlined some drawbacks to some extent."* (TPhD-2)

*"**On the other hand**, the differences between treatment groups in the quantitative analysis of the writing task were non-significant."* (NPhD-7)

The case of *at the end of + (the)* is also one of the lexical bundles which showed a difference in frequency across three sub-corpora. It occurred 567 times in Turkish theses, 154 times in native theses, and 42 times in the research articles.

*"**At the end of the** study, the results suggested positive implications of integrating technology in the language classroom for reading instruction and vocabulary development."* (TMA-29)

*"**At the end of the** third week, the students took the second 10-minute impromptu timed writing assessment."* (NMA-10)

4.3. Statistical Significance

To see whether a bundle in a sub-corpus is statistically significantly overused or underused with reference to another sub-corpus, KeyWord function of WordSmith was used as described in the method chapter. Since it can compare only two corpora in terms of the statistical differences in the frequency of words/word combinations, Turkish and native English postgraduate students' bundles were compared with reference to those of native English scholars (See Table 12). In addition, native postgraduate students' and scholars' bundles were then compared with reference to Turkish postgraduate students (See Table 15). Table 12 below may include shared or not shared bundles (as indicated in Table 9) which were statistically significantly overused and underused. Those that were not shared by Turkish and native writers and statistically significantly differed in frequency were shaded in bold.

Table 12. Key lexical bundles in TPMPT and NPMPT with NSRA as the reference corpus

($p < .001$)

Corpus	Level	Key lexical bundles	
TPMPT	Overuse	end of the study (177,43)	**to find out whether (43,50)**
		at the end of + (the) (151,95)	**it can be concluded + (that) (41,78)**
		beginning of the study (113,32)	in other words the (39,49)
		(at) + the beginning of the (108,31)	**it was found that (39,49)**
		on the other hand (97,55)	**that most of the (38,92)**

		as a result of + (the) (88,46) **with the help of (73,83)** **the findings of the + (study) (69,82)** **in the light of (68,11)** to be able to (62,95) one of the most (62,38) **to find out the (58,38)** **is one of the (52,65)** **is considered to be (47,50)** results of the study (47,50) **on the use of (46,93)** **by the help of (46,93)** **in order to find (45,78)** **in order to see (45,21)** the analysis of the (44,80) **the aim of the (44,07)**	**it can be said + (that) (37,77)** that there was a (36,06) **in line with the (36,06)** **of the fact that (34,91)** **that the use of + (the) (34,91)** the results of the (34,77) **according to the results (33,77)** **in addition to this (33,77)** **the findings of this (32,05)** **the number of the (31,48)** in terms of the (31,16) **it is seen that (30,33)** **findings of this study (29,76)** **it was seen that (29,19)** **it can be claimed (29,19)**
	Underuse	---	
NPMPT	Overuse	the use of the (44,26) to be able to (40,58) through the use of (37,50) the rest of the (33,20) in a variety of (33,20)	in other words the (32,58) for the purpose of (30,74) in the following example (30,12) a great deal of (29,51) in the next section (28,89)
	Underuse	in the current study (-70,15)	

When native scholars' articles taken as reference, 41 bundles in Turkish postgraduate students' theses were statistically significantly overused while only 10 bundles were overused and 1 bundle was underused in native postgraduate students' theses. Again, it could be argued that native postgraduates' use of lexical bundles were closer to that of native scholars, compared to Turkish postgraduate students. The repetitive pattern in Turkish students' texts can be observed here as well; the keyness scores (indicated in parentheses) are much higher in Turkish students' bundles.

As emphasized in bold, there are 27 bundles that were not shared by neither native postgraduate students and native scholars, and overused by Turkish students.

Although some of these can be regarded as variants of similar bundles that were already shared such as *in addition to this* (shared bundle: *in addition to the*) and *of the fact that* (shared bundle: *the fact that the*), these bundles seem to be unique to Turkish postgraduate students, and clearly not employed by their native peers and native scholars. Some of these bundles such as *by the help of, with the help of* and *the aim of the* were discussed beforehand. As for the rest, for instance, there are two bundles that are variants of *in order to*:

> *"In addition,* ***in order to find*** *out whether the items were clear and the time was enough, they were applied to five 6th class learners."* (TMA-1)

> *"****In order to see*** *whether these differences were statistically significant, a non-parametric Mann-Whitney U test was performed by means of SPSS 13.00 program."* (TPhD-6)

Other examples of such bundles include anticipatory-*it* structure such as *it can be concluded + (that), it can be said + (that), it is seen that* and *it can be claimed + (that)*:

> *"****It can be concluded that*** *Turkish ELT department students do not use a large variety of connectives."* (TPhD-20)

> *"Accordingly,* ***it can be claimed that*** *there was a slight difference between the experimental group's vocabulary success level (9,08 %) and the control group's success level (8 %)."* (TMA-29)

> *"When the reasons of the results of the study are searched for,* ***it can be said that*** *there are more than one factor affecting how connectives are used by learners."* (TPhD-20)

In addition to the examples given, there was also a simple comparison with similar studies made to see whether these bundles thought to be only preferred by Turkish postgraduate students, not by native students or scholars in the research corpus were used by writers of different L1. Table 13 summarizes this comparison.

Table 13. Comparison of key bundles in Turkish postgraduate theses with writers of different L1

Bundles	Hyland (2008b) (Cantonese)	Chen & Baker (2010) (Chinese)	Wei & Lei (2011) (Chinese	Adel & Erman (2012) (Swedish)	Bal (2010) Turkish (L1)
with the help of				✓	
the findings of the + (study)			✓		
in the light of		✓*			
to find out the	✓				
is one of the	✓	✓	✓	✓	✓
is considered to be		✓			
on the use of					
by the help of					
in order to find				✓	
in order to see				✓	
the aim of the				✓	
to find out whether					
it can be concluded + (that)					
it was found that	✓				✓
that most of the					
it can be said + (that)					
in line with the					✓
of the fact that				✓	✓
that the use of + (the)					
according to the results					✓
in addition to this					
the findings of this					
the number of the					
it is seen that					
findings of this study					
it was seen that					
it can be claimed					

* This bundle was found in the list of native writers.

As can be seen, only 11 of 27 bundles statistically significantly overused in Turkish students' texts and not occurred in native texts were used by writers of different

L1 in four similar studies. The remaining 16 bundles seem to be used only by Turkish postgraduate students based on the literature. The study examining Turkish scholars' published research articles were also added to this comparison, but among those not occurring in the texts of other L1 writers, only two bundles (*in line with the, according to the results*) were found to be shared by the Turkish postgraduate students in the current study. At this point, it can be argued that the bundles that seem to be used only in the texts produced by the Turkish postgraduate students may be a transfer from Turkish. In this regard, Yıldız and Aksan (2013) identified frequently used verbs in a one-million corpus of Turkish academic texts in 15 disciplines. Table 14 presents 10 most frequently used verbs in academic Turkish.

Table 14. 10 most frequently used verbs in academic Turkish (Yıldız & Aksan, 2013)

Verb	Frequency	Translation
görülmektedir	813	***It is seen***
göstermektedir	655	*It shows*
bulunmuştur	541	***It was found***
gerekmektedir	475	*It should…*
bulunmaktadır	431	*It is found*
görülmüştür	403	***It was seen***
belirlenmiştir	369	*It was identified*
saptanmıştır	334	*It was determined*
söylenebilir	296	***It can be said***
gerekir	292	*It should…*

The verbs and their English translations were shaded with bold since they had been identified as being unique to the Turkish postgraduate students in the current study. Based on the table, it may be claimed that Turkish postgraduate students transferred some Turkish expressions to various lexical bundles in English, and consequently differed from native English postgraduates and scholars.

As for the second significance analysis, the key bundles in native postgraduate students' and scholars' texts were determined with reference to Turkish students' texts. In other words, this analysis reveals the bundles statistically significantly overused or underused in native texts when compared to Turkish texts. The findings are summarized in Table 15. The bundles significantly overused or underused in both native

postgraduate students' and scholars' texts with reference to Turkish students' texts were shaded in bold.

Table 15. Key lexical bundles in NPMPT and NSRA with TPMPT as the reference corpus

($p < .001$)

Corpus	Level	Key lexical bundles	
NPMPT	Overuse	**the total number of (100,01)**	for the purposes of (60,30)
		in a variety of (79,42)	as well as a (58,83)
		in the current study (79,42)	as a result the (58,83)
		in the case of (77,95)	**for each of the (57,36)**
		is important to note + (that) (73,54)	I was able to (57,36)
		in the following example (72,07)	has been shown to (55,89)
		it is important to (71,53)	in an effort to (55,89)
		a great deal of (70,60)	**are more likely to (54,42)**
		at the time of (70,60)	as part of the (51,48)
		in the next section (69,13)	**the course of the (50,01)**
		the majority of the (66,18)	**the ways in which (50,01)**
		on the part of + (the) (64,71)	**the context of the (48,53)**
		the purpose of this + (study) (64,71)	in order to determine (48,53)
		the way in which (64,71)	**it is possible that (48,53)**
		can be found in (63,24)	by the end of (45,59)
		in an attempt to (61,77)	as a way to (45,59)
		in a way that (60,30)	
	Underuse	**in terms of the (-47,00)**	**the analysis of the (-101,75)**
		of the present study (-70,02)	**at the beginning of + (the) (-108,28)**
		the results of the (-98,09)	**on the other hand (-190,19)**
		as a result of (-99,87)	**(at) + the end of the (-207,23)**
NSRA	Overuse	**in the current study (230,90)**	**the context of the (41,73)**
		the extent to which (169,70)	these results suggest that (38,95)
		in the case of (127,97)	**(is) + important to note that (38,95)**
		it is possible that (100,15)	should be noted that (38,95)
		for each of the (97,37)	was found to be (36,16)
		of the current study (77,89)	a greater number of (36,16)
		as a function of (69,55)	in the absence of (36,16)

		it is clear that (69,55)	**are more likely to (36,16)**
		with respect to the (66,76)	**the ways in which (36,16)**
		the total number of (66,76)	**the focus of the (36,16)**
		were more likely to (61,20)	to the extent that (36,16)
		over the course of (58,42)	beyond the scope of (33,38)
		with the exception of (55,64)	it is likely that (33,38)
		the effect of the (55,64)	play a role in (33,38)
		to ensure that the (52,86)	that the number of (33,38)
		are presented in table (52,86)	the size of the (30,60)
		the degree to which (50,07)	to be related to (30,60)
		in a way that (50,07)	**in an attempt to (30,60)**
		in contrast to the (47,29)	in any of the (30,60)
		at the time of (47,29)	from the current study (30,60)
		a number of studies (47,29)	in a study of (30,60)
		there was also a (47,29)	it may be that (30,60)
		are summarized in table (41,73)	it is difficult to (30,60)
	Underuse	**in terms of the (-31,16)**	**at the beginning of + (the) (-94,75)**
		the results of the (-34,77)	**on the other hand (-97,55)**
		the analysis of the (-44,80)	**(at) + the end of the (-129,71)**
		as a result of (-88,46)	

Native postgraduate students overused a total of 33 bundles and underused 8 bundles; as for native scholars, they overused 46 bundles, and underused 7 bundles when compared to Turkish students. As mentioned in the overall findings, the number of bundle types in native scholars' articles was 83, and native postgraduate students 75. Therefore, considering the number of key bundles, it can be argued that Turkish postgraduate students considerably differed from native postgraduate students in their use of lexical bundles.

