

KARADENİZ TECHNICAL UNIVERSITY * THE INSTITUTE OF SOCIAL SCIENCES

DEPARTMENT OF WESTERN LANGUAGES AND LITERATURE

APPLIED LINGUISTICS

**PHRASEOLOGICAL AWARENESS OF TERTIARY LEVEL TURKISH EFL LEARNERS
AS EVIDENCED IN NATIVE AND NON-NATIVE CORPORA**

MASTER'S THESIS

Hakan TAŞKAYA

MAY-2019

TRABZON

KARADENİZ TECHNICAL UNIVERSITY * THE INSTITUTE OF SOCIAL SCIENCES

DEPARTMENT OF WESTERN LANGUAGES AND LITERATURE

APPLIED LINGUISTICS

**PHRASEOLOGICAL AWARENESS OF TERTIARY LEVEL TURKISH EFL LEARNERS
AS EVIDENCED IN NATIVE AND NON-NATIVE CORPORA**

MASTER'S THESIS

Hakan TAŞKAYA



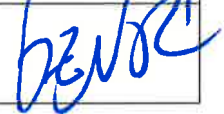
Supervisor: Assist. Prof. Dr. Ali Şükrü ÖZBAY

MAY-2019

TRABZON

APPROVAL

Upon the submission of the dissertation, **Hakan TAŞKAYA** has defended the study **Phraseological Awareness of Tertiary Level Turkish EFL Learners As Evidenced in Native and Non-Native Corpora** in partial fulfillment of the requirements for the degree of Master of Arts in English Language and Literature at Karadeniz Technical University, and the study has been found fully adequate in scope and quality as a thesis by **unanimous/majority** vote on **01/10/2019**.

Committee Member		Decision		Signature
Title - Name and SURNAME	Mission	Accept	Refuse	
Assist. Prof. Dr. Ali Şükrü ÖZBAY	Supervisor	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Prof. Dr. Kemalettin YİĞİTER	Member	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Assist. Prof. Dr. Öznur SEMİZ	Member	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

Approval of the Graduate School of Social Sciences.

Prof. Dr. Yusuf SÜRMEŒ

Director

DECLARATION OF ORIGINALITY

I, **Hakan TAŞKAYA**, hereby confirm and certify that;

- I am the sole author of this work and I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution,
- This work contains no material that has been submitted or accepted for a degree or diploma in any other university or institution,
- All data and findings in the work have not been falsified or embellished,
- This is a true copy of the work approved by my advisor and thesis committee at Karadeniz Technical University, including final revisions required by them.
- I understand that my work may be electronically checked for plagiarism by the use of plagiarism detection software and stored on a third party's server for eventual future comparison,
- I take full responsibility in case of non-compliance with the above statements in respect of this work.

Signature.....

Date: 20/05/2019

ACKNOWLEDGEMENT

Foremost, I would like to express my deepest appreciation and gratitude to my thesis supervisor Assist. Prof. Dr. Ali Şükrü ÖZBAY who paved the way for me, for all his enthusiasm, support, encouragement and patience. I always admire his endless ambition and self-sacrificing character. His never-ending energy always encouraged me do more for my studies and thesis and I owe a debt of gratitude to him. It was a honour to meet him for me.

I am sincerely grateful to Assist. Prof. Dr. Öznur SEMİZ, Assoc. Prof. Dr. M. Zeki ÇIRAKLI and Assoc. Prof. Dr. Mustafa Naci KAYAOĞLU for their great contributions to my academic life. They always supported us with their valuable comments during our lectures. Also, I present my special appreciations to Assist. Prof. Dr Öznur SEMİZ for her valuable feedbacks on my thesis.

I am also grateful to my committee member Prof. Dr. Kemalettin YİĞİTER for his valuable feedbacks on my thesis.

I would also like to thank Res. Assist. Tuncer Aydemir for his help to my thesis. His help while using software which analyzes corpora cleared uncertain points in my mind.

I am grateful to my dear wife Burcu Taşkaya for her endless encouragements and support on my thesis. Finally, I present my warmest and endless love to my daughter Duygu Nisa. Her smile always encouraged me to work enthusiastically.

