

11-30-2010

Analysis of Four-word Lexical Bundles in Published Resesarch Articles Written by Turkish Scholars

Betul Bal
Georgia State University

Follow this and additional works at: http://scholarworks.gsu.edu/alesl_theses

Recommended Citation

Bal, Betul, "Analysis of Four-word Lexical Bundles in Published Resesarch Articles Written by Turkish Scholars." Thesis, Georgia State University, 2010.
http://scholarworks.gsu.edu/alesl_theses/2

This Thesis is brought to you for free and open access by the Department of Applied Linguistics and English as a Second Language at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Applied Linguistics and English as a Second Language Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

ANALYSIS OF FOUR- WORD LEXICAL BUNDLES IN PUBLISHED RESEARCH
ARTICLES WRITTEN BY TURKISH SCHOLARS

by

BETUL BAL

Under the Direction of Viviana Cortes

ABSTRACT

This study investigated the use of lexical bundles in research articles written in English by Turkish scholars. For the purpose of the study, a corpus of published research articles produced by Turkish scholars in six different academic disciplines was collected. The four-word lexical bundles that appeared at least twenty times in this one million word corpus were identified and further analyzed both structurally and functionally based on the previous taxonomies developed by Biber, Johansson, Leech, Conrad and Finegan (1999) and Biber, Conrad and Cortes (2004). The results of this study revealed that the lexical bundles found have structural correlates as well as strong functional features that help to construct discourse in academic writing. The conclusions drawn from this study could be applied to the teaching of academic genres to researchers in English as a Foreign Language context and are expected to provide insights for further corpus-based studies in academic writing.

INDEX WORDS: Lexical bundles, Research articles, Corpus, Academic writing, Corpus-based studies

ANALYSIS OF FOUR- WORD LEXICAL BUNDLES IN PUBLISHED RESEARCH
ARTICLES WRITTEN BY TURKISH SCHOLARS

by

BETUL BAL

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Art Education

in the College of Arts and Sciences

Georgia State University

2010

Copyright by

Betul Bal

2010

ANALYSIS OF FOUR- WORD LEXICAL BUNDLES IN PUBLISHED RESEARCH
ARTICLES WRITTEN BY TURKISH SCHOLARS

by

BETUL BAL

Committee Chair: Viviana Cortes

Committee: Diane Belcher

Eric Friginal

YouJin Kim

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2010

ACKNOWLEDGEMENTS

It is a pleasure to thank the many people who made this thesis possible.

First, I wish to express my gratitude to the Fulbright Commission for the support that they gave me in order to study in the U.S.A. and to the Department of Applied Linguistics and ESL at Georgia State University for helping me write a master's thesis.

It is difficult to overstate my gratitude and appreciation to Dr. Viviana Cortes, Chair of my committee and my thesis advisor, who has been and always will be an inspiration for me in my academic career. It would have been impossible for me to complete this thesis without her patience, constructive feedback, and insightful advice. I have benefited a lot from her stimulating ideas and suggestions. I am truly lucky to be her student.

Besides, I would like to thank to my committee members, Dr. Diane Belcher, Dr. Eric Friginal, and Dr. YouJin Kim, for the time they have dedicated to the reading of my thesis and all the valuable comments they have made. I have learned a lot from each of them during my time at Georgia State, and I know I will continue to do so in the future.

Moreover, my heartfelt thanks go to my family for their invaluable support and encouragement. I am grateful to my parents, Hikmet Bal and Saim Bal, who not only raised me, taught me, and loved me but also believed in me and supported every step I took in my life. I am also thankful to my sister and brother for their presence and moral support.

Last but not least, I am forever grateful to my significant other, Cenk, for his encouragement, understanding, endless patience, and for his unconditional love when it was most required.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS	ix
CHAPTER 1. INTRODUCTION	1
1.1 Purpose of the Study	2
1.2 Research Questions	3
1.3 Organization of the Study	3
CHAPTER 2. LITEARUTE REVIEW	4
2.1 Definition of Corpus and Corpus-Based Studies	5
2.2 Formulaic Language and Corpora	6
2.3 Lexical Bundles and Register Variations	9
CHAPTER 3. METHODOLOGY	14
3.1 The TSRA Corpus	15
3.2 Concordancing Software: AntConc	17
3.3 Structural and Functional Taxonomies	21
CHAPTER 4. RESULTS and DISCUSSION	25
4.1 TSRAC Lexical Bundles	26
4.2 Structural Analysis of TSRAC Lexical Bundles	27
4.3 Functional Analysis of TSRAC Lexical Bundles	28
4.3.1 Stance Bundles	31
4.3.2 Discourse Organizers	33
4.3.3 Referential Expressions	33

CHAPTER 5. CONCLUSION	37
5.1 Summary of the Results	37
5.2 Limitations	38
5.3 Implications	39
5.4 Suggestions for Further Research	39
REFERENCES	41
APPENDIX A: Journals Used in the TSRAC	49
APPENDIX B: TSRAC Lexical Bundles	54

LIST OF TABLES

Table 2.1 Major studies on lexical bundles

Table 3.1 Disciplines in the TSRAC

Table 3.2 Structural types of lexical bundles (Biber et al., p.1015)

**Table 3.3 Functional classification of lexical bundles (Biber, Conrad and Cortes, 2004
p.384)**

Table 4.1 Lexical bundles in TSRAC according to their functions in context

LIST OF FIGURES

Figure 1. AntConc screenshot showing the TSRAC bundles (Anthony, 2007)

Figure 2. AntConc screenshot showing the concordances (Anthony, 2007)

Figure 3. Structural distribution of TSRAC lexical bundles

LIST OF ABBREVIATIONS

TSRA: Turkish Scholars' Research Articles

TSRAC: Turkish Scholars' Research Articles Corpus

LSWE: Longman Spoken and Written English

T2K-SWAL: TOEFL 2000 Spoken and Written Academic Language

BASE: British Academic Spoken English

MICASE: Michigan Corpus of Academic Spoken English Corpus

CHAPTER 1. INTRODUCTION

Writing for academic purposes is a challenging journey since creating texts to convey one's ideas in this environment requires special attention and effort. As stated by Zamel (1998), academic discourse has its distinguishing features "because it appears to require a kind of language with its own vocabulary, norms, sets of conventions, and modes of inquiry, academic discourse has come to characterize a separate culture..." (p.187). Therefore, throughout the history of language studies, there have been many investigations that focused on finding these distinguishing features of academic writing. As cited by Biber (2006), the majority of these studies focused on different aspects of academic writing such as expressions of stance (Charles, 2003; Crompton, 1997; Grabe & Kaplan, 1997; Holmes, 1986; Hyland, 1994, 1996a, b; Meyer, 1997; Myers, 1989, 1990; Salager-Meyer, 1994; Silver, 2003; Varttala, 2003); academic registers (Flowerdew, 2002; Hewings, 2001); verb classes (e.g., Hunston, 1995), and the organization of discourse (Ferguson, 2001), to mention only a few. Academic vocabulary is also one of the features that attracted attention, and analyzing academic vocabulary has been the purpose of numerous studies (Coxhead, 2000; Nation, 1990, 2001; Schmitt & McCarthy, 1997). Lately, there has been a shift from studying single lexical items to studying multi-word expressions. Therefore, studies have begun to go beyond the analysis of single lexical items and focused on formulaic expressions (Altenberg, 1998; Biber, Johansson, Leech, Conrad and Finegan, 1999; Nattinger and DeCarrico, 1992; Pawley and Syder 1983). All these studies highlight the significance of these fixed expressions which perform particular structural forms and strong discourse functions. In the light of previous research on the presence and significance of formulaic language in academic prose, the present study focuses on a particular multi-word expression which is called "lexical bundle" (Biber et al., 1999). Altenberg (1998) is considered

to be one of the first researchers to study recurrent word combinations using empirical-based methods. Drawing on his work, Biber et al. (1999) focused on the study of recurrent expressions that they called lexical bundles. Lexical bundles have been the focus of various further studies (Biber et al., 1999, 2003, 2004; Butler, 1997). Studies on lexical bundles in various English registers have presented noteworthy and prominent results looking from different perspectives in different registers. The common conclusion drawn from these studies on lexical bundles in academic writing is that lexical bundles constitute a large part of academic texts and they have structural correlates as well as significant discourse functions that help to construct the text itself (Biber et al. 1999, 2003, 2004; Biber, Conrad & Cortes, 2003; Cortes 2002, 2004).

Most of these studies on lexical bundles are based on lexical bundles in English with the exception of two recent studies: Cortes (2008) includes Spanish in her analysis of lexical bundles in academic history writing, and Kim (2009) analyzes the use of lexical bundles both in academic and spoken registers in Korean. However, although there are studies that go beyond the use of lexical bundles in languages other than English, little is known about the use of lexical bundles by non-native speakers of English when they speak or write in English. Therefore, the idea to investigate the lexical bundles produced by non-native speakers of English in their published academic writing became the impetus for this study.

1.1 Purpose of the Study

The main objective of the present study is to identify the four-word lexical bundles used by Turkish scholars who are non-native speakers of English. The academic texts used for the purpose of the study are published research articles in international journals from six different academic disciplines written by Turkish scholars. The lexical bundles identified are compared

with the bundles previously identified in several studies from the literature that analyzed lexical bundles in different academic registers (Biber et al. 1999; Biber and Conrad, 1999; Biber, Conrad and Cortes, 2003, 2004; Cortes, 2004; 2008). Moreover, using both quantitative and qualitative analyses, this study aims to further investigate these lexical bundles in terms of their structures and functions based on the taxonomies that have been previously designed and used for the classification of lexical bundles (Cortes 2002, 2004; Biber, Conrad & Cortes, 2003).