It is clear that there are some lexical bundles unique to Turkish postgraduates and some bundles unique to native postgraduates and scholars. Although Turkish students seem to have shared bundles with their native counterparts and scholars, which may show their high level of English and familiarity with academic writing, even the raw frequencies of these bundles differ to a large extent, which points to the verbose or redundant nature in their writing. Furthermore, the bundles unique to Turkish students

and not even employed by other L1 writers in similar studies such as *it can be said that* or *it was seen that* could be due to their effort to directly translate what they have in mind in Turkish to English rather than trying to be native-like and more academic. On the other hand, the bundles observed to be unique to native texts in the current study such as *in an attempt to, it is clear that* and *was found to be* seem to reveal the bundles distinguishing them from Turkish texts.

Regarding the different bundle types employed by student and published writers, Cortes (2004) argues that although students might have encountered such bundles in their academic reading, simple exposure does not necessarily result in the acquisition of these expressions. Therefore, identifying the key bundles unique to non-native writers and those unique to native writers is important in this respect.

4.4. Structures of Lexical Bundles

The lexical bundles identified in the research corpus were then analyzed based on their structures using the same taxonomy with Biber et al. (1999), as discussed in the methodology chapter. The overall findings are presented in Table 16 in the form of percentage of that structure among all structures and all cases.

Table 16. Structures of the bundles in the three sub-corpora

Structure	Turkish PGSs		Native PGSs		Native Scholars	
	% of all structures	% of all cases	% of all structures	% of all cases	% of all structures	% of all cases
Noun phrase with of-phrase fragment	24	31	28	28	19	20
Noun phrase with other post-modifier fragments	1	1	4	3	5	7
Prepositional phrase with embedded of-phrase fragment	14	20	20	21	20	22
Other preposition phrase (fragment)	17	18	22	24	24	27
Anticipatory it + verb phrase/adjective phrase	7	4	2	4	9	9
Passive verb + prepositional phrase fragment	7	3	4	2	6	4
Copula be + noun phrase/adjective phrase	1	1	2	2	2	1
(Verb phrase +) that-clause fragment	11	7	4	2	3	2
(Verb/Adjective +) to-clause fragment	9	7	7	5	3	2
Adverbial clause fragment	2	1	0	0	1	1
Pronoun/noun + be + (...)	0	0	1	1	2	1
Other expressions	6	3	4	5	3	3
Total	100	100	100	100	100	100

When the table is examined, the distribution of bundle structures seems to be similar with only minor differences. In this sense, the chi-square test did not reveal any statistically significant difference between the three groups of writers and 12 structural

categories, X^2 (22, N = 291) = 23.75, p = .36 (See Appendix III). The chart below also shows the distribution of structures in clustered bars.

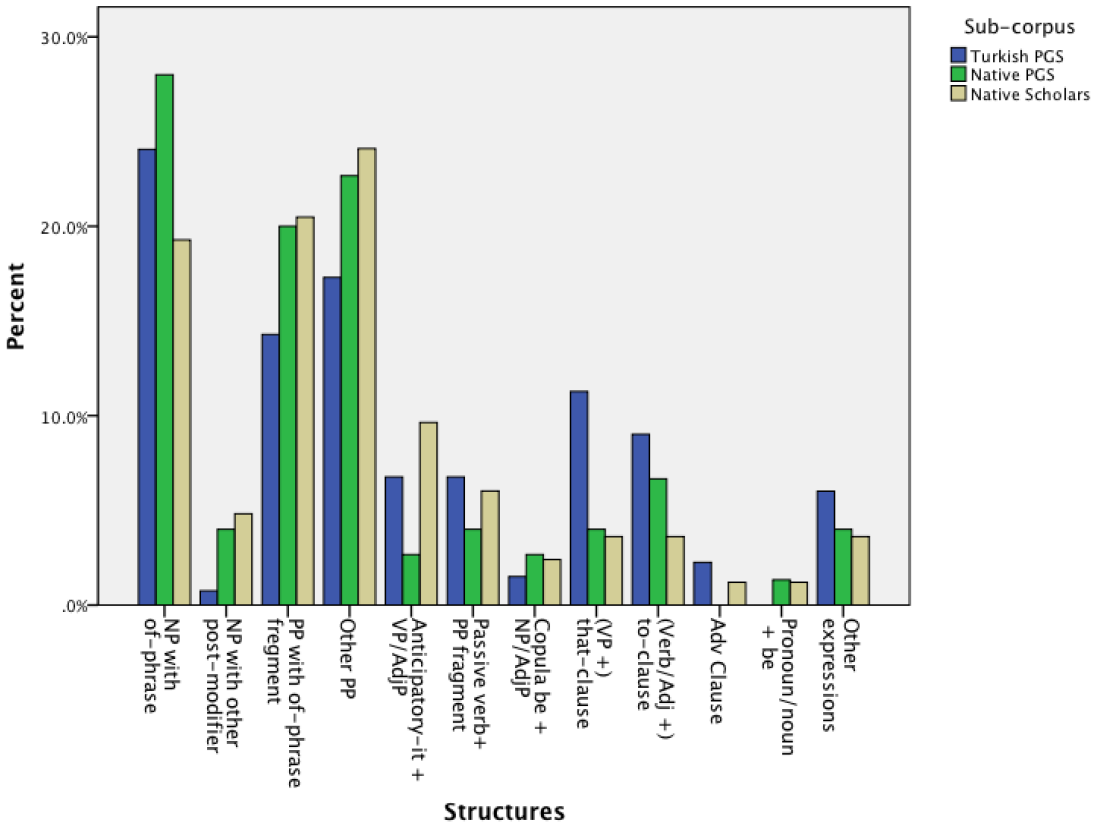


Figure 12. Structural distribution of bundles used by three groups of writers

Considering NP with of-phrase fragment and NP with other post-modifier fragments, native texts included slightly more NP-based bundles (e.g. *the results of the, the extent to which*). Likewise, PP-based bundles (e.g. *at the end of, with respect to the*) also occurred more in native texts than that of Turkish texts. As for VP-based bundles, anticipatory-it + VP/AdjP (e.g. *it is important to, it can be concluded*) and passive verb + PP (e.g. *can be seen in, are summarized in table*) fragment structures were used more in Turkish and native scholar texts than native student texts. Copula be + NP/AdjP bundles (e.g. *is important to note, is one of the*) were distributed almost equally and formed a very small proportion. Two types of structures that Turkish students employed more frequently were (Verb phrase +) that-clause fragment (e.g. *the results showed that, we can say that*) and (Verb/Adjective +) to-clause fragment (*to be able to, are more likely to*).

If both columns are examined in Table 15, some differences can be realized. For example, NP with of-phrase fragments in Turkish texts constituted 20% of all structure, but 28% of all cases (i.e. the percentage in the total token frequency of the bundles in that structural category). Similarly, 14% of all structures and 20% of all cases in Turkish texts was PP with embedded of-phrase fragment. However, as for native texts, this difference between the percentage of structures and all cases does not exist. This difference also support the earlier finding that there seems to be a redundancy in texts of Turkish students, but this time it can also be argued that this redundancy is observed in NP and PP-based bundles.

In spite of the fact that the redundancy or the repetitive nature in Turkish postgraduate students' theses is also observed here, they used similar structures since there was no statistically significant difference in the distribution of structural categories. This finding does not seem to support Chen and Baker (2010) and Hyland (2008b) where the difference between the groups of writers in their studies was larger. For instance, in both studies, research articles included much more NP with of-phrase fragments than non-native student texts. On the other hand, the finding of the current study is consistent with Wei and Lei's (2011) indicating similar distribution of structures in non-native postgraduate texts and professional writing. Perhaps this is due to the fact that both the current study and Wei and Lei's study included texts from disciplines (i.e. foreign language teaching and applied linguistics, respectively) that require a high level of English even at undergraduate level. Therefore, the writers of these texts presumably have advanced English proficiency. In this sense, it may be that writers of advanced English proficiency use similar proportion of structures of lexical bundles with native writers although the types and tokens of bundles they use differ to a large extent.

4.5. Functions of Lexical Bundles

Finally, the bundles were categorized functionally using Biber and Barbieri's (2007) taxonomy which is the most updated taxonomy based on Biber et al. (1999). Figure 13 shows the distribution of functional categories across three groups of writers.

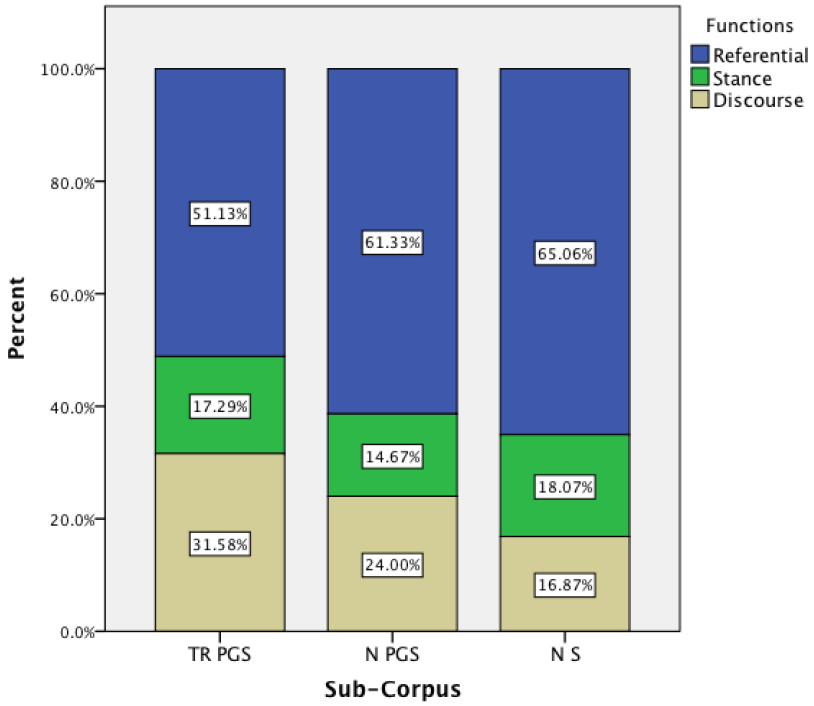


Figure 13. Functional distribution of lexical bundles (types)

As can be seen, native postgraduate students and scholars used more referential bundles to make reference to entities, either physical or abstract, or to the textual context itself (e.g. *in the current study, at the end of, can be seen in, the total number of*).