May, 2019

Hakan TAŞKAYA

TABLE OF CONTENTS

ACKNOWLEDGEMENT	IV
TABLE OF CONTENTS	V
ÖZET	VII
ABSTRACT	VIII
LIST OF TABLES	IX
LIST OF GRAPHICS	X
LIST OF ABBREVIATIONS	XI
INTRODUCTION	1-7
CHAPTER ONE	8
1. LITERATURE REVIEW	8-42
1.1. Terminology in Phraseology	9
1.2. Features of Phraseology	10
1.2.1. Idiomaticity	11
1.2.2. Fixedness	12
1.2.3. Semantic Unity.....	13
1.4. Phraseology in Language Learning.....	15
1.4.1. Lexical Priming Theory	17
1.4.2. Pattern Grammar	19
1.4.3. Idiom and Open-Choice Principles	20
1.5. Phraseology in EFL Context in Turkey	20
1.5.1. EFL Learners in Turkish Curriculum.....	21
1.5.2. Problems Faced in Teaching and Learning English.....	23
1.5.3. Limitations of The Study	25
1.6. Summary	26
1.7. Literature Review of Lexical Bundles	27
1.8.1. Lexical Bundles in Academic Prose and Conversation	29
1.8.1.1. Lexical Bundles in Conversations.....	30
1.8.1.2. Lexical Bundles in Academic Prose.....	31
1.9. Classification of Lexical Bundles	33

1.9.1. Structural Classification of Lexical Bundles	33
1.9.2. Functional Classification of Lexical Bundles	33
1.9.3. Types of Lexical Bundles	36
1.9.4. Target Bundles in Current Research	36
1.10. Performance of Tertiary Level Students on Lexical Bundles	41
1.10.1. Earlier Research on Lexical Bundles	41
CHAPTER TWO	43
2. METHODOLOGY	43-51
2.1. Frequency Based Analysis	44
2.2. Corpus-Based Approaches	45
2.3. Contrastive Interlanguage Analysis	46
2.4. Target Corpora, KTUCALE and TICLE.....	47
2.5. Reference Corpora: BAWE and LOCNESS	48
2.6. Corpus Tool: Sketch Engine	50
CHAPTER THREE	52
3. FINDINGS AND DISCUSSIONS	52-74
3.1. Categorization of the Most Common Lexical Bundles in Four Corpora	53
3.2. The Most Common Lexical Bundles in KTUCALE and BAWE	53
3.3. Common Lexical Bundles in BAWE and KTUCALE.....	57
3.4. The Most Common Lexical Bundles in LOCNESS and TICLE.....	58
3.5. Common Bundles in TICLE and LOCNESS	62
3.6. The Most Common 50 Lexical Bundles in Reference Corpus BAWE.....	63
3.7. 50 Most Common Lexical Bundles in a Reference Corpus LOCNESS	67
3.8. Common Bundles in Biber's, Salazar's and Davis' Corpora and Their Values in BAWE and KTUCALE.....	69
3.9. Common Bundles in Biber's, Salazar's and Davis' Corpora and Their Values in LOCCNESS and TICLE	72
CONCLUSION AND SUGGESTIONS FOR FURTHER RESEARCH	75
REFERENCES	78
CURRICULUM VITAE	84

ÖZET

Bu çalışma, Türk öğrencilerin akademik yazılarındaki sözcük seçimlerinin kullanımına dikkat çekmek için dil ailesinin büyüyen alanı Öbekbilimine odaklanmaktadır. Kalıplaşmış dil olarak adlandırılan öbek bilim, kullanıcıların dili olması gerektiği gibi doğal olarak kullanabilmelerini sağladığı için, son zamanlarda ilgi çekmeye başlamıştır. Susan Hunston (2011: 5) Öbekbilimini “...bazı çevrelerde diğerlerinden daha çok kullanılan kelime gruplarının eğilimlerini tanımlayan genel bir tanım” olarak ifade eder. Birçok dilbilimci anlatım biçimlerinin dil ailesinde ki önemini belirtmiştir ve bu konuda birçok araştırma yapmışlardır. Öbekbilimin geçmişi 1950’li yıllara dayansa da uzun yıllar boyunca gözardı edilen anlatım biçimleri, son zamanlarda dilbilimde araştırılan temel odaklardan biri haline gelmiştir. Kelime kombinasyonları dili zenginleştirdiği için öbek bilim farkındalığının dil kullanıcıları için çok önemlidir ve bu çerçevede mevcut araştırma Türk öğrencilerinin sözcük kullanımının, hedef dildeki öğrencilerin sözcük kullanımlarıyla karşılaştırarak, nasıl kullandıklarına dikkat çekmeye çalışmaktadır. Öbekbilim alanının oldukça geniş olması sebebiyle, bu çalışmada önemli alt alanlarından biri olan sözcük kalıpları alanında çalışma yapılmıştır. Sözcük kalıpları, en az üç kelimedenden oluşan kendini tekrar eden kelime kombinasyonları olarak geçmektedir. Biber (2006: 174) ve Barbieri (2007) “sözcük öbekleri söylem oluşturma konusunda çok önemlidir” diye ifade etmişlerdir. Sözcük öbekleri etkili ve doğru bir biçimde kullanılmazsa, söylem problemleri ortaya çıkabilir. Biber (1999) sözcük öbeklerinin akademik yazıların %20’sini oluşturduğunu ifade etmiştir. Farklı dilbilimciler tarafından farklı şekilde kategori edilen 4 kelimedenden oluşan sözcük öbekleri bu araştırmanın temel odaklarından biridir. Metodoloji bölümünde iki referans derlem BAWE ve LOCNESS ve iki Türk derlemi KTUCALE ve TICLE aralarında ki benzerlikleri ve farklılıkları bulmak için analizler yapılmıştır. Referans derlemlerde ki en çok görülen sözcük öbeklerini ve Biber’in (1999) en sık görülen sözcük öbekleri listesini baz alarak karşılaştırmalar yapılmıştır. Sıklıklarını ve olasılık değerlerini hesaplayarak yabancı dil öğrenen Türk öğrencilerin sözcük öbekleri farkındalıklarını ortaya çıkaran bu araştırma, aynı zamanda da Türk öğrencilerin karakteristik seçimlerini de ortaya çıkarmaktadır.