1.2 Research Questions

In order to reach a comprehensive analysis of lexical bundles used by Turkish scholars when they write research articles in English, this study will explore the following research questions:

1. What are the most common four-word lexical bundles found in published research articles written by Turkish scholars?
2. How much do these lexical bundles have in common with those bundles previously identified in the literature?
3. What are the structural and functional features of the lexical bundles found in this study?

1.3 Organization of the Study

To address these research questions, Chapter 2 will provide background information on the meaning of corpora and how corpus-based studies are conducted. Then the significance of formulaic language in academic writing will be presented as well as a description of lexical bundles and recent corpus-based studies on these expressions. In Chapter 3, will introduce the procedures followed for the compilation of corpus data, the computer software used together

with the quantitative and qualitative analyses conducted, and the taxonomies used for the analysis of the lexical bundles identified in this study. The characteristics of the TSRAC lexical bundles will be introduced in Chapter 4, together with a detailed report of the results of the analyses. To conclude, Chapter 5 will offer a brief summary of the study and its results, followed by its limitations. Then implications for language teachers and researchers, and suggestions for further studies on lexical bundles will also be provided in this final chapter.

CHAPTER 2. LITERATURE REVIEW

This chapter will provide background information for the present study by presenting two sections: first, an introduction to corpora and corpus-based studies and second, a literature review on studies of formulaic language in academic discourse followed by a detailed review of recent corpus-based studies on lexical bundles in academic prose which are closely related to the present study.

2.1 Definition of Corpus and Corpus-Based Studies

As a Latin rooted word, *corpus* means *body* which, when used in the linguistics field, refers to a “body of texts”. In today’s world, however, in the field of Applied Linguistics, the term corpus is related to a large collection of machine-readable texts. As cited by McEnery and Wilson (1996), some corpus-based studies were conducted in the past centuries (Eaton, 1940; Fries and Traver, 1940; Preyer 1889; Kading, 1897). The actual meaning of corpus-based research, however, refers to studies where a machine-readable corpus is created and computer software is used to analyze it.

As Conrad (1996) states, there are certain important characteristics of corpus-based investigations that need to be emphasized. Corpus-based studies

- (a) are based on principled collections of naturally occurring texts (the corpus),
- (b) use computers for both automatic and interactive analyses, and
- (c) include both quantitative analyses and functional interpretations in order to describe patterns in language features.

As these features suggest, in a corpus-based study, once the corpus is collected, a concordancing program, for example, may be used to allow the researcher to search the target

item or items in the corpus. These programs provide lists of lines/concordances in which the target item occurs, which enables further analysis. When automatic quantitative analyses such as frequency lists, collocations etc. are retrieved, more qualitative interpretations are made based on these findings. The target item in the corpus depends on the purpose of the study. While it can be a specific language feature such as complex noun phrases (Vande Kopple, 1992), it can also be writer attitude as in the example presented by Salager - Meyer (1992).

One of the first modern corpus-based analysis projects was begun by Francis and Kucera at Brown University in 1961. This project deserves mention as it is the first major computational corpus project. It was a one million word corpus known as the *Brown Corpus* drawn from randomly sampled materials written in American English in 1961 in a variety of genres. It has inspired many other corpus studies as representing a significant step from non-digital to digital corpus-based investigations.

2.2 Formulaic Language and Corpora

In recent years, an increasing number of studies have made use of corpus data to analyze formulaic expressions used in different registers. Academic registers¹ have become one of the registers that attracted attention of linguists. Research on defining and processing formulaic language in academic prose has been the purpose of many studies, starting with the study of Pawley and Syder (1983), followed by Nattinger and DeCarrico (1992) and more recently Biber et al. (1999), Wray (2000, 2002) and Cortes (2002, 2004, 2008) to mention only a few. The latest trend in the study of formulaic language in academic writing has focused on a particular type of

¹ All the analyses conducted in this study used a register-based perspective, defining register as a situationally defined variety of the language (Biber et al., 1999. p.15). It is necessary to point out that this perspective is different from other perspectives on text types used for text analysis and classification. In addition, the register-based perspective has been used by numerous corpus-based studies to categorize texts.

recurrent expressions called lexical bundles (Biber et al., 1999) which will be defined in the following section.

For many years throughout the literature, groups of words that frequently occur together in a language have been studied and described under different labels such as; *recurrent word combinations* (Altenberg, 1998; De Cock, 1998), *n-grams* (Banerjee & Pedersen, 2003), *lexical bundles* (Biber & Conrad, 1999; Biber, Johansson, Leech, Conrad, & Finegan, 1999; Stubbs, 2007a, 2007b), *prefabricated patterns* (Granger, 1998), *formulas* (Granger and Meunier 2008; Sinclair 1991; Wray 2002), *clusters* (Hyland, 2008a; Schmitt, Grandage & Adolphs, 2004), *phrasal lexemes* (Moon, 1998), *prefabs or lexical phrases* (Nattinger & DeCarrico, 1992), *sentence stems* (Pawley & Syder, 1983), *formulaic sequences* (Schmitt & Carter, 2004), among others. These studies focused on different types of word combinations and used different research methods. The present study will focus on a particular type of word combinations called *lexical bundles* which were first defined in the Longman Grammar of Spoken and Written English (Biber et al., 1999). Lexical bundles are fixed group of words that occur together in a language and are commonly used in particular registers, that is in different situationally defined varieties of the language. As stated by Biber et al. (1999) lexical bundles are “recurrent expressions, regardless of their idiomaticity, and regardless of their structural status” (p. 990). In order for a word combination to count as a bundle, it has to meet a set of defining criteria as explained by Biber (1996). First, since frequency is the defining characteristics of the lexical bundles, these expressions must occur frequently in a register. They are simply the most frequently occurring sequences of words in a sub-corpus of texts from a single register. The frequency cut-off point may vary from study to study. Biber et al. (1999) concluded that to be a lexical bundle, a four-word expression had to recur ten times per million words and had to

appear in more than five texts. On the other hand, the criterion for Biber, Conrad, & Cortes (2004) was that a lexical bundle had to occur forty times in a one-million word corpus; whereas, Cortes (2004) decided to set the cut-off point at twenty times in one million words. These higher cut-off points were chosen to be more conservative in the frequency of these expressions and to ensure that the object of analysis in these studies consisted of unit expressions that were used in extremely high frequencies. Second, in addition to frequency, lexical bundles must be used in at least five different texts. This prevents focusing on idiosyncratic uses by the authors of the texts in the corpus under consideration. Third, it should be noted that lexical bundles are not idiomatic in meaning. Although a lexical bundle functions as a whole unit, unlike idioms, its meaning could be clearly understood from the words that construct the bundle. Finally, lexical bundles do not represent complete structural units. In fact, Biber et al. (1999) found that in academic writing more than 95% of the lexical bundles were not complete units. The argument is further supported by Cortes (2004): “Lexical bundles are identified empirically, rather than intuitively, as word combinations that recur most commonly in a register, and therefore, lexical bundles are usually not complete structural units, but rather fragmented phrases or clauses with new fragments embedded” (p. 400).

In the study of lexical bundles, computer software and corpus tools have been essential for researchers to complete these studies where the purpose is to reach empirical conclusions and to analyze the collected data. The present study also utilizes computer software in order to conduct the study. The concordance program AntConc, which is used in this study, will be introduced in detail in Chapter 3.

Lexical bundles have attracted attention in language studies. Many corpus-based studies were conducted looking at frequencies of lexical bundles or comparing lexical bundles in

different registers, in different contexts, or in the products of writers with different proficiency levels (novice vs. experienced authors). Among many other results of these studies, it is found that lexical bundles can be easily related to various discourse functions. In the next section, some prominent studies on lexical bundles will be presented.

2.3 Lexical Bundles and Register Variations

Over the last few decades, there has been a sharp shift in the study of formulaic expressions toward to study of recurrent expressions identified empirically and frequency-based. An increasing number of studies on lexical bundles have been conducted. Most of these studies have reported results on the distribution and use of lexical bundles in English. These studies have had various purposes and looked at different registers. While some of these studies investigated the lexical bundles in spoken vs. written registers, others looked at academic vs. non-academic registers. In addition, there are studies that investigate lexical bundles in languages other than English or comparing two languages (English vs. Spanish). Examining lexical bundles for pedagogical purposes has also been the focus of a few studies on lexical bundles.

Table 2.1 below provides an overview of these previous corpus-based studies on lexical bundles with different corpora, and research focus and purposes which were conducted in the past decades. Further explanation of the purposes, findings and results for each of these studies will be provided in the following paragraphs.

Table 2.1 Major studies on Lexical bundles

Author	Year	Corpus	# Corpus Size
Biber, Johansson, Leech, Conrad, & Finegan	1999	LSWE Corpus	Over 40,000,000
Cortes	2002	Native freshmen compositions (311 papers)	360,704
Cortes	2004	Published writings and student writings	Published writings: 1,992,531; Student writings: 904,376
Biber, Conrad, & Cortes	2004	T2K-SWAL Corpus	2,009,400
Scott & Tribble	2006	MA dissertations (POZ_LIT) and BNC World English Edition	POZ_LIT: 352,258 BNC: 1,500,000
Nesi & Basturkmen	2006	BASE corpus and MICASE	1,270,798
Biber & Barbieri	2007	T2K-SWAL and LSWE	T2K-SWAL: 2,541,795 LSWE Academic: 5,330,000
Cortes	2008	Published history writing in English and Spanish	English: 1,001,012 Spanish: 1,003,264
Hyland	2008a	Research articles, doctoral dissertations and master's theses	3,400,400
Hyland	2008b	Research articles, doctoral dissertations and master's theses	3,500,000
Kim	2009	Korean Lexical Bundles in Conversation and Academic Texts	The Sejong Corpus: Conv.: 2,604,054 Acad.: 3,407,020

Table 2.1 shows a list of some corpus-based studies on lexical bundles. The results of these studies emphasize the importance of these linguistic features in different registers, contexts and languages. The first study shown on the table is by Biber et al. (1999) which was based on a large corpus of both American and British English conversation and academic prose. Biber et al. (1999) coined the term lexical bundles for "...word forms often co-occur in longer sequences, called *lexical bundles*" (p.989). In the same chapter, it is stated that "both conversation and academic prose use a large stock of different lexical bundles" (p.993). This claim has become a springboard for further studies on lexical bundles in different registers Biber et al. (2004) conducted another extensive study by looking at the use of lexical bundles in university classroom teaching and textbooks in comparison with the LSWE corpus previously mentioned. They discovered that the lexical bundles in their corpora differ dramatically from other linguistic features, and that university lectures use twice as many lexical bundles than conversation and four times as many lexical bundles as textbooks. The structural and functional taxonomies structured in these two studies (Biber et al. 1999, 2004) will also be used in the present study and will be described in detail in the methodology chapter.