"*In fact, all of the variables employed **in the current study** were relevant for both data sets.*" (NRA-1)

"***At the end of the** twelve-week period, the participants completed a follow-up questionnaire, a standardized posttest, and a post-treatment writing task.*" (NPhD-7)

"*As **can be seen in** Table 3.1, these three students, Andrew, Phil, and Tina had submitted 10 DAVI uploads to our course website.*" (NMA-5)

Higher use of referential bundles by native writers was also found in Chen and Baker (2010) and although very slightly in Adel and Erman (2012). Leaving out the size of difference, the proportions of discourse and stance bundles are also similar. In these two studies, native texts included more stance bundles to express writer attitude or assessment of certainty (e.g. *it is possible that, it may be that*) which was also the case in the current study.

"***It is possible that*** *the first group of students places a higher priority on developing a native-like accent and thus wants a NS pronunciation model to follow.*" (NPhD-13)

"***It may be that*** *the raters' global impression scores were based, in part, on the communication skills subcomponent of L2 oral ability, whereas fluency, accuracy, and complexity were not.*" (NRA-33)

However, Turkish postgraduate students used more discourse organizers to reflect relationships between prior and coming discourse (e.g. *on the other hand, as a result of, in addition to the*), similar to the findings of the current study.

"***On the other hand****, some practitioners indicate newspaper articles written for native speakers are not always appropriate for ESL students although they agree that newspapers can represent useful tools in the classes.*" (TMA-29)

"***As a result of*** *these studies, it is widely accepted that the AWL basically comprises vocabulary that is common across a range of different academic fields also the applicability of AWL to variety of disciplines has been confirmed to a large extent.*" (TPhD-10)

"***In addition to the*** *studies on teachers in their initial years of service, the literature also involves studies that concentrate on more experienced teachers and the comparison of beginning and experienced teachers in many respects.*" (TPhD-14)

Although Chen and Baker (2010) did find a significant difference, the chi-square test in the current study did not reveal any statistically significant differences between the distribution of functional categories and three groups of writers, X^2 (4, N = 291) = 6.67, p = .15 (See Appendix IV).

The reason that Chen and Baker (2010) found a significant difference and this study did not could be attributed to the previously mentioned characteristic of the research corpus used in this study: the non-native students (i.e. Turkish postgraduates) were actually theses writers with advanced level English, not essay writers at undergraduate level like in their study. Therefore, this could explain why the chi-square test did not find a significant difference.

Perhaps it should be noted that although there was no significant difference in the distribution of functions, the Turkish postgraduates employed different bundles, particularly stance bundles, in the same discourse function. Table 17 below compares the stance bundles of native English and Turkish writers.

Table 17. Comparison of stance bundles in order of frequency

Turkish postgraduate theses	Native postgraduate theses	Native scholars
to be able to	**it is important to**	**it is important to**
is considered to be	**to be able to**	it is possible that
the fact that the	**to the fact that**	it is clear that
it can be concluded + (that)	is important to note + (that)	the fact that the
it can be said + (that)	the fact that the	were more likely to
of the fact that	i was able to	**it should be noted**
to the fact that	due to the fact	**to the fact that**
it is possible to	are more likely to	(is) + important to note that
it can be claimed + (that)	it is possible that	are more likely to
it is necessary to		it is likely that
be taken into consideration		it is difficult to
it is important to		it may be that
due to the fact		
it is believed that		

we can say that are considered to be be claimed that the will be able to may be due to **it should be noted**		

As can be seen, although the proportion of stance bundles in functional categories was similar, Turkish postgraduate students employed different bundle types than those of native students and scholars. This finding is consistent with the significance analysis of key bundles which revealed the bundles unique to each groups of writers. To sum up, although Turkish students seem to use similar proportion of functions of lexical bundles, the bundles types they use differ from native students and scholars to a large extent.

CHAPTER 5. CONCLUSION, IMPLICATIONS AND SUGGESTIONS

5.1. Summary of the Study

This study aimed to examine the use of lexical bundles in a corpus of MA and PhD theses produced by Turkish and native English postgraduate students, and published research articles by native English scholars in the area of foreign language teaching research. As a result of the analysis, a total of 271 4-word combinations occurring at 25 per million words and appearing in at least 5 different texts were identified in the research corpus. The highest number of bundle types was found in Turkish students' texts including 125 bundles while native students' texts contained 69 bundles and scholars' 77 bundles. Although it was hypothesized in the previous literature that non-native writers would produce fewer bundles overall (Erman, 2009; Howarth, 1998) and less varied ones (Granger, 1998; Lewis, 2009) than native writers, the current study revealed a different finding in this respect. The Turkish postgraduate students in the research corpus was observed to employ a much wider range of lexical bundle types than the native students and scholars, which is consistent with Hyland's (2008b) and Wei and Lei's (2011) studies. This consistence is argued to be due to the fact that both studies and the current study contained postgraduate theses and dissertations in the research corpora. On the other hand, studies such as Chen and Baker's (2010) and Adel and Erman's (2012) focusing on university-level argumentative essays supported the aforementioned hypothesis. Therefore it can be concluded that variety in lexical bundle use may be affected by writing expertise since theses writers employed a wider range of bundles while constructing their texts compared to their native peers. Moreover, considering the 50 most frequent lexical bundles, almost half of the bundles in Turkish students' texts were either similar to or variants of those found in native students' and students' text, which can be interpreted as Turkish postgraduate students being familiar to the bundles used by their native peers and scholars to a certain extent.

Although there were similar bundles shared by three groups of writers, Turkish students extremely overused most of these bundles when compared to native students and scholars. This finding with regard to redundancy in non-native texts is also supported by Chen and Baker (2010) and Hyland (2008). It can be inferred that despite being familiar with the frequently used bundles, Turkish postgraduate students use more

varied bundles than native English students and scholars in a way more repetitive nature.

In terms of the significant differences in the frequency of actual bundles types, the current study revealed key findings. Firstly, 42 bundles were found to be statistically significantly overused by Turkish postgraduate students and 27 of these such as *it can be said that* and *it was seen that* were the bundles not shared with native English postgraduate students and scholars and argued to be unique to Turkish students. A comparison of the lexical bundles in similar studies showed that 11 of the 27 bundles overused by Turkish students but rarely or never used by native students and scholars were not employed by non-native writers of different L1, either. This finding could be explained by some expressions in Turkish academic writing being transferred to English by the Turkish postgraduate students. For example, *it can be said + (that)* that was used by the Turkish students seems to be the English equal for one of the 10 most frequent verbs in academic Turkish, *söylenebilir*. Secondly, when compared to Turkish postgraduate students, native postgraduate students statistically significantly overused a total of 32 bundles and underused 9 bundles; as for native scholars, they overused 46 bundles, and underused 7 bundles. In other words, the current study revealed lexical bundles unique to Turkish postgraduate students and those unique to native postgraduate students and scholars. As a result, it can be concluded that in their use of lexical bundles while structuring their texts, Turkish postgraduate students, to a large extent, differed from their native peers and scholars in the area of foreign language teaching research.

As for the structural and functional analysis, the current study did not reveal any statistically significant differences between the three groups of writers included in the research corpus. There are only slight differences in the distribution of lexical bundles in both structural and functional categories, but these were also observed in Wei and Lei (2011). This finding may be due to the Turkish students' presumably high level of English owing to their area of study, i.e. foreign language teaching research. However, the extreme repetitive nature in the Turkish students' text was also observed here. Moreover, in spite of employing similar percentages of functions, they employed different bundles, especially stance bundles. Therefore, it can be deduced that Turkish

postgraduate students employ similar proportions of structures and functions, but they make redundant use of bundles and employ different bundles although they seem to be using lexical bundles functionally and structurally at similar proportions.

5.2. Implications and Suggestions for Teaching

Based on the results of the current study, several implications can be drawn. Although Biber et al. (1999) argues that lexical bundles are very common and easily acquired in the natural discourse of language learning, Turkish postgraduate students whose MA and PhD theses were included in the research corpus seem not to have acquired certain lexical bundles used by native English postgraduates and scholars. According to Cortes (2004), this difference might be due to the lack of formal instruction that students in different disciplines on the frequency and function of such expressions. Regarding formal instruction, Eriksson (2012) suggested that while presenting lexical bundles in class, disciplinarity and specialisation need to be considered when deciding what bundles to include. In this sense, the bundles identified to be commonly used by native students and scholars in the current discipline-specific study can be incorporated in academic writing courses of ELT programs.

Similarly, those bundles found to be used by only Turkish postgraduate students can also be integrated in these courses in a way to make students notice that they can sometimes produce such bundles which may not seem native-like or academic.

As discussed above, Turkish postgraduate students also made redundant use of certain bundles. Incorporating the key bundles reported in studies such as the current study in academic writing classes can enhance students' repertoire of lexical bundles, which may decrease the level of redundancy in their use of lexical bundles.

Several practices can be seen in Cortes (2006) and Eriksson (2012) on how such bundles can be incorporated in teaching. In this regard, functionally related lexical bundles taken from texts in a specific discipline can be introduced to students in contextualized examples. Students can be asked to analyse the functions and possible uses of these bundles. This can be followed by some application exercises including filling in the blanks, multiple choice or inappropriate use correction (Cortes, 2006).

Different from these, students can be asked for their beliefs about usage of lexical bundles. For instance, they can be asked to choose which lexical bundle they think is commonly used in their discipline for a specific function. They can then be asked to use lexical bundles in the context of their own writing (Eriksson, 2012).

5.3. Suggestions for Further Research

As lexical bundles are extremely common in academic prose and their use varies in different disciplines, further studies can investigate lexical bundles in different disciplines so as to guide student writers in their writing processes. Furthermore, the bundles unique to non-native writers or students with the same L1, as revealed in this study, can be investigated elaborately to identify whether it is simply transfer from L1. In addition to using a corpus including texts only in English, a parallel corpus in Turkish can also be combined in a further research, which may explain possible unique uses of Turkish writers in English can be attributed to the nature of Turkish in terms of commonly used words or expressions.