Anahtar Kelimeler: Derlem, Öbek bilim, Sözcük Öbekleri, Yabancı Dil Olarak İngilizce

ABSTRACT

The current research focuses on phraseology as a part of the linguistic field so as to draw attention to use of phraseology in academic prose by Turkish EFL learners. Phraseology which can be also called as formulaic language has gained dramatical interest recently as it enables learners to use the language as naturally as it has to be. Susan Hunston (2011: 5) states that “Phraseology is a general term used to describe tendency of words and a group of words, to occur more frequently in some environments than others”. Many linguists underline the importance of phraseology for language learning environment and many studies are conducted in the linguistics field of the phraseology. Dating back to the 1950s, phraseology has recently been one of the main interests in the linguistics field although it was neglected many years. Languages are enhanced by word combinations and it can be emphasized that phraseological awareness of learners is crucial for a language user and current research tries to draw attention to tertiary level Turkish EFL students’ phraseological performances in academic prose by researching their writings and comparing them with native speakers’ writings. Regarding that phraseology is a broad area of linguistic inquiry, in the current study, one of the most crucial branches of the phraseology, lexical bundles, defined as at least three-word recurrent combinations mostly occur in a sentence or context are the main interest. Biber (2006: 174) states that “lexical bundles are crucial for constructing a discourse in university register”. Apart from Biber, Barbieri (2007) states that “lexical bundles are important blocks of discourse”. Four-word lexical bundles which are categorized by different linguists into different fields are the main concern in the current study. In the methodology part while carrying out the research both two native speakers’ corpora, BAWE and LOCNESS and two Turkish corpora KTUCALE and TICLE are analyzed so as to find out the how Turkish EFL learners use bundles. Based on the most common lexical bundles in reference corpora and in a corpus analyzed by Biber (1999), current study aims to see usage of lexical bundles in Turkish EFL learners’ academic writings compared to performance of native students’ performance over the lexical bundles. By analyzing the frequencies of four word more bundles with their LL scores it enables to see how effective do the Turkish EFL learners use the lexical bundles in academic prose compared to native students’ writings and what are the characteristics of Turkish EFL learners’ use of lexical bundles. The investigation of lexical bundles showed that Turkish EFL learners commonly overuse and underuse the lexical bundles comparing to native learners and their preferences of lexical bundles are not the same with preferences of native learners.