In addition, using the same corpus, the T2K-SWAL, Biber and Barbieri (2007) looked at the use of lexical bundles in non-academic university registers and core instructional registers. In contrast with previous studies which showed that lexical bundles were more common in speech than in writing, they found that lexical bundles were very common in instructional written course texts such as course syllabi.

Cortes (2002) analyzed freshman compositions in terms of lexical bundle use. After collecting 311 student writings and using a specially-designed computer program, she found 93 different lexical bundles. Further analysis, however, showed that in terms of structure these

lexical bundles looked like the lexical bundles used in academic prose while functionally these expressions served as temporal or locative markers which created redundancy in students' writings. This study showed that lexical bundles should be analyzed elaborately both structurally and functionally and further studies should be done in students' written production at different levels and in different disciplines. Following this argument in her next study, Cortes (2004) compared the written productions of university students who were native speakers of English with published journal articles. Her corpus of over 2 million words consisted of two main disciplines; history and biology. This study revealed that students rarely used the lexical bundles identified in the corpus of published writing. Similarly, Scott and Tribble (2006) also looked at student writings and professional writings and concluded that apprentice writers used less varied and less sophisticated lexical bundles.

Going beyond studies that focused on English, four years later Cortes (2008) published another study aimed at comparing published history articles in English and in Spanish. After collecting history articles from journals both in American English and Argentinean Spanish, Cortes compared the lexical bundles identified in those corpora and analyzed them in terms of both structure and function. It was clear that even though the number of lexical bundles found was different, there was a certain degree of agreement in the expressions identified in each language. Another recent study exploring another language than English has been published by Kim (2009). Investigating a large corpus of Korean texts consisting of academic prose and conversation, she found that lexical bundles are important expressions in Korean with the function as discourse frames for new information.

As also shown in table 2.1, Nesi and Basturkmen (2006) used 160 monologic lectures from the BASE corpus and MICASE. This study focused on the function of lexical bundles in

academic lectures and revealed that lexical bundles can play a discourse signaling role in lectures and it is important to raise students' awareness of this use of lexical bundles.

The two other corpus-based studies on lexical bundles that deserve to be mentioned here are by Hyland (2008a, 2008b) who has done many studies on the analysis of various linguistic features frequently found in academic discourse. In these two studies based on findings from two corpora of research articles, doctoral dissertations and master's theses, Hyland emphasized that postgraduate students tended to employ more formulaic expressions than native academics and there was disciplinary variation in the use of lexical bundles.

In addition to comparing registers or novice or experienced writers, there are also a few studies that focused on a more pedagogical aspect of lexical bundles (Cortes 2006, Neely & Cortes, 2009). Cortes (2006) reported the results of a study in which she explicitly taught lexical bundles to students in a writing intensive history class. After analyzing the effectiveness of the tasks she prepared for teaching lexical bundles by comparing students' writings, she concluded that students' use of target bundles was rare and uneven and having a few lessons that demonstrate some examples of lexical bundles in professional writing might not necessarily result in students using more lexical bundles in a more appropriate way. However, she also emphasized that this explicit teaching of lexical bundles might increase awareness of these expressions and might lead to more academically appropriate written productions.

It should be noted that most studies on lexical bundles focused on the production of the native speakers of a language, English or other language. So far, little is known about the lexical bundles used by non-native speakers of a language in their academic written production.

In this chapter some corpus-based studies on lexical bundles have been reviewed in detail. It is clear that the results obtained from these corpus-based studies reveal a lot of valuable

information about the significance of lexical bundles and how they differ both structurally and functionally in different academic registers and in different contexts. Additionally, they provide opportunities to explore lexical bundles in further studies, which was the impetus for the present investigation. In the light of these and other studies on lexical bundles, the following chapter will introduce the data collected for this study and the methodology used in this study.

CHAPTER 3. METHODOLOGY

This chapter describes the steps followed to conduct this study. First, the collection of the corpus created for the purpose of this study (a corpus of published articles written in English by Turkish scholars) will be introduced. In the second section, the concordancing program used to facilitate the search for lexical bundles in the corpus will be described, and in the last section the taxonomies used for structural and functional analysis of the identified lexical bundles will be discussed in detail.

3.1. The TSRA Corpus

In one of her works, Conrad (1996) begins describing the corpus for her study by saying that “In a corpus-based study, the design of the corpus is very important because the corpus must be suitable to the research questions being addressed” (p. 303). Since this study focuses on finding the lexical bundles used by Turkish scholars in their research articles written in English, the corpus needed to be carefully compiled to serve this purpose. Only research articles were included in the corpus because it is believed that including more than one type of academic prose could affect the results of the study as lexical bundles are register-bound. Therefore, instead of including a limited number of theses or dissertations from a limited number of researchers or including different types of academic texts, only research articles from different authors have been compiled which contributed to the reliability of the study. Using the library online database at Georgia State University Library, articles written between 1990 and 2010 by Turkish authors in six different disciplines were collected from various professional journals (see Appendix A for a complete list of journals). Table 3.1 presents more information on the disciplines included and the number of words for each discipline in the corpus. The articles collected for the Turkish

Scholars Research Articles Corpus used in this study, which hereafter will be referred to as TSRAC, were individually checked to ensure that the article was from a journal published in an English speaking country and was released within the time period previously established (1990-2010). It was also ensured that the nationality of the authors was Turkish and they were in Turkey while writing these articles. In addition, articles which had native speakers of English as co-authors were not included in the corpus collection. After all the electronic copies of the articles were collected, the process of erasing non-textual annotations such as the titles, page numbers, tables, statistical graphics, numerical data, formulations, and references was completed.

In terms of the size of the corpus for this study, the principle suggested by Biber (2006) was followed. According to Biber (2006) “A corpus must be large enough to adequately represent the occurrence of the features being studied”. He goes on explaining why corpus size matters by emphasizing that it depends on the purpose of the study. For example, if the target feature is a frequent grammatical structure such as nouns or verbs, the size of the corpus can be smaller because these features occur frequently. However, if less common features are the target of the study, then it is essential to work with a larger corpus. In this study; therefore, a one-million word corpus was required. It should also be noted that it is ensured that the number of words in each section of the corpus from different academic fields is almost equal.

Table 3.1 shows some information on corpus size and the disciplines the research articles selected for the corpus collection belong to. When the corpus reached 1,000,000 words and was ready to be further analyzed, the computer software AntConc was used.

Table 3.1 Disciplines in the TSRAC

Disciplines	# of Words	# of Articles
Economics	164,745	29
Education	167,541	32
History	169,299	20
Medicine	153,715	44
Psychology	164,358	50
Sociology	185,479	25
Total	1,005,137	200

3.2 Concordancing Software: AntConc

This present study aims to find the most common lexical bundles in TSRAC. It has been noted that different studies have set different criteria for the identification of lexical bundles, such as number of words within each bundle and the frequency and range cut-off points. In this study the criteria followed in establishing the cut-off points agrees with that by Cortes (2008) “a four-word combination has to occur twenty times in one million words, and has to appear in five or more texts” (p.46) to be considered a lexical bundle. The reason to focus on four-word lexical bundles is that, as Cortes (2004) observes, “many four-word bundles hold three-word bundles in their structures” (p. 401) and four-word bundles are, in many cases, much more frequent than five-word bundles. As also stated by Hyland (2008b), four-word lexical bundles are more common and present a wider range of structures and functions.

With the increase of corpus-based research studies in the field of Applied Linguistics and language teaching, new tools used to analyze language corpora have been developed. For the purpose of this study, AntConc, a useful text analysis tool created by Laurence Anthony (2007), was used. The reason why this software was chosen is because along with other features, it has word and keyword frequency generators, and tools for cluster and N-grams analysis. Particularly

in terms of lexical bundles, AntConc can be considered an efficient tool to identify word combinations after meeting the previously-established cut-off points for frequency. However, it does not allow range which had to be processed manually as explained in detail below. The procedure of finding lexical bundles began with first clearing the articles from non-textual content such as graphics, formulas, page numbers, references, tables, figures etc. Since AntConc requires plain text, all the articles are saved as plain texts before being uploaded to AntConc. Second, for retrieving lexical bundles from those integrated files, frequency counts of 4-grams using the “N-Grams” command in AntConc (Anthony, 2007) were conducted. This function performs a full extract of any n-grams from the whole corpus once “n” is specified. In addition, using the minimum n-gram frequency of AntConc, it is ensured that the expression found appears at least twenty times in the corpus. After running AntConc based on these settings, a list of four-word expressions is retrieved and the cut-off point for range had to be calculated manually. The way in which the file information is presented by the software makes it easy to manually count the number of texts in which an expression occurs in order for that expression to meet the cut-off point for range and be considered a lexical bundle. As the next step, each expression in the list had to be manually checked to find whether or not it appears in more than five texts in the corpus. Expressions that appeared in less than five texts are not considered to be lexical bundles and were, therefore, eliminated.