A final suggestion would be on including non-contiguous word combinations along with contiguous word combinations such as lexical bundles in corpus-based studies. For instance, *play a role in* was identified as a four-word lexical bundles in the current study, but since non-contiguous combinations was not our focus, we did not discuss variations such as *play a vital/important/crucial role in*. Since the study of non-contiguous word combinations does not ignore variations within clusters maximizing the uncovering of word associations, it has been very popular in the last few years. Such combinations can also have great pedagogical value as they can serve frames for student writers.

REFERENCES

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

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

Bal, B. (2010). *Analysis of Four-word Lexical Bundles in Published Research Articles Written by Turkish Scholars*. Unpublished MA Thesis. Georgia State University, U.S.A.

Becker, J. (1975). The phrasal lexicon. In R. Shank & B. L. Nash-Webber (Eds.) Theoretical issues in natural language processing. Cambridge, MA: Bolt, Beranek & Newman, 60-63.

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

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical Bundles in University Teaching and Textbooks (A. Wilson, P. Rayson, & T. McEnery, Eds.). *Applied Linguistics, 25(3),* 371–405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Harlow, England: Pearson Education Limited.

Biber, D., Conrad, S. Reppen, P. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Björkman, B. (2013). *English as an Academic Lingua Franca: An Investigation of Form and Communicative Effectiveness.* Berlin/New York: De Gruyter Mouton.

Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum* 1:1-14.

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

Chen, Y. -H. (2009). *Lexical Bundles across Learner Writing Development*. Unpublished

doctoral thesis. Lancaster University, Lancaster, UK.

Chen, Y., & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning & Technology, 14(2),* 30–49.

Cortes, V. (2002). Lexical bundles in Freshman composition. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.), Using corpora to explore linguistic variation (pp. 131–145). Amsterdam: John Benjamins Publishing Company.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. English for Specific Purposes, 23, 397–423.

Cortes, V. 2006. Teaching lexical bundles in the disciplines: An example from a writing intensive history class. Linguistics and Education 17: 391-406.

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

Coulmas, F. (1979). On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics,* 3:239-266.

Cowie, A. (1998). Introduction. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 1–20). Oxford: Oxford University Press.

De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. *Learner English on computer* (pp. 67–79). London: Longman.

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

Eriksson, A. (2012). Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry. In James Thomas & Alex Boulton (eds). Input, Process and Product: Developments in Teaching and Language Corpora. Brno: Masaryk University Press, 195-211.

Flowerdew, L. (2002). The exploitation of small learner corpora in EAP materials. In M.

Ghandessy, A. Henry and R. L. Roseberry (Eds.) Small Corpus Studies and ELT: Theory and Practice. Amsterdam: John Benjamins.

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

Gries, S. (2009). Quantitative Corpus Linguistics with R. New York: Routledge.

Howarth, P. (1998). Phraseology and second language proficiency. Applied Linguistics, 19(1), 24–44.

Hunston, S. (2002). Corpora in Applied Linguistics. Cambridge: Cambridge University Press.

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics, 32,* 150-169.

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

Hyland, K. (2008b). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18(1),* 41–62.

Karabacak, E. & Qin, J. (2012). Comparison of lexical bundles by Turkish, Chinese, and American university students. *Paper presented at Akdeniz Language Studies Conference at Akdeniz University*. Antalya: Turkey.

Koester, A. 2010. Building small specialised corpora. In McCarthy, M. and O'Keeffe, A. (eds), *The Routledge Handbook of Corpus Linguistics*. London: Routledge.

Lewis, M. (2009). The idiom principle in L2 English: Assessing elusive formulaic sequences as indicators of idiomaticity, fluency, and proficiency. Saarbrücken, Germany: VDM Verlag.

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes, 31(1),* 25–35.

McEnery, T. & Wilson, A. (1996). Corpus Linguistics: An Introduction. Edinburgh:

Edinburgh University Press.

Nelson, M. (2010). Building a written corpus: what are the basics? In A. O'Keeffe and M. McCarthy (Eds.) The Routledge Handbook of Corpus Linguistics. New York: Routledge.

Pang, W. (2010). Lexical bundles and the construction of an academic voice: A pedagogical perspective. *Asian EFL Journal, 47,* 1–13.

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. New York: Longman, 191-226.

Reppen, R. (2010). Using Corpora in the Language Classroom. New York: Cambridge University Press.

Ping, 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.

Schmitt, N., Grandage, S. & Adolphs, S. (2004). Are corpus-derived recurrent clusters pscholinguistically valid? In N. Schmitt (Ed.), *Formulaic Sequences* (pp. 127-152). Amsterdam: John Benjamins Publishing.

Scott, M. (2011). *WordSmith Tools version 6*. Liverpool: Lexical Analysis Software.

Scott, M., & Tribble, C. (Eds.). (2006). *Textual Patterns: Key Words and Corpus Analysis in Language Education.* Amsterdam and Philadelphia: John Benjamins B.V.

Simpson-Vlach, R., & Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*, 487–512.

Sinclair, J. (2004). Trust the Text: Language, Corpus and Discourse. London: Routledge.

Sinclair J., (2001). Preface. In Ghadessy M., Henry A. and Roseberry R. (Eds.) *Small corpus studies and ELT*. Amsterdam/Philadelphia: John Benjamins.

Stubbs, M. (2007). An example of frequent English phraseology: Distribution, structures and

functions. In R. Facchinetti (Ed.), *Corpus Linguistics 25 years on* (pp. 89-105). Amsterdam: Radopi.

Tribble, C. (2002). Small corpora and teaching writing. In M. Ghandessy, A. Henry and R. L. Roseberry (Eds.) Small Corpus Studies and ELT: Theory and Practice. Amsterdam: John Benjamins.

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

Yıldız, İ. & Aksan, M. (2013). Türkçe bilimsel metinlerde eylemler: Derlem temelli bir inceleme. 27. Ulusal Dilbilim Kurultayı, 2-4 Mayıs 2013, Antalya, Kemer.

Zamel, V (1998). Questioning Academic Discourse. In V. Zamel and R. Spack (Eds.), Negotiating academic literacies: Teaching and learning across languages and cultures (pp. 187-197). Mahwah, NJ: Erlbaum.

APPENDICES

Appendix I. List of the texts in the research corpus

Texts in Native English Postgraduate Students Sub-Corpus

Writer	Title	University	Code
Thomas Michael Lage (2008)	An exploratory study of computer assisted language learning (CALL) glosses and traditional glosses on incidental vocabulary learning and Spanish literature reading comprehension	Iowa State	NMA-1
Adam M. Russell (2010)	Assessment of strategy inventory of language learning (SILL) in students learning a second language.	Tennessee	NMA-2
Daniel J. Norris (2011)	Attitudes and motivations towards learning foreign languages: A survey of U.S. university students	Southern Illinois	NMA-3
Robert Poole (2011)	Concordance-based Glosses For Facilitating Semantization And Enhancing Productive Knowledge Of Academic Vocabulary	Alabama	NMA-4
Monica Grace Richards (2010)	Developing Academic Vocabulary Independently (DAVI): A usability study	Iowa State	NMA-5
James Robert Garner	Does data-driven learning lead to better academic writing?	Alabama	NMA-6
Sally J. Andrews (2009)	Educational Background As Predictor Of Lexical Richness Among Libyan And Saudi Arabian Esl Students	Pittsburgh	NMA-7
Cheryl Mooney (2010)	Effects of Peer-Tutoring on Vocabulary Recognition, Fluency and Interaction of Low SES ELL Students in a Second Grade Classroom	Caldwell College	NMA-8
Anna Beth Wilkerson (2010)	Electrate Language Learning: An Analysis Of Foreign Language Acquisition In Virtual Environments	Clemson	NMA-9
Tammy L. Johnson (2011)	How does explicit grammar instruction affect students' writing?	California State University	NMA-10
Debra L. Otterby (2009)	Instructional Strategies to Enhance English Language Learners' Vocabulary Acquisiton	Seattle Pasific	NMA-11
Lindsay Clark (2012)	Investigating the Syntagmatic-Paradigmatic Shift in Second Language Speaking Adults	William Paterson	NMA-12
Elissa Kaye Polley (2007)	Learner perceptions of small group and pair work in the ESL classroom: Implications for conditions in second language acquisition	Texas	NMA-13

Anne M. Desiderio (2011)	Pedagogical Implications Of Pragmatic Video Clips In An Efl Context With L1 Arabic Speakers	Michigan State	NMA-14
Kimberly Nicole Mcmillen (2012)	Perceptual Mismatches And The Place Of Culture And Politics In Teaching English: Perspectives Of Six U.s. Efl Teachers In Japan	Colorado	NMA-15
Nicole Lynn Anderson (2011)	Phoneme Awareness And Vocabulary Acquisition In A German-language Classroom	Purdue	NMA-16
Adrienne Marie Johnson (2012)	Processing of wh-movement by second language learners	Kansas	NMA-17
Holly Montgomery (2008)	Self-Reported Listening Strategies by Students in an Intensive English Language Program	Arkansas	NMA-18
Micah William Park (2011)	Teaching Intonation Patterns through Reading Aloud	Portland State	NMA-19
Claire M. Roof (2005)	Testing the Immediate Effects on the Reading Fluency and Comprehension of a Peer-Assisted Learning Strategies-Based Peer-Tutoring Program for English Language Learners	Clemson	NMA-20
Patricia Brannon Bradford (2010)	The acquisition of colloquial speech and slang in second language learners of English in El Paso, Texas	Texas	NMA-21
Wade Hasty (2011)	The Acquisition of Morphology among Child L2 Learners of Spanish: Comparing Pedagogical Interventions	South Carolina	NMA-22
Evelyn Shaw (2009)	The Effectiveness of Games and Activities in Teaching Vocabulary to Adult Learners of English as a Second Language (ESL)	Caldwell College	NMA-23
Sara Strickland Brathwaite (2009)	The efficacy of peer review in a university-level ESL writing class	Alabama	NMA-24
Sarah Huffman (2010)	The influence of collaboration on attitudes towards English vocabulary learning	Iowa State	NMA-25
Catherine E. Showalter (2012)	The Influence Of Novel Orthographic Information On Second Language Word Learning: The Case Of Native English Speakers Learning Arabic	Utah	NMA-26