Keywords: Phraseology, Lexical Bundles, Corpus, Turkish EFL Learners

LIST OF TABLES

<u>Table Nr.</u>	Table Name	Page Nr.
1	Terms Used For ‘ Sentence-Like ’ and ‘ Word-Like ’ Combinations	10
2	Priming Hypothesis by Hoey (2005)	18
3	Solutions for Language Learning Problems	25
4	Categorization of Lexical Bundles in Conversations	31
5	Categorizations of Lexical Bundles in Academic Prose.....	32
6	Functional Classification of Lexical Bundles	34
7	Hyland’s Classification of Lexical Bundles in Academic Writing.....	35
8	Lexical Bundles Typology of Altenberg	36
9	Lexical Bundles Taxonomy in Current Research	37
10	Target Four-Word and More Bundles of Current Research	38
11	Davis’ Study of Most Common Three-Word Lexical Bundles	39
12	Salazar List of Most Common Three-Word Bundles In Academic Prose.....	40
13	Major Studies on Lexical Bundles.....	42
14	Categorization of Corpora	44
15	Study Design of Current Research	45
16	Design Criteria of KTUCALE.....	48
17	Distribution of Topics in KTUCALE	48
18	Design of BAWE Corpus	49
19	Design of LOCNESS.....	50
20	Values of Log Likelihood.....	51
21	The Most Common Lexical Bundles in KTUCALE and BAWE.....	54
22	Common Lexical Bundles In BAWE and KTUCALE.....	57
23	The Most Common Lexical Bundles In TICLE and LOCNESS.....	60
24	Common bundles in TICLE and LOCNESS	63
25	The Most Common 50 Lexical Bundles in Reference Corpus BAWE	64
26	The Most Common 50 Lexical Bundles in a Reference Corpus LOCNESS.....	67
27	Common Bundles in Biber’s Salazar’s and Davis’ Corpora	70
28	Common Bundles in Biber’s, Salazar’s and Davis’ Corpora and Their Values in LOCNESS and TICLE	72

LIST OF GRAPHICS

Graphic Nr.	Graphic Name	Page Nr.
1	Continuum of Fixedness in Phraseology	12
2	Percentage of Recurrent and Non- Recurrent Words in Conversation	29
3	Percentages of Recurrent And Non-Recurrent Words in Academic Prose.....	30



LIST OF ABBREVIATIONS

BAWE	: British Academic Written Corpus
BNC	: British National Corpus
EFL	: English as a Foreign Language
KTUCALE	: Karadeniz Technical University Corpus of Learner English
L1	: First Language
L2	: Foreign or Second Language
LL	: Log-Likelihood
LOCNESS	: Louvain Corpus of Native English Essays
OECD	: The Organisation for Economic Co-operation and Development
SLA	: Second Language Acquisition
TICLE	: Turkish International Corpus of Learner English

INTRODUCTION

Linguistics can be defined as the scientific study of the language. The aim in linguistics is to understand the language and its nature. It helps learners in a way that they can understand the origins of the words and languages with their historical backgrounds. Considering the fact that linguistics as a scientific field has been researched for a considerable amount of time in various respects it seems that there is a need to explore many other sub-topics at stake. It broadens learners' point of view to a language and gives them an insight to have look at in a different perspective. It is certain that linguistics is one of the fundamental components of language and it is inevitable for researchers to study linguistics. Current study presents a detailed study in the field of linguistics which enables learners to understand the nature of the language and to incorporate so many areas such as sociology, history, culture, philosophy, psychology, politics, art. In other words, it can be concluded that linguistics is the study of human and it enables a chance to understand language environment with the help of the linguistics.

As the linguistics and language coincide with each other it can be assumed that linguistics holds crucial value in language learning and teaching. Halliday (1966) underlines the critical role of linguistics in language environment especially in language learning areas, by implying that a teacher is not teaching linguistics in a class but he has to know linguistics so as to use teaching methods more effectively. Corder (1968: 74) states that "a teacher cannot teach a language by any of current techniques without linguistics knowledge, and that he does make constant use of what are basically linguistic concepts in this teaching". Based on this quotation it is possible to conclude that linguistics is vital in language teaching and learning and all language learners or teachers have to apply linguistics information into the process. Traditional methods in language teaching like grammar teaching method have been losing their popularity. Instead, new methods are increasing which means that language is taught in new methods and while applying, linguistics has to be known for better learning. Kjellmer (1991: 112) notes that "That language to a large extent relies on combinations of words that customarily occur" and it is now accepted as a general view in linguistics. As in an analysis of Kjellmer (1991), phraseology is the main subject in linguistics as languages are mostly based on phraseological terms.