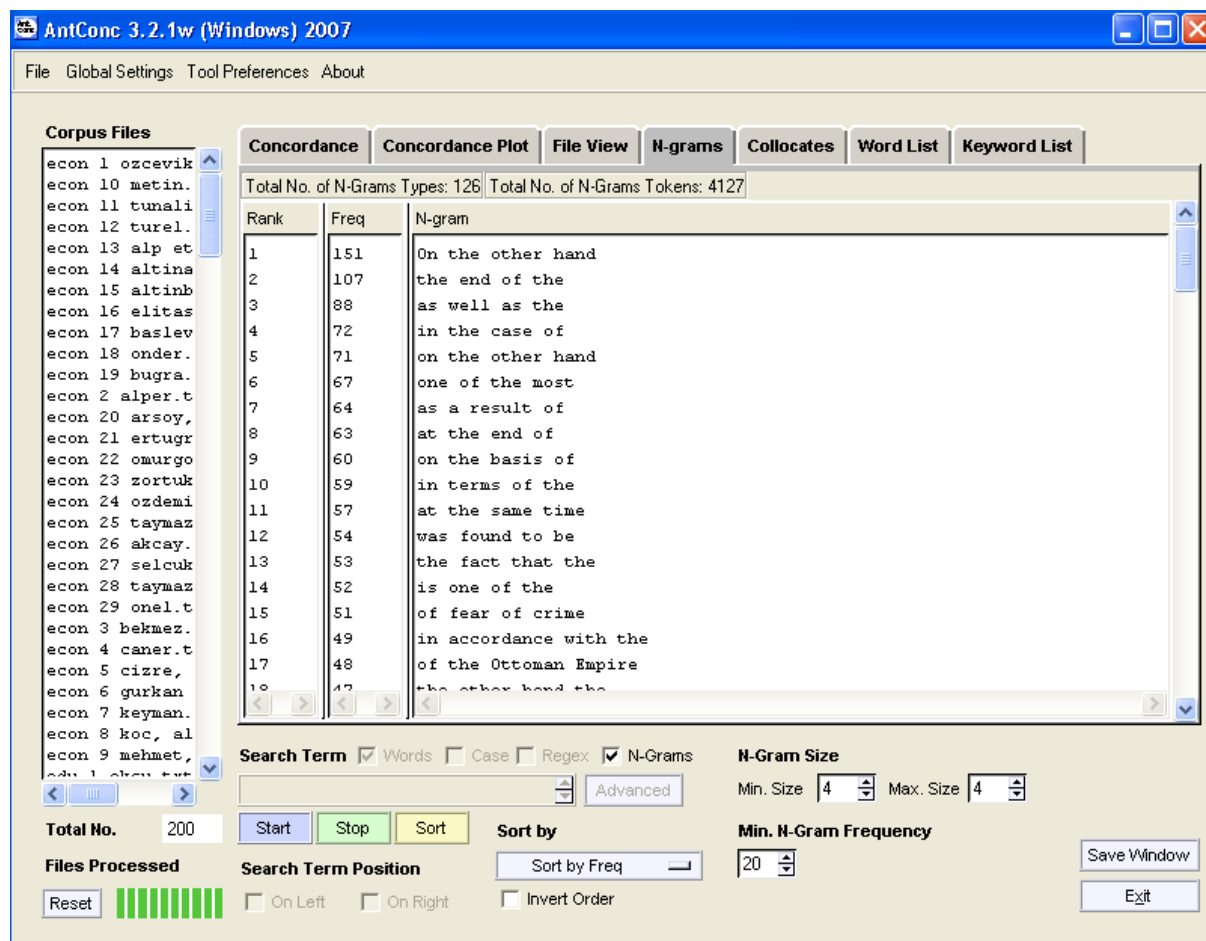


Figure 1. AntConc screenshot showing the TSRAC bundles (Anthony, 2007)

AntConc not only helped with the quantitative part of this study, providing a frequency list as shown in *Figure 1*, but also provided the information required for the qualitative interpretation of the results which is one of the aims of the study: the description of structural and functional types of lexical bundles identified in the TSRAC. In previous studies on lexical bundles, it was clearly stated that lexical bundles show variety in terms of their grammatical structures and their functionalities (Biber et al. 1999, 2003, 2004). Therefore, each bundle was analyzed elaborately in its context in order to reach a conclusion about the functional type of a bundle. As the last

step, the concordancing tool of AntConc was used to get a clear viewing of the sentences in all the texts in which the bundle occurred, which is also shown in *Figure 2*.

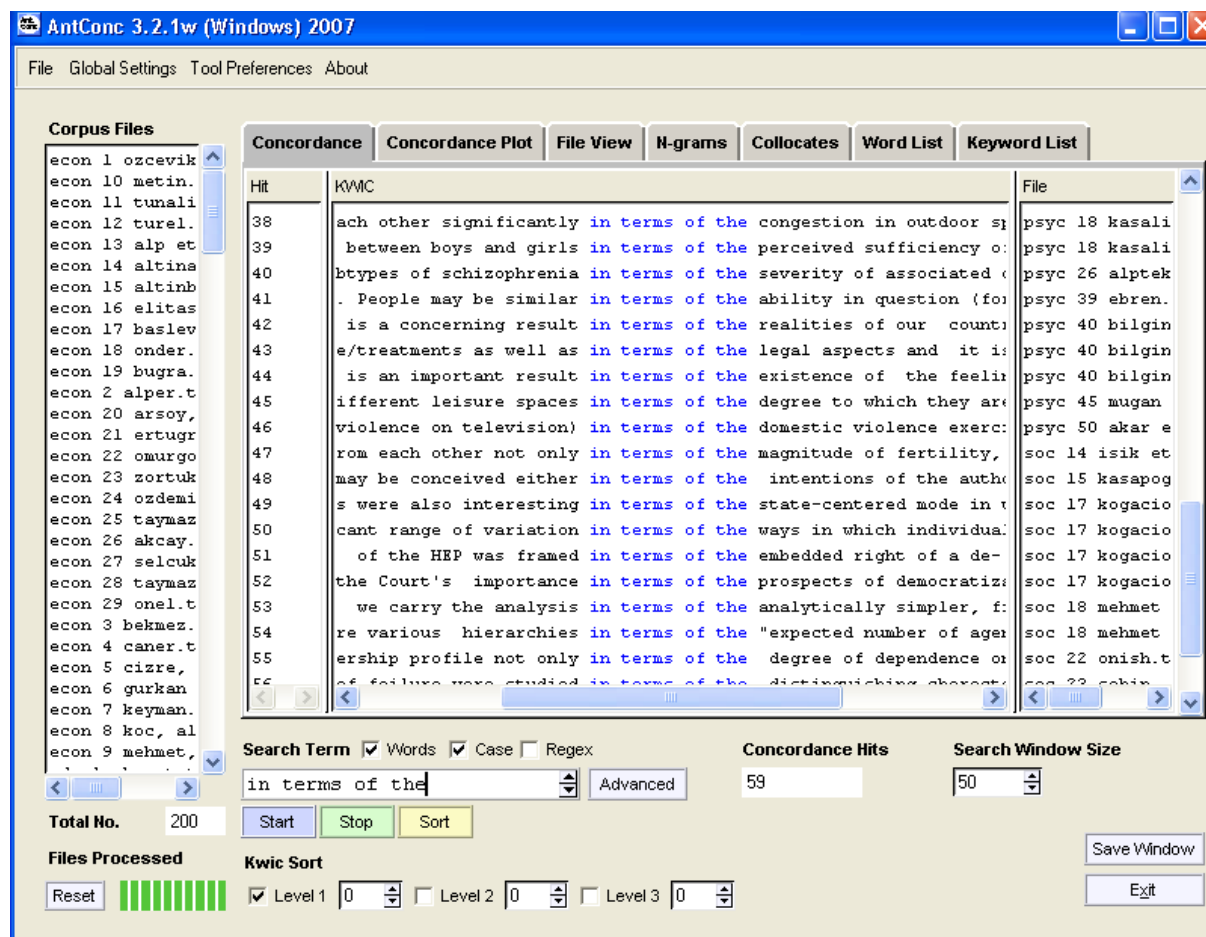


Figure 2. AntConc screenshot showing the concordances (Anthony, 2007)

Based on these analyses, lexical bundles that had similar grammatical structures and functions were grouped together using the structural and functional taxonomies according to their use and meaning in context. It should be noted that in order to reach a complete and more reliable conclusion, a second rater helped with the identification and classification of the lexical bundles found.

3.3 Structural and Functional Taxonomies

The structural classification of lexical bundles in the Longman Grammar of Spoken and Written English (Biber et al., 1999) has been widely relied on in the studies on lexical bundles in the field (Cortes, 2002, 2004; Hyland, 2008a, 2008b). A revised version of this classification was used for the purpose of this study (see Table 3.2). According to this taxonomy, lexical bundles were divided into 12 major structural categories which can be seen in Table 3.2. However, for the purpose of this study, a slight change has been applied to this model by placing these classification into two broader categories; phrasal and clausal. For the phrasal bundles, three subcategories were distinguished: “Noun-Phrase (NP) based,” “Preposition Phrase (PP) based,” and “Verb Phrase (VP) based.” NP-based bundles include any noun phrases with post-modifier fragments, such as *the role of the* or *the way in which*; PP-based bundles refer to bundles starting with a preposition plus a noun-phrase fragment or another prepositional phrase fragment, such as *at the end of* or *in relation to the*. Lastly, VP-based bundles are those with any word combination with a verb component, such as *in order to make* or *was one of the*. Clausal lexical bundles, on the other hand, can be a verb or adjective followed by a to-clause fragment as in the example of *is likely to be*, or a verb phrase followed by a that-clause fragment such as *should be noted that*. Lexical clauses that incorporate that-clause (*can be seen that*), to-clause (*are more likely to*), or adverbial clause (*if there is a*) are categorized in one broad group as clausal. Although Biber et al. (1999) does not classify the lexical bundles into phrasal and clausal in the taxonomy modeled, for the purpose of this study these two categorizations are used as seen in Table 3.2.