Jennifer Wood Shand (2008)	The use of drama to reduce anxiety and increase confidence and motivation towards speaking English with two groups of English language learners	Arizona	NMA-27
Kelly Phillip (2009)	Twenty-seven authentically-based ESL grammar supplements: Shifting from form to function	Arkansas	NMA-28
Michelle Marie Priester (2011)	Using Song Lyrics in the Preschool ESL Classroom to Assist Students' English Vocabulary Retention and Use	Caldwell College	NMA-29
Amy Lucile Hammom (2011)	Wondrous Words: Explicit Vocabulary Instruction for Kindergarten English Language Learners	California	NMA-30
Elizabeth Taylor Walden (2008)	A Case Study Of Beliefs And Culturally Relevant Practices Of Four Kindergarten Teachers And Esl Reading Achievement	Capella	NPhD-1
Janet L. Pierce	A co-construction of space trilogy- Examining how ESL teachers, English language learners, and classroom designs interact	Indiana	NPhD-2
Paul Edmunds (2009)	ESL speakers' production of English lexical stress: The effect of variation in acoustic correlates on perceived intelligibility and nativeness	New Mexico	NPhD-3
Jean Louise Ferguson (2009)	Explicit second language vocabulary learning: An investigation of a gloss-embedded text plus form, meaning, and use exercises	Pennsylvania State	NPhD-4
Mary Pyron (2007)	I Hear You, But I Don't Understand You: The Effects Of Peer Tutoring For Helping Secondary Esl Students Achieve Academic Success	Louisiana State	NPhD-5
Kathryn A. Brooks (2006)	In Search of Academic Voice: The Immpact of Instructional Grouping Configurations on English Languagze Learner Academic Language Production	Kansas State	NPhD-6
Jonathan Smart (2012)	Innovative Approaches To Esl Grammar Instruction	Northern Arizona	NPhD-7
Elizabeth A. Specker (2008)	L1/L2 Eye Movement Reading of Closed Captioning: A Multimodal Analysis of Multimodal Use	Arizona	NPhD-8
Laura Jeanne Smith (2009)	Motivation and Long-Term Language Achievement: Understanding Motivation to Persist in Foreign Language Learning	Maryland	NPhD-9
Kara Grace Johnson (2012)	Peer And Self Review: A Holistic Examination Of Efl Learners' Writing And Review Process	Arizona	NPhD-10

Barbara B. Booker (2012)	Perceptions of Female Hispanic ESL Students Toward First-Year College Writing Courses: A Phenomenological Examination of Cultural Influences	South Florida	NPhD-11
Judith L. Otte (2009)	Real Language for Real People: A Descriptive and Exploratory Case Study of the Outcomes of Aural Authentic Texts on the Listening Comprehension of Adult English-as-a-Second Language Students Enrolled in an Advanced ESL Listening Course	Loyola	NPhD-12
Angela Ferguson (2005)	Student Beliefs About Their Foreign Language Instructors: A Look At The Native-speaker/non-native Speaker Issue	Arizona	NPhD-13
Natalie Hudson (2011)	Teacher Gesture In A Post-secondary English As A Second Language Classroom: A Sociocultural Approach	Nevada Las Vegas	NPhD-14
John Patrick Madden (2004)	The Effect of Prior Knowledge on Listening Comprehension in ESL Class Discussions	Texas	NPhD-15
K. James Hartshorn	The effects of manageable corrective feedback on ESL writing accuracy	Brigham Young	NPhD-16
Martine Sabine Sylvain	The Language of Success: A Case Study of the Academic Achievement of ESL Students who Thrive in Spite of Language Barriers	Capella	NPhD-17
Michael David Hubert (2008)	The Relationship Between Writing and Speaking in the U.S. University Spanish Language Classroom	Purdue	NPhD-18
Courtney George (2008)	Toward Political And Ideological Clarity And Care: First Year Esl Teachers And Culturally Responsive Pedagogy	North Carolina	NPhD-19
Duane Eric Paul Leonard (2011)	Why we teach ―ESL‖ Writing: A Socio-Historic Discussion of an Undergraduate ESL Program	California	NPhD-20

Research Articles in Native English Scholars Sub-Corpus

Writer	Title	Vol/Issue	Code
Kimberly L. Geeslin	A Comparison of Copula Choice: Native Spanish Speakers and Advanced Learners	53(4)	NRA-1
Holly Krech Thomas & Alice F. Healy	A Comparison of Rereading Benefits in First and Second Language Reading	62(1)	NRA-2
Norbert Francis	A Componential Approach for Bilingual Reading and Comparative Writing System Research: The Role of Phonology in Chinese Writing as a Test Case	60(4)	NRA-3
Martin Lamb	A Self System Perspective on Young Adolescents' Motivation to Learn English in Urban and Rural Settings	62(4)	NRA-4
Johanne Paradis	Bilingual Children's Acquisition of English Verb Morphology: Effects of Language Exposure, Structure Complexity, and Task Type	60(3)	NRA-5
Nancy Bell	Comparing Playful and Nonplayful Incidental Attention to Form	62(1)	NRA-6
Gerald P. Berent, Ronald R. Kelly, Jeffrey E. Porter & Judith Fonzi	Deaf Learners' Knowledge of English Universal Quantifiers	58(2)	NRA-7
Jette G. Hansen Edwards	Deletion of /t, d/ and the Acquisition of Linguistic Variation by Second Language Learners of English	61(4)	NRA-8
Joe Barcroft	Effects of Opportunities for Word Retrieval During Second Language Vocabulary Learning	57(1)	NRA-9
Marilyn L. Abbott	ESL Reading Strategies: Differences in Arabic and Mandarin Speaker Test Performance	56(4)	NRA-10
Robert Nelson	Expanding the Role of Connectionism in SLA Theory	63(1)	NRA-11
Alison Mackey, Rebecca Adams, Catherine Stafford & Paula Winke	Exploring the Relationship Between Modified Output and Working Memory Capacity	60(3)	NRA-12

Ron I. Thomson	Improving L2 Listeners' Perception of English Vowels: A Computer-Mediated Approach	62(4)	NRA-13
Tracy Hirata-Edds	Influence of Second Language Cherokee Immersion on Children's Development of Past Tense in Their First Language, English	61(3)	NRA-14
R. C. Gardner, A.-M. Masgoret, J. Tennant, L. Mihic	Integrative Motivation: Changes During a Year-Long Intermediate-Level Language Course	54(1)	NRA-15
Alison Mackey, Rhonda Oliver & Jennifer Leeman	Interactional Input and the Incorporation of Feedback: An Exploration of NS-NNS and NNS-NNS Adult and Child Dyads	53(1)	NRA-16
Michael J. Leeser	Learner-Based Factors in L2 Reading Comprehension and Processing Grammatical Form: Topic Familiarity and Working Memory	57(2)	NRA-17
Roderick Edwards & Laura Collins	Lexical Frequency Profiles and Zipf's Law	61(1)	NRA-18
Scott Crossley, Tom Salsbury & Danielle McNamara	Measuring L2 Lexical Growth Using Hypernymic Relationships	59(2)	NRA-19
Alister Cumming	Multiple Dimensions of Academic Language and Literacy Development	63(1)	NRA-20
Alison Mackey and Rebecca Sachs	Older Learners in SLA Research: A First Look at Working Memory, Feedback, and L2 Development	62(3)	NRA-21
Larry Vandergrift	Orchestrating Strategy Use: Toward a Model of the Skilled Second Language Listener	53(3)	NRA-22
Daniel G. Tight	Perceptual Learning Style Matching and L2 Vocabulary Acquisition	60(4)	NRA-23
John N. Williams & Peter Lovatt	Phonological Memory and Rule Learning	53(1)	NRA-24

Antoine Tremblay, Bruce Derwing, Gary Libben & Chris Westbury	Processing Advantages of Lexical Bundles: Evidence From Self-Paced Reading and Sentence Recall Tasks	61(2)	NRA-25
Victoria A. Murphy & Jennifer Hayes	Processing English Compounds in the First and Second Language: The Influence of the Middle Morpheme	60(1)	NRA-26
Paul D. Toth	Processing Instruction and a Role for Output in Second Language Acquisition	56(2)	NRA-27
Carrie Jackson	Proficiency Level and the Interaction of Lexical and Morphosyntactic Information During L2 Sentence Processing	58(4)	NRA-28
John McE. Davis	Resistance to L2 Pragmatics in the Australian ESL Context	57(4)	NRA-29
Kim McDonough & Alison Mackey	Responses to Recasts: Repetitions, Primed Production, and Linguistic Development	56(4)	NRA-30
Thomas Holtgraves	Second Language Learners and Speech Act Comprehension	57(4)	NRA-31
Murray J. Munro & Tracey M. Derwing	Segmental Acquisition in Adult ESL Learners: A Longitudinal Study of Vowel Production	58(3)	NRA-32
Gary Ockey	Self-consciousness and Assertiveness as Explanatory Variables of L2 Oral Ability: A Latent Variable Approach	61(3)	NRA-33
Patrick Bolger & Gabriela Zapata	Semantic Categories and Context in L2 Vocabulary Learning	61(2)	NRA-34
Gregory D. Keating	Sensitivity to Violations of Gender Agreement in Native and Nonnative Spanish: An Eye-Movement Investigation	59(3)	NRA-35
Jeremy Cross	Social-Cultural-Historical Contradictions in an L2 Listening Lesson: A Joint Activity System Analysis	61(3)	NRA-36
Bryan Donaldson	Syntax and Discourse in Near-Native French: Clefts and Focus	62(3)	NRA-37
Paul Seedhouse	Task as Research Construct	55(3)	NRA-38

Susan Gass, Alison Mackey & Lauren Ross-Feldman	Task-Based Interactions in Classroom and Laboratory Settings	55(4)	NRA-39
Paul D. Toth	Teacher- and Learner-Led Discourse in Task-Based Grammar Instruction: Providing Procedural Assistance for L2 Morphosyntactic Development	58(2)	NRA-40
Gillian Stevens	The Age-Length-Onset Problem in Research on Second Language Acquisition Among Immigrants	56(4)	NRA-41
Rod Ellis	The Definition and Measurement of L2 Explicit Knowledge	54(2)	NRA-42
Scott Crossley, Tom Salsbury & Danielle McNamara	The Development of Polysemy and Frequency Use in English Second Language Speakers	60(3)	NRA-43
Luke Plonsky	The Effectiveness of Second Language Strategy Instruction: A Meta-analysis	61(4)	NRA-44
Ryan Deschambault	Thinking-Aloud as Talking-in-Interaction: Reinterpreting How L2 Lexical Inferencing Gets Done	62(1)	NRA-45
Diane Larsen-Freeman	Transfer of Learning Transformed	63(1)	NRA-46
Rick Dale & Michael J. Spivey	Unraveling the Dyad: Using Recurrence Analysis to Explore Patterns of Syntactic Coordination Between Children and Caregivers in Conversation	56(3)	NRA-47
Shawn Loewen	Uptake in Incidental Focus on Form in Meaning-Focused ESL Lessons	54(1)	NRA-48
Batia Laufer & Tina Waldman	Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English	61(2)	NRA-49
Stuart Webb & Michael P. H. Rodgers	Vocabulary Demands of Television Programs	59(2)	NRA-50