Background of the Study

Phraseology which is one of the concepts in linguistics, is a widespread linguistics field to define and it is divided into many subfields such as phrasal verbs, minimal pairs, idioms, clusters,

lexical bundles. Pawley (2001) defines phraseology as “study of conventional phrases” in which phrase means multi-word combinations. Pawley also (2001) states that phraseology is crucial especially in the field of writing and speaking as there are many phraseological combinations used in well-documented languages. Cowie (1994) also underlines the importance of phraseology by stating that it is a major field of applied and pure research. Grenger and Bestgen (2014) state that phraseology is a reliable field in which it gives a chance to distinguish native learners from L2 learners and for that reason phraseology enables opportunities to analyze language learners. Despite the fact that phraseology has been on the scene, it has been neglected until recently by the linguists. However, there has been an incredible interest rising to phraseological subjects to get in the language to see the differences between L1 and L2 learners. Phraseology is in the center of the language and phraseological word combinations are mostly used to express ourselves and it is affected by several situations such as personal life of the learner, culture and habits. In another study, Gledhill (2000: 202) stated that;

The notion of phraseology implies much more than inventories of idioms and systems of lexical patterns. Phraseology is a dimension of language use in which patterns of wording (lexico-grammatical patterns) encode semantic views of the world, and at a higher level idioms and lexical phrases have rhetorical and textual roles within a specific discourse.

Position of the phraseology in linguistics cannot be underestimated and if a language is used in a most natural way as it is stated above, phraseology gives insight to understand the world and customs. For example; when someone forms phraseological sentences in target language, the world of linguistic features may sound different to native learners; however, quantity of their exposure to target language may enhance the performances of L2 learners. To be able to use a language as effective as native learners, being exposed to a language is vital and quantity of experiences in real life situations may enable L2 learners to be effective in target language. It can therefore be assumed that exposing a language is one of the most important ways of using the language naturally. Considering the fact that phraseological words are formed naturally based on intuitions and experience, it may take a long time for L2 learners to use the language in an effective way. Langacker (1987) explains that when someone experiences a structure thoroughly enough, forming word combinations becomes automatic which means that there is no need to spend any effort on how to put the structures together.

Phraseology is an area which consists of many sub-fields mainly lexical bundles, phrasal verbs, minimal pairs, collocations, idioms, clusters and more, and for that reason it is quite difficult to analyze all of the sub-fields as they also contain their own sub-fields which makes analysis more difficult for a research. Lexical bundles are crucial indicators in a language to see how language learners are successful within a specific discourse and they are also important for the construction of discourse. Allen (2010: 105) defines the bundles as “empirically derived formulaic units of language which are register-specific and perform a variety of discourse functions”. Lexical bundles

can also be defined as group of word bundles which consist of at least three words. Lexical bundles are mostly seen as three-word and four-word combinations though there are many examples of five-word and six-word combinations. Lexical bundles do not consist of structural units wholly but rather, they are part of the clauses or fragments.

Some linguists such as Hyland (2008) clearly illustrates the point that when the lexical bundles are longer, it means that frequency is lower. Lexical bundles are mostly used in speaking and writing inspired by the language environment and lexical bundles differ from idioms in a way that they do not contain fixed meanings. While meaning of a bundle can be inferred from the bundles, it cannot be inferred from idioms without looking context. Linguists such as Biber et al. (1999, 2003, 2004), Biber, Conrad & Cortes (2003), Cortes (2002, 2004) underline that lexical bundles are commonly seen in academic prose and they have significant discourse functions as well as structural correlates. It can be referred that lexical bundles are must for better academic writing skills. Lexical bundles are seen in academic writing and there are not any lexical bundles structuring complete units but seen as noun phrase or prepositional phrase. Biber (2009) differentiates lexical bundles seen in academic writing by stating that they are distinctive in academic prose.

Purpose of the Study

Lexical bundles are not used effectively in L2 learners' academic prose when we have a look at to the previous studies on bundles which show the fact that quite a few and inappropriate bundles are used. For example, Cortes (2004) made a research on bundles and found that lexical bundles are not used appropriately by L2 learners compared with L1 learners and they often make repetitions in their writing. Despite the fact that there are also some situations in which L2 learners use relatively more bundles than native speakers, it can be accepted that L2 learners have problems related to use of bundles in academic prose. L2 learners need to make use of lexical bundles appropriately in order to understand and develop discourse. Wide range of usage of lexical bundles helps language learners to achieve naturalness in language especially in their written texts. Academic discourse was defined by Hyland (2009: 1) as "the ways of thinking and using language which exists in the academy". It can be concluded that lexical bundles contribute academic discourse positively when they are used appropriately and effectively.