Table 3.2 Structural Types of Lexical Bundles (Biber et al., p.1015)

Category	Example
A. Phrasal	
1. NP-based	
(connector +) NP with of- phrase fragment	<i>the end of the</i>
NP with other post modifier fragment	<i>the way in which</i>
2. PP-based	
PP with embedded of-phrase fragment	<i>as a result of</i>
Other Prepositional Phrase (fragment)	<i>at the same time, on the other hand</i>
3. VP-based	
Anticipatory it + VP/adjective P + comp. cl.	<i>it is possible to</i>
Passive verb +PPf	<i>is based on the</i>
Copula be + noun phrase/adjective phrase	<i>is one of the, is due to the</i>
Pronoun/NP + be	<i>this is not the, there are a number of</i>
B. Clausal	
(verb/adjective +) to-clause fragment	<i>is likely to be, to be able to</i>
(VP +) that-clause fragment	<i>should be noted that</i>
Adverbial clause fragment	<i>as shown in figure, if there is a</i>
C. Other Expressions	<i>as well as the</i>

With regard to the functional categorization of the lexical bundles in this study, the taxonomy designed by Cortes (2002) and improved by Biber and his colleagues (Biber et al., 2003, 2004 and 2007) was used. In this taxonomy three major categories were distinguished: “stance bundles”, “discourse organizers,” and “referential expressions” (see Table 3.3).

Stance Bundles are groups of words that reveal the writer’s attitude, judgment, perspective in terms of certainty or uncertainty, and proposition or ability as in *it is important to*, *to come up with*, or *the fact that the*. On the other hand, as their name suggests, “*discourse organizers*” help to compose and structure the text itself. They have various functions such as introducing a topic, clarifying or elaborating on the topic (e.g., *a little bit about*, *as well as the*). Finally, “*referential expressions*”, which are very frequent in academic texts, are those that relate to a given attribute, a condition or refer to number, amount, size or quantity. Furthermore, expressions which reveal information about time and place are also included in this broad category. The bundles that can express different referential functions in different contexts are categorized as multi-functional referential expressions. For example the bundle *at the end of* can both refer to place and time as seen in the example, “*at the end of this paper*” or “*at the end of the 19th century*”.

Table 3.3 Functional classification of lexical bundles (Biber, Conrad and Cortes, 2004 p.384)

Categories	Example
1. Stance Expressions	
A. Epistemic Stance	
Personal	<i>I think it was</i>
Impersonal	<i>are more likely to</i>
B. Attitudinal/ Modality Stance	
B.1) Desire	<i>if you want to</i>
B.2) Obligation/ Directive	
Personal	<i>you look at the</i>
Impersonal	<i>it is necessary to</i>
B.3) Intention/Prediction	
Personal	<i>what we are going to</i>
Impersonal	<i>is going to be</i>
B.4) Ability	
Personal	<i>to be able to</i>
Impersonal	<i>it is possible to</i>
2. Discourse Organizers	
A. Topic Introduction/Focus	<i>in this chapter we</i>
B. Topic Elaboration/ Clarification	<i>on the other hand</i>
3. Referential Expressions	
A. Identification/ Focus	<i>one of the most</i>
B. Imprecision	<i>and things like that</i>
C. Specification of Attributes	
C.1) Quantity Specification	<i>a lot of people</i>
C.2) Tangible Framing Att.	<i>in the form of</i>
C.3) Intangible Framing Att.	<i>in the case of</i>
D. Time/Place/Text Reference	
D.1) Place Reference	<i>in the United States</i>
D.2) Time Reference	<i>at the same time</i>
D.3) Text Deixis	<i>as shown in Figure N</i>
D.4) Multi-functional Ref.	<i>at the end of</i>

In this chapter, the details of how the texts were collected and the corpus was compiled, followed by a brief description of computer software used to analyze these texts were presented. Finally, the chapter introduced the two taxonomies developed by Biber et al. (1999) and Biber et al. (2004) that will be used for the structural and functional analysis of lexical bundles found in the TSRAC. Based on the data and procedures just described, the next chapter will present the lexical bundles identified in this study together with their structural and functional classifications.

CHAPTER 4: RESULTS AND DISCUSSION

This chapter introduces the lexical bundles identified in the TSRAC. In addition, the results of the quantitative and qualitative analyses will be presented as well as a discussion for these results.

4.1 TSRAC Lexical Bundles

A total of ninety-nine lexical bundles were identified in the TSRAC (see Appendix B for a complete list). 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*, all of which are also identified as frequent lexical bundles in the literature. In the Longman Grammar of Spoken and Written English, Biber et al. (1999) state that the two most common four-word lexical bundles are *in the case of* and *on the other hand*, which are also extremely frequent in the TSRAC.

Fourteen out of ninety-nine lexical bundles occurred more than fifty times per million words, which shows a highly frequent use of these recurrent expressions. The first nine of these fourteen frequently used bundles had been also identified by Biber et al. (2004), and Cortes (2004, 2008). When individually compared to the lexical bundles identified before in the literature, it was found that 53 of the total 99 lexical bundles had not been identified before.

As one of the purposes of this study is to find the structural and functional features of the lexical bundles produced in the TSRAC, in the next two sections the structural and functional analyses will be presented. These classifications will be followed by a detailed analysis of those bundles that were exclusively found in the corpus collected for the present study, the TSRAC.

4.2 Structural Analysis of TSRAC Lexical Bundles

First of all, in parallel with what Biber et al. (1999) and Cortes (2004) argued in their studies, the lexical bundles found in the TSRAC are not grammatically complete units as shown by expressions from the TSRAC such as *one of the most*, *the end of the*, *this study was to*, *to the results of*, etc. Even though lexical bundles are not complete units, they can be grouped according to their structural characteristics. Overall, there are two broad types of lexical bundles, phrasal and clausal. Phrasal lexical bundles are divided into sub-categories as noun phrase-based, prepositional phrase-based and verb phrase-based. Clausal bundles, on the other hand, are formed for example by a that-clause fragment and a verb followed by a to-clause fragment. The third group in addition to phrasal and clausal fragments is called other expressions which is further explained by Biber et al. (1999) as “lexical bundles that do not fit neatly into any of the other categories” (p.1024). As shown in *Figure 3*, the largest part of the lexical bundles is comprised of prepositional phrases (PP). The forty-eight lexical bundles in this group are made up of prepositional phrases followed by thirty-three lexical phrases made up of noun phrases. Examples of these prepositional phrases are: *in the context of*, *at the time of*, *in this study the*, *in line with the*, *in terms of their* etc. Lexical bundles that are formed by noun phrases (NP) are expressions such as *aim of this study*, *results of this study*, *an increase in the*, *the second half of*, and others. Verb phrase (VP)-based bundles are relatively rare and examples are expressions such as *it was found that*, *it is necessary to*, *participate in the study*, *it is possible to* etc. There are only two lexical bundles that have clausal fragments (CF) in this corpus and they are *that there is a*, and *to be able to*. Finally, lexical bundles called other expressions, as explained above, are; *as well as the*, *as well as in* and *than half of the*.

The lexical bundles that had not been identified before show structural varieties. There are bundles from each of these five groups except CF; PP (*in accordance with the, according to the results, of the most important, in line with the*); NP (*a result of the, the role of the, the purpose of this, the second half of, the establishment of the*); VP (*it was determined that, to participate in the, were included in the, participate in the study*); and other expressions (*than half of the, as well as in*).

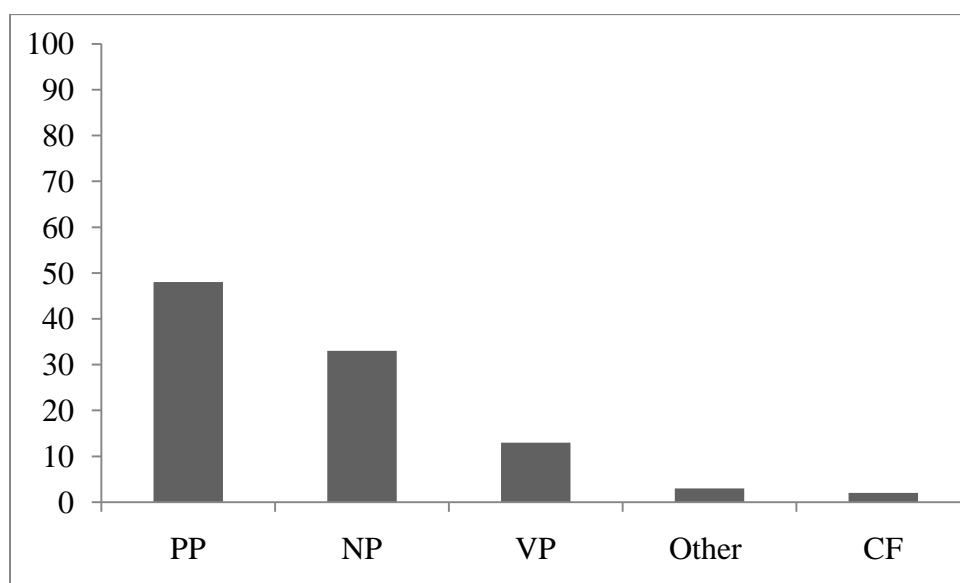


Figure 3. Structural Distribution of TSRAC Lexical Bundles

4.3 Functional Analysis of TSRAC Lexical Bundles

As explained in Chapter 3, the taxonomy used for the functional analysis of lexical bundles was developed by Biber et al. (2004) and included three broad categorizations; stance expressions, discourse organizers and referential expressions with various subcategories for each. It was found that overall, lexical bundles used by Turkish scholars perform functions similar to those performed by bundles previously identified in the literature. In addition to the categories

from Biber et al. (2004), the sub-category of referential expressions called *institute bundle* (Cortes, 2008) had to be added to classify some bundles for the TSRAC. In addition a group of bundles identified in the TSRAC which had not been identified before were performing a function that did not match any of the functions in the existing taxonomy used for the classification. Thus, a new category labeled as *research referential* had to be created within referential expressions to classify these expressions. Table 4.1 presents the functional classification of all the four-word lexical bundles identified in the TSRAC.