Texts in Native English Postgraduate Students Sub-Corpus

Writer	Title	University	Code
Çetin Yıldız, H. (2008)	A Comparative Study Into The Effects Of Two Different Techniques Used To Learn Vocabulary By Turkish Learners Of English At Primary Level	ÇOMU	TMA-1
Tokaç, A. (2005)	A Comparison Of Computer-Assisted Vocabulary Instruction And Teacher-Led Vocabulary Instruction	Bilkent	TMA-2
Karakuş, S. (2005)	A Study On Two Different Grammar Teaching Methods Comparison Of Sentence Level And Context-Based Grammar Teaching	Mersin	TMA-3
Yılmaz, H. (2007)	Comparison Of Teacher-Provided Keyword And Context Methods On Retention Of Vocabulary	Selçuk	TMA-4
Cellat, S. (2008)	Computer Assisted Vocabulary Learning- A Study With Turkish 4th Grade Efl Learners	Anadolu	TMA-5
Kayael, R. (2007)	Do Turkish Teacher Trainees Avoid English Phrasal Verbs?- A Study With The Students Of Elt Department, Anadolu University	Anadolu	TMA-6
Öztuna, S. (2009)	Effects Of Input Flood And Negative Evidence On Learning Of Make/Do Collocations- A Study With Seventh Grade Turkish Efl Students	Anadolu	TMA-7
Aksar, M (2010)	Formulaic Sequences In English Tv Series	Uludağ	TMA-8
Bircan, P. (2010)	Lexical Approach In Teaching Vocabulary To Young Language Learners	Anadolu	TMA-9
Bilgin, Z. (2010)	Long-Term Potentiation In Teaching Vocabulary In Foreign Language A Case Study	METU	TMA-10
Biricik, E. (2010)	Motivating Very Young Learners Of English In A Classroom Setting	Çukurova	TMA-11
Herkemn Şahbaz, Z. (2005)	Needs Assessment Of Academic Reading Tasks And Close Analysis Of Academic Reading Texts For Reading Difficulty And Vocabulary Profile	Bilkent	TMA-12
Tuncer, F. (2005)	Processing Instruction Through Structured Input Activities And Output Practice Activities- A Study On Causative Insruction	Anadolu	TMA-13
Saygı, Ş. (2010)	Reading Motivation In L1 And L2 And Their Relationship With L2 Reading Achievement	METU	TMA-14
Gök, O. (2006)	Rote Learning Versus Deep Processing- The Effect On Vocabulary Learning And Retention	Anadolu	TMA-15

Özlü, H. G. (2009)	Shall We Teach Vocabulary In Lexical Sets, Thematically Related Sets Or Unrelated Sets?	Anadolu	TMA-16
Alptekin (2010)	The Acquisition Of Aorist Passive Voice In Turkish Efl Context- A Comparison Between Processing Instruction And Meaningful Output-Based Instruction	Anadolu	TMA-17
Çiftçi, H. (2009)	The Effect Of Blog Peer Feedback On Turkish Efl Students' Writing Performance And Their Perceptions	Yeditepe	TMA-18
Kesin, M. (2008)	The Effect Of Classroom Learning Environment On Intrinsic Motivation Of Students Learning English As A Foreign Language In Freshman Class At Atilim University	Hacettepe	TMA-19
Yardım, S. (2011)	The Effect Of Computer Assisted And Teacher-Led Storytelling On Vocabulary Learning Of 5th Grade Students	Gazi	TMA-20
Odacı, T. (2006)	The Effect Of Explicit Listening Comprehension Strategy Training On Listening Comprehension Strategy Use And Listening Proficiency Level	Anadolu	TMA-21
Kütük, R. (2007)	The Effect Of Mnemonic Vocabulary Learning Strategy And Story Telling On Young Learners' Vocabulary Learning And Retention	Çukurova	TMA-22
Kasap, B. (2005)	The Effectiveness Of Task-Based Instruction In The Improvement Of Learners' Speaking Skills	Bilkent	TMA-23
Özdem, Z. (2010)	The Effects Of Exposure Frequency And Grammatical Classes Of Words On Receptive And Productive Vocabulary Knowledge Of Efl Learners	Anadolu	TMA-24
Akyıldız Uygun, A. (2009)	The Effects Of Receptive And Productive Tasks On Vocabulary Retention	Anadolu	TMA-25
Mutlu, A. (2008)	The Role Of Call In Promoting Learner Autonomy	METU	TMA-26
Kuru Gönen, S. İ. (2005)	The Sources Of Foreign Language Reading Anxiety Of Students In A Turkish Efl Context	Anadolu	TMA-27
Eylek, Y. (2011)	The Use Of Realia With Turkish Efl Learners	Uludağ	TMA-28
Gedikoğlu, G. (2009)	Using Authentic Newspaper Texts In Teaching Intermediate Vocabulary	Muğla	TMA-29

Güleç, E. (2012)	Using Story Telling Supported By Nlp Techniques In The Teaching Of Vocabulary To Young Learners	Gazi	TMA-30
İstifçi, İ. (2006)	A Descriptive Study On The Styles Of Supervisors In Pre-Observation Conferences	Anadolu	TPhD-1
Caner, M. (2009)	A Study On Blended Learning Model For Teaching Practice Course In Pre-Service English Language Teacher Training Program	Anadolu	TPhD-2
Sarı, R. (2003)	A Suggested English Language Teaching Program For GüLhane Military Medical Academy	METU	TPhD-3
Şimşek, H. (2007)	A Teacher Development Program For Young Learners Of English An Action Research	Çukurova	TPhD-4
Genç, B. (2007)	An Analysis Of Communication Strategies Employed By Turkish-Speakers Of English	Çukurova	TPhD-5
Tokdemir Demirel, E. (2009)	An Investigation Of A Complementary Feedback Model For L2 Writing Peer And Teacher Feedback Versus Teacher Feedback	METU	TPhD-6
Koç, E. M. (2008)	An Investigation Of Cooperating Teachers' Roles As Mentors During The Teaching Practicum At Distance B.A. Program In Elt At Anadolu University Open Education Faculty	Anadolu	TPhD-7
Kafes, H. (2009)	Authorial Stance In Academic English- Native And Non-Native Academic Speaker Writers' Use Of Stance Devices (Modal Verbs) In Research Articles	Anadolu	TPhD-8
Noral, Y. (2009)	Classroom Power Relations In An English As A Foreign Language Setting From A Critical Pedagogical Perspective	İstanbul	TPhD-9
Yüksel, İ. (2012)	Cross-Sectional Evaluation Of Turkish Elt Majors' General And Academic Lexical Competence And Performance	Anadolu	TPhD-10
Yaygın Ersanlı, C. (2010)	Developing Prospective English Language Teachers Comprehension Of Texts With Humorous Elements	Gazi	TPhD-11
Razı, S. (2010)	Effects Of A Metacognitive Reading Program On The Reading Achievement And Metacognitive Strategies	DEU	TPhD-12
Köse, N. (2006)	Effects Of Portfolio Implementation And Assessment On Critical Reading And Learner Autonomy Of Elt Students	Çukurova	TPhD-13
Çopur Şallı, D. (2008)	Teacher Effectiveness In Initial Years Of Service A Case Study On The Graduates Of Metu Foreign Language Education Program	METU	TPhD-14

Erice, D. (2008)	The Impact Of E-Portfolio On The Writing Skills Of Foreign Language Learners Studying At Abant Izzet Baysal University Basic English Program	Gazi	TPhD-15
Bardakçı, M. (2010)	The Impact Of Raising Awareness About Reasoning Fallacies On The Development Of Critical Reading	Gazi	TPhD-16
Özkan, Y. (2005)	The Role Of Input Enhancement In Elt	Çukurova	TPhD-17
Kahraman, A. (2009)	The Role Of L1 Use In Improving Affective And Cognitive Factors In English Language Classrooms	Hacettepe	TPhD-18
Durak Üğüten, S. (2009)	The Use Of Writing Portfolio In Preparatory Writing Classes To Foster Learner Autonomy	Çukurova	TPhD-19
Altunay, D. (2009)	Use Of Connectives In Written Discourse: A Study At An Elt Department In Turkey	Anadolu	TPhD-20

Appendix II. List of the lexical bundles

Native English Postgraduate Students

Word	Freq.	%	Texts	%
THE END OF THE	154,00	0,01	32,00	64,00
IT IS IMPORTANT TO	153,00	0,01	35,00	70,00
AT THE SAME TIME	151,00	0,01	26,00	52,00
AS WELL AS THE	150,00	0,01	34,00	68,00
AT THE END OF	140,00	0,01	33,00	66,00
ON THE OTHER HAND	139,00	0,01	34,00	68,00
THE RESULTS OF THE	125,00	0,01	33,00	66,00
AS A RESULT OF	105,00		28,00	56,00
AT THE BEGINNING OF	92,00		26,00	52,00
THE BEGINNING OF THE	83,00		23,00	46,00
IN THE PRESENT STUDY	82,00		19,00	38,00
IN THE FORM OF	77,00		25,00	50,00
THE RESULTS OF THIS	74,00		27,00	54,00
THE USE OF THE	72,00		24,00	48,00
THE TOTAL NUMBER OF	68,00		19,00	38,00
TO BE ABLE TO	66,00		27,00	54,00
THE PURPOSE OF THIS	62,00		28,00	56,00
THROUGH THE USE OF	61,00		18,00	36,00
TO THE FACT THAT	59,00		21,00	42,00
IN ADDITION TO THE	58,00		29,00	58,00
USED IN THIS STUDY	57,00		22,00	44,00
IN TERMS OF THE	57,00		20,00	40,00
IN A VARIETY OF	54,00		20,00	40,00
THE REST OF THE	54,00		17,00	34,00
IN THE CURRENT STUDY	54,00		14,00	28,00
IN OTHER WORDS THE	53,00		21,00	42,00
IN THE CASE OF	53,00		20,00	40,00
FOR THE PURPOSE OF	50,00		20,00	40,00
IS IMPORTANT TO NOTE	50,00		14,00	28,00
RESULTS OF THIS STUDY	49,00		25,00	50,00
IN THE FOLLOWING EXAMPLE	49,00		7,00	14,00
AT THE TIME OF	48,00		22,00	44,00
THE FACT THAT THE	48,00		21,00	42,00
A GREAT DEAL OF	48,00		14,00	28,00
OF THE PRESENT STUDY	47,00		15,00	30,00
IN THE NEXT SECTION	47,00		11,00	22,00
THE MAJORITY OF THE	45,00		20,00	40,00
THE ROLE OF THE	45,00		15,00	30,00
PURPOSE OF THIS STUDY	44,00		27,00	54,00
IN THE CONTEXT OF	44,00		21,00	42,00
ON THE PART OF	44,00		16,00	32,00
THE WAY IN WHICH	44,00		11,00	22,00
CAN BE FOUND IN	43,00		18,00	36,00
IN AN ATTEMPT TO	42,00		13,00	26,00