Table 4.1 Lexical Bundles in TSRA Corpus according to their functions in context

Category	Sub-category	Bundles
Stance	a) Epistemic Stance	
	Personal	
	Impersonal	<i>the fact that the, to the fact that*, of the fact that*</i>
	b) Attitudinal/Modality Stance	
	Desire	
	Obligation/Directive	
	Personal	
	Impersonal	<i>of the most important*, it is necessary to, the importance of the*</i>
	Intention/ Prediction	
	Personal	<i>are more likely to</i>
Impersonal		
Ability		
Personal	<i>to be able to, it is possible to</i>	
Impersonal		
Discourse Organizers	a) Topic Introduction/Focus	<i>in the present study, that there is a, with respect to the</i>
	b) Topic Elaboration/ Clarification	<i>on the other hand, as well as the, in accordance</i>

Table 4.1 Lexical Bundles in TSRA Corpus according to their functions in context (cont'd)

Category	Sub-category	Bundles
Discourse Organizers		
	b) Topic Elaboration/ Clarification	<i>with the*, it was determined that*, it was found that*, on the one hand, that there was a*, with the help of* as well as in*, were found to be, was found to be, in addition to the</i>
Referential Expressions		
	a) Identification Focus	<i>one of the most, is one of the</i>
	b) Imprecision	
	c) Specification of Attributes	
	Quantity Specification	<i>the majority of the, the rest of the, the total number of, for the first time, than half of the*, the second half of*</i>
	Tangible Framing Attr.	<i>on the part of, in line with the*, the size of the</i>
	Intangible Framing Attr.	<i>in the case of, as a result of, on the basis of, in terms of the, a result of the*, the beginning of the, in the context of, the basis of the*, an important role in, the case of the*, in terms of their*, the nature of the, the course of the, in the form of, an increase in the, Turkish version of the, the ways in which, in the number of, the establishment of the*, at the level of*, in the face of, in the field of*, the characteristics of the*, the relationship between the, the role of the*</i>
	d) Time/Place/Text Reference	
	Time Reference	<i>at the same time, at the time of, in the early #s*, the #s and #s, in the #s and*, in the late #s, during the course of*, at the end of</i>
	Place/ Event Reference	<i>in the Ottoman Empire*, in the city of*</i>

Table 4.1 Lexical Bundles in TSRA Corpus according to their functions in context (cont'd)

Category	Sub-category	Bundles
Referential Expressions	Text Deixis	<i>in accordance with the*</i> , <i>of this study was*</i> , <i>this study was to*</i> , <i>according to the results*</i> , <i>are presented in Table*</i> , <i>of the present study*</i>
	Institution Reference	<i>the Ministry of Education*</i> , <i>of the Ministry of*the Turkish Republic*</i> , <i>Ministry of National Education*</i> , <i>by the Ministry of*</i> , <i>at the university of</i>
	Multi-Func. Reference	<i>the end of the, of the Ottoman Empire*</i> , <i>at the beginning of</i>
	e) Research Reference	<i>to participate in the*</i> , <i>to the result of*</i> , <i>the results of the, the aim of this*</i> , <i>purpose of this study*</i> , <i>aim of this study*</i> , <i>the purpose of this*</i> , <i>results of this study*</i> , <i>in this study the*</i> , <i>of this study is*</i> , <i>in a study by*</i> , <i>of the patients were*</i> , <i>were included in the*</i> , <i>for the purpose of*</i> , <i>participate in the study*</i>

* is used for lexical bundles that had not been identified before in the literature

4.3.1 Stance Bundles:

According to Biber (2006), stance bundles express personal feelings, attitudes, perspective, certainty, uncertainty etc. Stance bundles can be divided into two sub-groups: epistemic stance bundles and attitudinal/modality stance bundles. Epistemic stance bundles are those expressions that reveal information about certainty (impersonal) and uncertainty (personal). The lexical bundles in the TSRAC that show impersonal epistemic stance bundles are expressions such as *of the fact that* and *to the fact that* as shown in the following example:

Although the researcher is aware *of the fact that* the universities involve various levels, this study only deals with the perceptions of the faculty members. (Edu.)

Another bias is related *to the fact that* a large part of the available evidence pertains to state intervention in the economy of the capital city, which should not be construed as evidence of conditions elsewhere in the empire. (Hist.)

The second sub-category of stance bundles is attitudinal/ modality stance with four major further sub-categories: desire, obligation/directive, intention/prediction, and ability. As the names suggest, these lexical bundles express personal attitudes. Examples of these bundles can be found in the following excerpts from the TSRAC:

Furthermore our results on the difference between single and married women clearly indicate *the importance of the* gender based division of labor in the household, indicated by the slower and weaker response of married women to the macroeconomic changes. (Econ.)

She was one *of the most important* names in mobilizing the women's vote for the party in the March 1994 local elections, which brought the party to power in major municipalities including Ankara and Istanbul. (Soc.)

By tracing how these books have been actively appropriated and filtered through the conceptual grid of prevailing controversies and ongoing events in the national arena, I hope *to be able to* say something about the changing contours of the discipline. (Soc.)

4.3.2 Discourse Organizers:

The lexical bundles in this group either introduce a topic or elaborate/clarify the topic introduced. The majority of the lexical bundles found were used for elaboration and clarification purposes as shown in the examples below:

On the other hand, inflation adjustments made after January 1, 2004 will affect the tax calculation (Pricewaterhousecoopers, 2004b). (Econ.)

Thus, *in addition to the* effects of demographical and organizational characteristics, the effects of the variations in cultural orientations were tested by using rigorous analysis techniques. (Soc.)

4.3.3 Referential Expressions:

As the last broad group, referential expressions play an important role in the identification of functions of lexical bundles. As Biber et al. (2004) state, the bundles in this category “generally identify an entity or single out some particular attribute of an entity as especially important” (p.393).

This group has four sub-categories; identification/focus, imprecision, specification of attributes, and time/place/text reference. Similar to what Cortes (2008) did in her classification, a further sub-group called institution was added to the referential category. These lexical bundles referring to institutions are expressions such as *the Ministry of Education**, *of the Ministry of**, *of the Turkish Republic**, *Ministry of National Education**, *by the Ministry of**, *at the university of*.

In the overall analysis of lexical bundles, it was found that, except for the imprecision category, every type of referential bundles occurred in the TSRAC. Moreover, when the lexical

bundles found were further analyzed a new category named “research referential” had to be added to the taxonomy used. This new classification included a large number of lexical bundles found in the TSRA corpus. In his study, Hyland (2008) introduced a research-oriented category which he explained as “helping writers to structure their activities and experiences of the real world” (p. 13). In this group, he included bundles such as *at the beginning of, the role of the, the size of the, in the present study* etc. However, when compared to the research-referential bundles in the TSRAC, his categorization was found very general and none of the bundles found in the TSRAC except for *purpose of this study* had been identified and included in Hyland’s group of research-oriented bundles. Unlike the bundles mentioned by Hyland, the bundles identified as research referential in the TSRAC refer specifically to the study itself and provide information about the purpose, procedure, results, or participants of the study as shown in Table 4.1. Furthermore, research referential lexical bundles are also different from text deixis in that lexical bundles in that text deixis refer to the paper (article or report) that presents the study and not to the investigation. However, when the lexical bundles in the research referential group are analyzed, it was found that these lexical bundles refer to more general features of the study rather than referring to the text itself. As seen in the examples *were included in the, participate in the study, to participate in the*, these research referential bundles do not refer to the text but to the study, to the actions needed to conduct the study or to describe the participants involved in the study, as shown in the following examples.

Thus, 223 teacher educators and 2,116 prospective teachers were selected from these schools in May 2005 and invited *to participate in the* study by completing the

questionnaire. Follow-up questionnaires were sent in June and July 2005 to those who did not respond to the first query. (Edu.)

Supervisors from eight different cities *were included in the* survey data. (Edu.)

In this study the respondents completed the same instrument again after four weeks. (Med.)

The aim of this study was to evaluate current use of surgical antibiotic prophylaxis in Turkish hospitals and to identify factors associated with appropriate prophylaxis. (Med.)

Although some *of the patients were* not available in the third stage, dissociative disorder NOS or dissociative identity disorder was confirmed in all of the patients who were admitted for an evaluation by the study clinician. (Psyc.)

It should also be noted that with the only exception of the lexical bundle *the results of the*, none of these fifteen research- referential lexical bundles had been identified before in the literature.