IN A WAY THAT	41,00	19,00	38,00
FOR THE PURPOSES OF	41,00	17,00	34,00
AS WELL AS A	40,00	23,00	46,00
ONE OF THE MOST	40,00	19,00	38,00
AS A RESULT THE	40,00	17,00	34,00
FOR EACH OF THE	39,00	20,00	40,00
THE PART OF THE	39,00	16,00	32,00
IMPORTANT TO NOTE THAT	39,00	15,00	30,00
I WAS ABLE TO	39,00	12,00	24,00
IN AN EFFORT TO	38,00	19,00	38,00
HAS BEEN SHOWN TO	38,00	16,00	32,00
DUE TO THE FACT	38,00	14,00	28,00
ARE MORE LIKELY TO	37,00	18,00	36,00
THAT THERE IS A	37,00	17,00	34,00
IT WAS FOUND THAT	37,00	11,00	22,00
THE NATURE OF THE	36,00	19,00	38,00
AS PART OF THE	35,00	14,00	28,00
THE COURSE OF THE	34,00	19,00	38,00
IN ORDER TO UNDERSTAND	34,00	16,00	32,00
THE WAYS IN WHICH	34,00	13,00	26,00
THE CONTEXT OF THE	33,00	17,00	34,00
IT IS POSSIBLE THAT	33,00	15,00	30,00
IN ORDER TO DETERMINE	33,00	14,00	28,00
THE ANALYSIS OF THE	33,00	13,00	26,00
THE PURPOSE OF THE	32,00	23,00	46,00
IN ORDER TO BE	32,00	18,00	36,00
ANALYSIS OF THE DATA	32,00	15,00	30,00
CAN BE SEEN IN	32,00	13,00	26,00
THAT THERE WAS A	31,00	19,00	38,00
AS A WAY TO	31,00	17,00	34,00
BY THE END OF	31,00	16,00	32,00

Native English Scholars

Word	Freq.	%	Texts	%
IN THE CURRENT STUDY	83,00	0,02	17,00	34,00
IN THE PRESENT STUDY	65,00	0,01	22,00	44,00
THE EXTENT TO WHICH	61,00	0,01	25,00	50,00
THE RESULTS OF THE	54,00	0,01	24,00	48,00
ON THE OTHER HAND	47,00	0,01	25,00	50,00
IN THE CASE OF	46,00	0,01	24,00	48,00
THE END OF THE	42,00		21,00	42,00
IT IS IMPORTANT TO	39,00		22,00	44,00
ON THE BASIS OF	38,00		20,00	40,00
THE NATURE OF THE	37,00		21,00	42,00
IT IS POSSIBLE THAT	36,00		22,00	44,00
AT THE END OF	36,00		19,00	38,00
FOR EACH OF THE	35,00		21,00	42,00
AT THE SAME TIME	35,00		20,00	40,00
IN THE CONTEXT OF	32,00		20,00	40,00
THE RESULTS OF THIS	30,00		16,00	32,00
IN THE FORM OF	28,00		13,00	26,00
OF THE CURRENT STUDY	28,00		13,00	26,00
AS WELL AS THE	27,00		19,00	38,00
RESULTS OF THIS STUDY	26,00		14,00	28,00
IT IS CLEAR THAT	25,00		13,00	26,00
AS A FUNCTION OF	25,00		10,00	20,00
OF THE PRESENT STUDY	24,00		14,00	28,00
THE TOTAL NUMBER OF	24,00		12,00	24,00
WITH RESPECT TO THE	24,00		10,00	20,00
THE FACT THAT THE	22,00		13,00	26,00
WERE MORE LIKELY TO	22,00		8,00	16,00
OVER THE COURSE OF	21,00		11,00	22,00
AS A RESULT OF	20,00		14,00	28,00
IN ADDITION TO THE	20,00		14,00	28,00
WITH THE EXCEPTION OF	20,00		14,00	28,00
THE EFFECT OF THE	20,00		5,00	10,00
TO ENSURE THAT THE	19,00		16,00	32,00
ARE PRESENTED IN TABLE	19,00		11,00	22,00
IN A WAY THAT	18,00		11,00	22,00
THE DEGREE TO WHICH	18,00		9,00	18,00
IN CONTRAST TO THE	17,00		14,00	28,00
IN THE SAME WAY	17,00		13,00	26,00
AT THE TIME OF	17,00		12,00	24,00
USED IN THIS STUDY	17,00		12,00	24,00
A NUMBER OF STUDIES	17,00		11,00	22,00
IN RELATION TO THE	17,00		9,00	18,00
THERE WAS ALSO A	17,00		7,00	14,00
AT THE BEGINNING OF	16,00		14,00	28,00
THAT THERE IS A	16,00		13,00	26,00
IT SHOULD BE NOTED	16,00		12,00	24,00
IN TERMS OF THE	16,00		11,00	22,00
CAN BE SEEN IN	16,00		10,00	20,00
THE PURPOSE OF THIS	16,00		10,00	20,00

THE PURPOSE OF THE	15,00	12,00	24,00
IN THE FIELD OF	15,00	10,00	20,00
TO THE FACT THAT	15,00	10,00	20,00
THE CONTEXT OF THE	15,00	9,00	18,00
AS CAN BE SEEN	15,00	7,00	14,00
IT IS POSSIBLE TO	15,00	7,00	14,00
ARE SUMMARIZED IN TABLE	15,00	6,00	12,00
SHOULD BE NOTED THAT	14,00	11,00	22,00
THE BEGINNING OF THE	14,00	10,00	20,00
IMPORTANT TO NOTE THAT	14,00	9,00	18,00
THESE RESULTS SUGGEST THAT	14,00	9,00	18,00
ARE MORE LIKELY TO	13,00	11,00	22,00
A GREATER NUMBER OF	13,00	10,00	20,00
IS IMPORTANT TO NOTE	13,00	9,00	18,00
THE FOCUS OF THE	13,00	9,00	18,00
TO THE EXTENT THAT	13,00	9,00	18,00
IN THE ABSENCE OF	13,00	8,00	16,00
THE ANALYSIS OF THE	13,00	8,00	16,00
WAS FOUND TO BE	13,00	8,00	16,00
THE WAYS IN WHICH	13,00	6,00	12,00
BEYOND THE SCOPE OF	12,00	12,00	24,00
PLAY A ROLE IN	12,00	11,00	22,00
IN THE USE OF	12,00	9,00	18,00
THAT THE NUMBER OF	12,00	7,00	14,00
IT IS LIKELY THAT	12,00	5,00	10,00
IT IS DIFFICULT TO	11,00	10,00	20,00
THE SIZE OF THE	11,00	10,00	20,00
WITH REGARD TO THE	11,00	10,00	20,00
IT MAY BE THAT	11,00	9,00	18,00
IN AN ATTEMPT TO	11,00	8,00	16,00
IN ANY OF THE	11,00	7,00	14,00
IN A STUDY OF	11,00	6,00	12,00
TO BE RELATED TO	11,00	6,00	12,00
FROM THE CURRENT STUDY	11,00	5,00	10,00

Turkish Postgraduate Students

Word	Freq.	%	Texts	%
AT THE END OF	567,00	0,04	40,00	80,00
THE END OF THE	553,00	0,04	39,00	78,00
ON THE OTHER HAND	503,00	0,04	47,00	94,00
THE RESULTS OF THE	357,00	0,03	47,00	94,00
THE BEGINNING OF THE	325,00	0,02	37,00	74,00
AS A RESULT OF	324,00	0,02	40,00	80,00
AT THE BEGINNING OF	311,00	0,02	38,00	76,00
END OF THE STUDY	310,00	0,02	16,00	32,00
BEGINNING OF THE STUDY	198,00	0,01	20,00	40,00
THE ANALYSIS OF THE	182,00	0,01	32,00	64,00
OF THE PRESENT STUDY	177,00	0,01	26,00	52,00
IN TERMS OF THE	166,00	0,01	33,00	66,00
IN THE PRESENT STUDY	155,00	0,01	28,00	56,00
A RESULT OF THE	147,00	0,01	24,00	48,00
WITH THE HELP OF	129,00		24,00	48,00
AT THE SAME TIME	129,00		34,00	68,00
THE FINDINGS OF THE	122,00		34,00	68,00
IN THE LIGHT OF	119,00		28,00	56,00
TO BE ABLE TO	110,00		39,00	78,00
ONE OF THE MOST	109,00		40,00	80,00
IN THE USE OF	104,00		18,00	36,00
TO FIND OUT THE	102,00		31,00	62,00
THAT THERE IS A	101,00		33,00	66,00
IS ONE OF THE	92,00		38,00	76,00
AS CAN BE SEEN	87,00		21,00	42,00
CAN BE SEEN IN	86,00		25,00	50,00
AS WELL AS THE	86,00		30,00	60,00
RESULTS OF THE STUDY	83,00		28,00	56,00
IS CONSIDERED TO BE	83,00		21,00	42,00
IN ADDITION TO THE	83,00		26,00	52,00
ON THE USE OF	82,00		23,00	46,00
BY THE HELP OF	82,00		17,00	34,00
IN ORDER TO FIND	80,00		25,00	50,00
IN ORDER TO SEE	79,00		25,00	50,00
IN THE FIELD OF	78,00		26,00	52,00
THE FACT THAT THE	77,00		31,00	62,00
THE AIM OF THE	77,00		29,00	58,00
TO FIND OUT WHETHER	76,00		28,00	56,00
IN THE FORM OF	74,00		24,00	48,00
IT CAN BE CONCLUDED	73,00		23,00	46,00
THE RESULTS OF THIS	69,00		27,00	54,00
IT WAS FOUND THAT	69,00		16,00	32,00
IN OTHER WORDS THE	69,00		29,00	58,00
THAT MOST OF THE	68,00		25,00	50,00
THE PURPOSE OF THE	66,00		30,00	60,00
IT CAN BE SAID	66,00		26,00	52,00
CAN BE SAID THAT	65,00		26,00	52,00
CAN BE CONCLUDED THAT	64,00		22,00	44,00
THAT THERE WAS A	63,00		23,00	46,00

IN LINE WITH THE	63,00	16,00	32,00
THAT THE USE OF	61,00	23,00	46,00
OF THE FACT THAT	61,00	25,00	50,00
THE USE OF THE	59,00	24,00	48,00
IN ADDITION TO THIS	59,00	19,00	38,00
ACCORDING TO THE RESULTS	59,00	23,00	46,00
TO THE FACT THAT	58,00	26,00	52,00
FINDINGS OF THE STUDY	57,00	26,00	52,00
THE FINDINGS OF THIS	56,00	21,00	42,00
THE NUMBER OF THE	55,00	27,00	54,00
THE NATURE OF THE	55,00	25,00	50,00
ON THE BASIS OF	55,00	25,00	50,00
RESULTS OF THIS STUDY	54,00	22,00	44,00
IT IS POSSIBLE TO	54,00	23,00	46,00
IT IS SEEN THAT	53,00	16,00	32,00
FINDINGS OF THIS STUDY	52,00	20,00	40,00
IT WAS SEEN THAT	51,00	16,00	32,00
IT CAN BE CLAIMED	51,00	7,00	14,00
THE ROLE OF THE	49,00	24,00	48,00
THE RESULTS SHOWED THAT	49,00	18,00	36,00
ANALYSIS OF THE DATA	49,00	19,00	38,00
A WIDE RANGE OF	49,00	21,00	42,00
IT CAN BE SEEN	48,00	15,00	30,00
IN THE SENSE THAT	48,00	11,00	22,00
IN THE PROCESS OF	48,00	26,00	52,00
CAN BE CLAIMED THAT	48,00	7,00	14,00
BE DUE TO THE	48,00	13,00	26,00
AS SEEN IN TABLE	48,00	12,00	24,00
THE RESULTS INDICATED THAT	47,00	16,00	32,00
IT IS NECESSARY TO	47,00	23,00	46,00
USED IN THE STUDY	46,00	20,00	40,00
IN ORDER TO MAKE	46,00	27,00	54,00
IN ORDER TO BE	46,00	28,00	56,00
BE TAKEN INTO CONSIDERATION	46,00	22,00	44,00
THE STUDY SHOWED THAT	45,00	10,00	20,00
ORDER TO FIND OUT	45,00	18,00	36,00
IT WAS OBSERVED THAT	45,00	13,00	26,00
IT IS IMPORTANT TO	45,00	26,00	52,00
FOR THE PURPOSE OF	45,00	21,00	42,00
A PART OF THE	45,00	19,00	38,00
USED IN THIS STUDY	44,00	22,00	44,00
THROUGH THE USE OF	44,00	20,00	40,00
RESULT OF THE ANALYSIS	44,00	5,00	10,00
IN THE SAME WAY	44,00	23,00	46,00
IN ORDER TO UNDERSTAND	44,00	15,00	30,00
DUE TO THE FACT	44,00	21,00	42,00
AS ONE OF THE	44,00	24,00	48,00
IT IS BELIEVED THAT	43,00	19,00	38,00
WITH REGARD TO THE	42,00	13,00	26,00
FIRST OF ALL THE	42,00	24,00	48,00
AS IN THE FOLLOWING	42,00	13,00	26,00

IN ACCORDANCE WITH THE	40,00	20,00	40,00
WITH REFERENCE TO THE	39,00	11,00	22,00
THE REST OF THE	39,00	18,00	36,00
IN THE CONTEXT OF	39,00	25,00	50,00
BY MEANS OF THE	39,00	8,00	16,00
AS MUCH AS POSSIBLE	39,00	11,00	22,00
WE CAN SAY THAT	38,00	7,00	14,00
ORDER TO SEE THE	38,00	16,00	32,00
IN THE FOLLOWING SECTION	38,00	14,00	28,00
IN ADDITION TO THESE	38,00	20,00	40,00
A RESULT OF THIS	38,00	14,00	28,00
THE STUDY WAS CONDUCTED	37,00	22,00	44,00
CAN BE USED IN	37,00	22,00	44,00
ARE CONSIDERED TO BE	37,00	12,00	24,00
THE DESIGN OF THE	36,00	19,00	38,00
PURPOSE OF THE STUDY	36,00	20,00	40,00
CAN BE USED TO	36,00	22,00	44,00
AIM OF THIS STUDY	36,00	22,00	44,00
WITH THE USE OF	35,00	15,00	30,00
WITH THE AIM OF	35,00	15,00	30,00
IN RELATION TO THE	35,00	14,00	28,00
CAN BE SEEN FROM	35,00	10,00	20,00
BE CLAIMED THAT THE	35,00	5,00	10,00
WILL BE ABLE TO	34,00	13,00	26,00
WHETHER THERE WAS A	34,00	12,00	24,00
TO BE USED IN	34,00	19,00	38,00
THE PURPOSE OF THIS	34,00	20,00	40,00
MAY BE DUE TO	34,00	11,00	22,00
IT SHOULD BE NOTED	34,00	8,00	16,00
CAN BE REGARDED AS	34,00	10,00	20,00
BE SEEN IN TABLE	34,00	15,00	30,00
AS IT IS SEEN	34,00	11,00	22,00
AS IT CAN BE	34,00	9,00	18,00

Appendix III. Chi-square test for structural differences

sub_corpus * structure Crosstabulation

			structure											
			NP_of	NP_other	PP_of	other_PP	ant_it	pass_PP	copula_P	VP_that	V_A_to	adv	pronoun	others	Total
sub_corpus	TR_PGS	Count	32	1	19	23	9	9	2	15	12	3	0	8	133
		Expected Count	31.5	3.7	23.3	27.4	8.7	7.8	2.7	9.6	9.1	1.8	.9	6.4	133.0
		% within sub_corpus	24.1%	0.8%	14.3%	17.3%	6.8%	6.8%	1.5%	11.3%	9.0%	2.3%	0.0%	6.0%	100.0%
		% within structure	46.4%	12.5%	37.3%	38.3%	47.4%	52.9%	33.3%	71.4%	60.0%	75.0%	0.0%	57.1%	45.7%
		% of Total	11.0%	0.3%	6.5%	7.9%	3.1%	3.1%	0.7%	5.2%	4.1%	1.0%	0.0%	2.7%	45.7%
		Std. Residual	.1	-1.4	-.9	-.8	.1	.4	-.4	1.7	.9	.9	-1.0	.6	
	N_PGS	Count	21	3	15	17	2	3	2	3	5	0	1	3	75
		Expected Count	17.8	2.1	13.1	15.5	4.9	4.4	1.5	5.4	5.2	1.0	.5	3.6	75.0
		% within sub_corpus	28.0%	4.0%	20.0%	22.7%	2.7%	4.0%	2.7%	4.0%	6.7%	0.0%	1.3%	4.0%	100.0%
		% within structure	30.4%	37.5%	29.4%	28.3%	10.5%	17.6%	33.3%	14.3%	25.0%	0.0%	50.0%	21.4%	25.8%
		% of Total	7.2%	1.0%	5.2%	5.8%	0.7%	1.0%	0.7%	1.0%	1.7%	0.0%	0.3%	1.0%	25.8%
		Std. Residual	.8	.7	.5	.4	-1.3	-.7	.4	-1.0	-.1	-1.0	.7	-.3	
	N_S	Count	16	4	17	20	8	5	2	3	3	1	1	3	83
		Expected Count	19.7	2.3	14.5	17.1	5.4	4.8	1.7	6.0	5.7	1.1	.6	4.0	83.0
		% within sub_corpus	19.3%	4.8%	20.5%	24.1%	9.6%	6.0%	2.4%	3.6%	3.6%	1.2%	1.2%	3.6%	100.0%
		% within structure	23.2%	50.0%	33.3%	33.3%	42.1%	29.4%	33.3%	14.3%	15.0%	25.0%	50.0%	21.4%	28.5%
		% of Total	5.5%	1.4%	5.8%	6.9%	2.7%	1.7%	0.7%	1.0%	1.0%	0.3%	0.3%	1.0%	28.5%
		Std. Residual	-.8	1.1	.6	.7	1.1	.1	.2	-1.2	-1.1	-.1	.6	-.5	
Total		Count	69	8	51	60	19	17	6	21	20	4	2	14	291
		Expected Count	69.0	8.0	51.0	60.0	19.0	17.0	6.0	21.0	20.0	4.0	2.0	14.0	291.0
		% within sub_corpus	23.7%	2.7%	17.5%	20.6%	6.5%	5.8%	2.1%	7.2%	6.9%	1.4%	0.7%	4.8%	100.0%
		% within structure	100%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	23.7%	2.7%	17.5%	20.6%	6.5%	5.8%	2.1%	7.2%	6.9%	1.4%	0.7%	4.8%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	23.757[a]	22	.360
Likelihood Ratio	26.645	22	.225
Linear-by-Linear Association	3.282	1	.070
N of Valid Cases	291		

a. 17 cells (47.2%) have expected count less than 5. The minimum expected count is .52.

Appendix IV. Chi-square test for functional differences

sub_corpus * functions Crosstabulation

			functions			
			stance	discourse	referential	Total
sub_corpus	TR_PGS	Count	23	42	68	133
		Expected Count	22.4	33.8	76.8	133.0
		% within sub_corpus	17.3%	31.6%	51.1%	100.0%
		% within functions	46.9%	56.8%	40.5%	45.7%
		% of Total	7.9%	14.4%	23.4%	45.7%
	N_PGS	Count	11	18	46	75
		Expected Count	12.6	19.1	43.3	75.0
		% within sub_corpus	14.7%	24.0%	61.3%	100.0%
		% within functions	22.4%	24.3%	27.4%	25.8%
		% of Total	3.8%	6.2%	15.8%	25.8%
	N_S	Count	15	14	54	83
		Expected Count	14.0	21.1	47.9	83.0
		% within sub_corpus	18.1%	16.9%	65.1%	100.0%
		% within functions	30.6%	18.9%	32.1%	28.5%
		% of Total	5.2%	4.8%	18.6%	28.5%
Total		Count	49	74	168	291
		Expected Count	49.0	74.0	168.0	291.0
		% within sub_corpus	16.8%	25.4%	57.7%	100.0%
		% within functions	100.0%	100.0%	100.0%	100.0%
		% of Total	16.8%	25.4%	57.7%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6.678[a]	4	.154
Likelihood Ratio	6.876	4	.143
Linear-by-Linear Association	1.734	1	.188
N of Valid Cases	291		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 12.63.