To sum up, Turkish authors used lexical bundles frequently. While some of these lexical bundles had been previously identified in the literature, more than half of them had not been identified as frequent lexical bundles in the literature. Even though Turkish scholars used lexical bundles that were not frequently used by native speakers of English in their written productions, their writing was successful because the articles used in the TSRAC were all published articles from well-known journals in each of the disciplines included in the present study. This could be

an indicator of stylistic variation in the use of lexical bundles between native and non-native speakers of English writing for scholarly publication. It can be concluded that there is variation in the use of lexical bundles between this specific group of non-native speakers of English and native speakers of English in academic setting.

CHAPTER 5. CONCLUSION

The main purpose of this study was to explore the use of four-word lexical bundles in the research articles written by Turkish scholars in English. After the compilation of the corpus, the goal was to further analyze the lexical bundles found in comparison with the bundles previously identified in the related literature. Both structural and functional analyses were completed in order to highlight any similarities and differences. This chapter will present the summary of the results by answering the research questions previously posed, discuss the limitations of the study, and provide implications and suggestions for further study.

5.1 Summary of Results

The first research question posed referred to the most common four-word lexical bundles found in the published research articles written by Turkish scholars. According to the findings of the frequency analysis, it was found that overall Turkish scholars used ninety-nine frequent lexical bundles in research articles. (See appendix B for the complete list).

The second research question asked how many of the lexical bundles in TSRAC agree with those bundles identified by Biber et al. (2004) and Cortes (2004, 2008). First of all, it was found that more than half of the lexical bundles found in this study had not been identified before in the related literature. It was recorded that the most frequent lexical bundles that were used more than fifty times per one-million word in the TSRAC agreed with the lexical bundles in the literature. However, when the lexical bundles were compared to the lexical bundles that were found by Biber et al. (2004), some of the frequently used bundles did not occur in the TSRAC. Examples of these bundles that are not found in this study are for example, *in the absence of*, *the extent to which*, *in the presence of*, and *per cent of the*. In addition, 53 of the total 99 lexical bundles identified in the TSRAC had never been identified before in the related studies of lexical

bundles. Examples of these bundles are *in accordance with the*, *it was determined that*, and *during the course of*.

The last research question aimed to explore the structural and functional features of the lexical bundles found in this study based on the previous structural and functional taxonomies developed by Biber et al. (1999, 2004). It was found that there is a high level of agreement among the structural types of lexical bundles defined previously and those found in this study. All the lexical bundles found in this corpus fit into the structural categorizations previously defined. However, with regards to functional analysis, some modifications were needed. At the end of the functional analysis, a new group of lexical bundles that did not fit into to the previously defined groups were found and a new group called research referential bundles was created. Interestingly, these lexical bundles had not been identified before in any of the three studies by Biber et al. (2004) and Cortes (2004, 2008) with the exception of only one expression.

A possible reason for this discrepancy could be that the scholars in Turkey have been told to use expressions that emphasize the study itself while writing their research articles. Therefore, it could be beneficial to do a content analysis of academic writing classes in Turkey to see if there is a focus on fixed expressions used for research purposes in academic prose.

5.2 Limitations

The results of this study need to be treated with some caution since the TSRAC consisted of only six academic disciplines: they cannot be generalized to all the disciplines. Moreover, since the structural and functional analyses of lexical bundles were qualitatively conducted by hand, it is likely that there might be some possible inconsistencies. It is necessary to point out that some of the disciplines represented in this corpus had not been investigated before in the

study of lexical bundles (e.g. medicine, economics). This could have been the reason that originated the group of bundles that had never been identified in the literature as disciplinarity provides these frequent expressions with a high degree of specificity, making them strongly discipline bound.

5.3 Implications

From a pedagogical point of view, the findings of this study could be beneficial in designing more effective materials for academic writing purposes. Even though the use of TSRAC exclusive bundles produced successful writing that lead to publication, it is still important to raise awareness on how often and for which specific purposes lexical bundles are used in academic writing. As the findings suggest, lexical bundles constitute an important part of academic prose and this should be highlighted especially by writing teachers.

5.4 Suggestions for Further Research

This study has contributed to the existing knowledge of lexical bundles; however, further studies are needed on the use of lexical bundles especially in international settings. As a further analysis, it would be interesting to compare each bundle found in the TSRAC to the bundles identified before in the literature to see if the Turkish authors used the same lexical bundles in the same way, with the same purpose and function. Additionally, it would be beneficial to survey Turkish scholars to find out if they are aware of the use of lexical bundles and their significance in academic writing. Moreover, it would be useful to investigate the materials used in the teaching of academic writing in English to see if these frequently used lexical bundles which had been identified before but do not occur in this corpus exist in these academic writing sources for

this particular setting. The same materials should also be investigated to obtain information on the origin of these lexical bundles which are frequently used by Turkish scholars. For this reason, a corpus-based study of lexical bundles found in the academic writing books available to these scholars could be a starting point.

Finally, a study on lexical bundles used in research articles written in Turkish could be conducted to compare with the lexical bundles used in English by the same authors. This comparison could help to identify if there is L1 transfer in lexical bundle use.

REFERENCES

- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 99–122). Oxford: OUP.
- Anthony, L. (2004). AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit
- Anthony, L. (2007). Antconc 3.2.1w: Freeware corpus analysis toolkit. [on-line]. Retrieved from: <http://www.antlab.sci.waseda.ac.jp/>
- Banerjee, S. & Pedersen, T. (2003). Extended gloss overlap as a measure of semantic relatedness. In *Proc. of IJCAI-03*, pp. 805–810.
- Belcher, D. D. (2007). Seeking acceptance in an English-only research world. *Journal of Second Language Writing*, 16, 1–22.
- Biber, D. (1996). Investigating language use through corpus-based analyses of association patterns. *International Journal of Corpus Linguistics*, 1, 171-197.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, D., & Conrad, S. (1999). Lexical Bundles in Conversations and Academic Prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: studies in honour of Stig Johansson* (pp. 181–190). Amsterdam: Rodopi.
- Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: an initial taxonomy. In A. Wilson, P. Rayson & T. McEnery (Eds.), *Corpus linguistics by the Lune: a festschrift for Geoffrey Leech* (pp. 71–93). Frankfurt: Peter Lang.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371–405.

- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: Benjamin.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263–286.
- Butler, C. (1997). Repeated word combinations in spoken and written text: Some implications for functional grammar. In C. Butler, J. Connolly, R. Gatward, & M. Wismans (Eds.), *A fund of Ideas: Recent development in functional grammar* (pp. 60–77). Amsterdam: Institute for Functional Research into Language and Language Use.
- Charles, M. (2003). ‘This mystery. . .’: A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes*, 2, 313–326.
- 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.
- Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, 3, 43-57.
- Conrad, S. (1996). Investigating academic texts with corpus-based techniques: An example from biology. *Linguistics and Education*, 8, 299-326.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.

- Crompton, P. (1997). Hedging in academic writing: Some theoretical problems. *English for Specific Purposes*, 16, 271–287.
- De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics*, 3, 59–80.
- Eaton, H. (1940). *An English - French - German - Spanish Word Frequency Dictionary*. New York, NY: Dover Publications.
- Ferguson, G. (2001). If you pop over there: A corpus-based study of conditionals in medical discourse. *English for Specific Purposes*, 20, 61–82.
- Flowerdew, J. (Ed.). (2002). *Academic Discourse*. New York, NY: Longman.
- Fries, C. & Traver, A. (1940). *English word lists: a study of their adaptability and instruction*. Washington, DC: American Council of Education.
- Ghadessy, M. (1995). Thematic development and its relationship to registers and genres. In M. Ghadessy (Ed.), *Thematic development in English texts* (pp. 105–128). London: Pinter.
- Grabe, W. & Kaplan, R. B. (1997). On the writing of science and the science of writing: Hedging in science text and elsewhere. In R. Markkanen & H. Schroder (Eds.), *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts* (pp.151–167). Berlin: Walter de Gruyter & Co.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145–160). Oxford: Oxford University Press.
- Granger, S., & Meunier, F. (Eds.). (2008). *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins.

- Halliday, M. A. K. (1993a). The construction of knowledge and value in the grammar of scientific discourse: Charles Darwin's *The origin of the species*. In M. A. K. Halliday, & J. R. Martin (Eds.), *Writing science. Literacy and discursive power* (pp. 86–107). London: The Falmer Press.
- Halliday, M. A. K. (1993b). On the language of physical science. In M. A. K. Halliday, & J. R. Martin (Eds.), *Writing science. Literacy and discursive power* (pp. 54–68). London: The Falmer Press.
- Hewings, M. (Ed.). (2001). *Academic Writing in Context: Implications and Applications*. Birmingham: The University of Birmingham Press.
- Holmes, J. (1986). Doubt and certainty in ESL textbooks. *Applied Linguistics*, 9, 21–43.
- Hunston, S. (1995). A corpus study of some English verbs of attribution. *Functions of Language*, 2, 133–158.
- Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes*, 13, 239–256.
- Hyland, K. (1996a). Talking to the academy: Forms of hedging in science research articles. *Written Communication*, 13, 251–281.
- Hyland, K. (1996b). Writing without conviction? Hedging in science research articles. *Applied Linguistics*, 17, 433–454.
- Hyland, K. (1998). *Hedging in Scientific Research Articles*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Hyland, K. (2003). *Second language writing*. Cambridge: Cambridge University Press.
- Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4–21.

- Hyland, K. (2008b). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18, 41-62.
- Kading, J. (1879). *Häufigkeitswörterbuch der deutschen Sprache*. Steglitz: privately published.
- Kim, Y. (2009). Korean lexical bundles in conversation and academic texts. *Corpora*, 4, 135-165.
- Martinez, A. I. (2003). Aspects of theme in the method and discussion sections of biology journal articles in English. *Journal of English for Academic Purposes*, 2, 103-123.
- McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh Textbooks in Applied Linguistics.
- Myers, G. (1989). The pragmatics of politeness in scientific articles. *Applied Linguistics*, 10, 1-35.
- Myers, G. (1990). *Writing biology: Texts in the Social Construction of Scientific Knowledge*. Madison, WI: University of Wisconsin Press.
- Meyer, P. G. (1997). Hedging strategies in written academic discourse: Strengthening the argument by weakening the claim. In R. Markkanen & H. Schroder (Eds.), *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts* (pp. 21-41). Berlin: Walter de Gruyter & Co.
- Moon, R. (1998). *Fixed Expressions and Idioms in English*. Oxford: Oxford University Press.
- Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. New York, NY: Newbury House.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nattinger, J. R., & De Carrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

- Neely, E., & Cortes, V. (2009). A little bit about: analyzing and teaching lexical bundles in academic lectures. *Language Value*, 1, 17–38. Retrieved from <<http://www.e-revistas.uji.es/languagevalue>>.
- Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signaling in academic lectures. *International Journal of Corpus Linguistics*, 11, 283-304.
- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: native like selection and native like fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191-226). London: Longman.
- Preyer, W. (1889). *The Mind of a Child*. New York, NY: Appleton.
- Rica-Peromingo, J. P. (2009). The use of lexical bundles in the written production of Spanish EFL university students. *Applied Linguistics for Specialized Discourse. Conference Proceedings*. (pp 1–7). Riga: University of Latvia Publishing.
- Salager-Meyer, F. (1992). A text-type and move analysis study of verb tense and modality distribution in medical English abstracts. *English for Specific Purposes*, 11, 93-113.
- Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes*, 13, 149–170.
- Schmitt, N. & McCarthy, M. (Eds.). (1997). *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.
- Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences* (pp. 127–151). Amsterdam: Benjamins.

- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action - an introduction. In N. Schmitt (Ed.), *Formulaic sequences acquisition, processing, and use* (pp. 1- 22). Amsterdam; Philadelphia: John Benjamins Pub.
- Scott, M., & Tribble, C. (Eds.). (2006). *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam and Philadelphia: John Benjamins B.V.
- Silver, M. (2003). The stance of stance: A critical look at ways stance is expressed and modeled in academic discourse. *Journal of English for Academic Purposes*, 2, 359–374.
- Simpson, R. (2004). Stylistic features of academic speech: The role of formulaic expressions. In T. Upton and U. Connor (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp.37-64). Amsterdam: John Benjamins.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stubbs, M. (2007a). An example of frequent English phraseology: Distribution, structures and functions. In R. Facchinetti (Ed.), *Corpus Linguistics 25 years on* (pp. 89–105). Amsterdam: Radopi.
- Stubbs, M. (2007b). Quantitative data on multi-word sequences in English: The case of word ‘world’. In M. Hoey, M. Mahlberg, M. Stubbs & W. Teubert (Eds.), *Text, Discourse and Corpora: Theory and Analysis* (pp. 163–189). London: Continuum.
- Vande Kopple, W. J. (1992). Noun phrases and the style of scientific discourse. In S.P. Witte, N. Nakadate & R. D. Cherry (Eds.), *A rhetoric of doing: Essays on written discourse in honor of James L. Kinneavy* (pp. 328-348). Carbondale, IL: Southern Illinois University Press.
- Varttala, T. (2003). Hedging in scientific research articles: A cross-disciplinary study. In G. Cortese & P. Riley (Eds.), *Domain-Specific English: Textual Practices across Communities and Classrooms* (pp. 141–174). New York, NY: Peter Lang.

Wray, A. (2000). Formulaic sequences in second language teaching: principle and practice.

Applied Linguistics, 21, 463–489.

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

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.

APPENDIX A: Journals Used in the TSRAC**Economics Journals**

1. Critical Perspectives on Accounting
2. Disasters
3. Eastern European Economics
4. Economic Development and Cultural Change
5. Economic Modeling
6. Energy Economics
7. European Economic Review
8. International Journal of Urban and Regional Research
9. International Research Journal of Finance and Economics
10. Journal of Productivity Analysis
11. Journal of Asian Economics
12. Journal of Business & Economic Statistics
13. Physica
14. Public Choice
15. Review of International Political Economy
16. Russian and East European Finance and Trade
17. Small Business Economics
18. The Canadian Journal of Economics
19. Water Resources Development
20. World Development

Education Journals

1. Asia-Pacific Journal of Teacher Education
2. Education
3. Education Media International
4. Educational Studies in Mathematics
5. Educational Technology & Society
6. Educational Technology & Society
7. Environmental Education Research
8. European Journal of Education
9. Higher Education
10. International Research in Geographical and Environmental Education
11. International Review of Education
12. Internet and Higher Education
13. Journal of Adolescent & Adult Literacy
14. Journal of Documentation
15. Journal of Education for Teaching
16. Journal of Instructional Psychology
17. Models of Teacher Education
18. Religious Education
19. Review of Education
20. The Journal of Educational Research

Psychology Journals

1. Addictive Behaviors
2. Adolescence
3. Applied Developmental Psychology
4. Archives of Psychiatric Nursing
5. Child Abuse & Neglect
6. Children and Youth Services Review
7. Comprehensive Psychiatry
8. Eating Behaviors
9. European Neuropsychopharmacology
10. Issues in Mental Health Nursing
11. Journal of Applied Developmental Psychology
12. Journal of Clinical Forensic Medicine
13. Journal of Criminal Justice
14. Journal of Environmental Psychology
15. Journal of Loss and Trauma
16. Journal of Psychiatric and Mental Health Nursing
17. Journal of Psychiatric Research
18. Journal of Psychology
19. Journal of Psychosomatic Research
20. Learning and Individual Differences
21. Psychiatry and Clinical Neurosciences
22. Psychiatry Research
23. Soc. Psychiatry Epidemiology
24. Social Behavior and Personality
25. Social Science and Medicine
26. Technological Forecasting & Social Change
27. The Journal of Experimental Education
28. The Social Science Journal

Medicine Journals

1. Applied Developmental Psychology
2. Applied Nursing Research
3. Clinical Infectious Diseases
4. Culture, Health & Sexuality
5. European Journal of Epidemiology
6. European Journal of Oncology Nursing
7. Infection Control and Hospital Epidemiology
8. International Journal of Nursing Studies
9. Journal of Clinical Forensic Medicine
10. Journal of Midwifery & Women's Health
11. Journal of Professional Nursing
12. Journal of the Association of Nurses in AIDS Care
13. Nurse Education in Practice
14. Nurse Education Today
15. Pediatrics International
16. Quality of Life Research
17. Reproductive Health Matters
18. Safety Science
19. Social Science & Medicine
20. Social Indicators Research
21. Technological Forecasting & Social Change
22. The European Journal of Health Economics
23. The Journal of Infectious Diseases
24. Tobacco Control

History Journals

1. International Journal of Middle East Studies
2. Journal of Contemporary History
3. Journal of Interdisciplinary History
4. Journal of Social History
5. Journal of the Economic and Social History of the Orient
6. Law & Society Review
7. Middle Eastern Studies
8. The International History Review
9. The Journal of Economic History

Sociology Journals

1. Comparative Politics
2. Comparative Studies in Society and History
3. Contemporary Sociology
4. Ethnology
5. European Journal of Population
6. Fashion Theory
7. Feminist Studies
8. Human Studies
9. International Labor and Working-Class History
10. Journal of Black Studies
11. Journal of Law, Economics, & Organization
12. Journal of Medical Ethics
13. Law & Society Review
14. Middle East Journal
15. Middle Eastern Studies
16. Political Psychology
17. Social Indicators Research
18. Women's Studies Quarterly

APPENDIX B: TSRAC Lexical Bundles

Frequency	TSRAC Lexical Bundles
44	a result of the*
36	according to the results*
30	aim of this study*
30	an important role in
23	an increase in the
23	are more likely to
29	are presented in Table*
64	as a result of
34	as well as in*
88	as well as the
22	at the beginning of
63	at the end of
21	at the level of*
57	at the same time
31	at the time of
20	at the University of
23	by the Ministry of*
21	during the course of*
25	for the first time
20	for the purpose of*
24	in a study by*
49	in accordance with the *
28	in addition to the
27	in line with the*
59	in terms of the
27	in terms of their*
72	in the case of
25	in the city of*
33	in the context of
28	in the early s*

21 in the face of
21 in the field of*
24 in the form of
22 in the late s
22 in the number of
37 in the Ottoman Empire*
41 in the present study
23 in the #s and*
27 in this study the*
52 is one of the
24 it is necessary to
21 it is possible to
27 it was determined that*
27 it was found that*
25 Ministry of National Education*
26 of the fact that*
26 of the Ministry of*
35 of the most important*
48 of the Ottoman Empire*
21 of the patients were*
24 of the present study*
26 of the Turkish Republic*
27 of this study is*
43 of this study was*
60 on the basis of
24 on the one hand
151 on the other hand
29 on the part of
67 one of the most
20 participate in the study*
31 purpose of this study*
29 results of this study*

23	than half of the*
30	that there is a
20	that there was a*
34	the aim of this*
32	the basis of the*
34	the beginning of the
30	the case of the*
21	the characteristics of the*
26	the course of the
107	the end of the
22	the establishment of the*
53	the fact that the
20	the importance of the*
40	the majority of the
30	the Ministry of Education*
27	the nature of the
30	The purpose of this*
21	the relationship between the
37	the rest of the
35	the results of the
32	the role of the*
25	the s and s
23	the second half of*
23	the size of the
31	the total number of
23	the ways in which
40	this study was to*
26	to be able to
36	to participate in the*
27	to the fact that*
36	to the results of*
24	Turkish version of the*

54 was found to be
34 were found to be
21 were included in the*
22 with respect to the
20 with the help of*