Speaking, writing, listening and reading are the main skills in a language and lexical bundles in writing is the main focus in the current study. These four skills enable us to develop our capability of comprehending and using the language. These skills are categorized into two parts, one of them is input in which listening and reading skills exist and the other is output in which writing and speaking exist. Output means that writing and production can be related closely to each other and learners have to produce written texts based on their knowledge. Writing is a difficult

field in which there may be variety of reasons for a learner to not comprehend it. Considering problems of L2 learners in writing especially in a way that how learners apply lexical bundles in academic prose is an area in which there is a need for a research. Lexical bundles which are used often in academic prose by L1 and L2 learners are text-oriented bundles.

Despite the fact that writing is an important skill for a language production, it is considered as a difficult skill especially in ESL. Tench (2003: 139) explains that “Writing is one of the foundations of successful public relations practitioner and ability communicate messages clearly and concisely is one of differentiating skills”. As to be seen in above, writing skills are needed, not just in the academic area but also in public environment and skills are crucial aspect of academic performance thus learners have to use writing skills properly.

The fact that writing is one of the main skills and it is a part of a language directed me to make research about the relations between lexical bundles and writing. Differences between L1 and L2 learners have been observed in previous studies of lexical bundles in writing. Comparing the performances of across groups it has been shown that there are differences in overall frequencies and structures. Despite the fact that there are relatively more studies in other cultures about lexical bundles, there are restricted amount of studies focusing this matter in Turkish context and for that reason Turkish learners’ awareness in terms of using lexical bundles in writing courses is in the center of current study. Lexical bundles performances and preferences of Turkish EFL learners are the main interest in current study. Because of the fact that there is lack of studies about the lexical bundle performances of Turkish EFL learners, this study will shed a light to this problem. One of the previous studies carried out by Gungor (2016) in his doctoral disassertation about lexical bundles performances shows that Turkish EFL learners have deficiencies while using bundles.

In this part, rationale of the current study is given to make it clear why phraseological awareness of Turkish EFL learners is important. The current study aims to see how Turkish EFL learners use lexical bundles compared with L1 learners and pay attention to problems faced while using bundles in writing skills. Many linguists such as Cowie (1998), Biber (1999), Seidlhofer (2004), Hyland (2008) stated that native learners are applying lexical bundles much more than non-native learners and there are many reasons for this situation. As lexical bundles come from the experiences of someone with the culture and environment, it is possible to see that native learners use more effectively; however, the matter is that why non-native learners cannot use bundles appropriately. When someone applies the use bundles, the matter of being exposed to them comes to scene as native learners use them intuitively rather than consciously.

Significance of the Study

Lexical bundles can make one text more elaborate and fluent and when they are not used in texts, fluency may disappear. Texts are getting stronger when lexical bundles are used enough but gets weaker when they are not used. Thus, learners may have difficulty to understand. In that study, how effective lexical bundles are and how inevitable they are for writing skills so as to make EFL learners may use them effectively are studied. Comparing overall frequencies of lexical bundles preferences of Turkish EFL learners within a crosslinguistic way with native learners is the hearth of the current study. In that way, lexical bundles are analyzed for an appropriate language use.

Hill (2000) underlines that; without looking their names, they are considered to be essential parts of a language without which a learner may not gain native-like fluency and scholarly writing proficiency. Hoey (2000: 202) states that; “learn not just the meanings of the words but the environments they occur in” is one of the most important strategies in language use. When someone is exposed to the language it will be easier to use the features of the language more effectively. Also, Sinclair (1991) underlines that instead of always making new combinations of individual words, native learners are relying on stock of prefabricated word chunks. With the help of this corpus-based thesis, there may be a chance to see most common bundles native learners used.

Meunier (2012: 112) stating the importance of phraseology expressing that;

(...) adopting a formulaic approach to L2 teaching seems relevant for three main reasons at least: (a) formulaicity is ubiquitous in language; (b) formulaic language use has been shown to be a marker of proficiency in an L2...; and (c) studies have demonstrated that L2 language learners find formulaicity particularly challenging as it is impossible for them to use the innate native intuition usually associated with formulaic language use.

Many linguists have the idea that phraseology in other words formulacity is one of the key factors for learners to develop language skills and with the help of the current study awareness of the Turkish EFL learners is evaluated and using lexical bundles especially in writing skills is encouraged. Significance of lexical bundles is aimed to be shown in current study so as to make learners to concentrate on and make use of them.

Statement of the Problem

Formulaic language in other words multi-word combinations hold vital position in linguistic environment since they contribute to use the language more proficiently. Most of the studies show us that the more we use word combinations the more we are get in the languages thus it is crucial for L2 learners to use target language as much as possible. Many studies have been carried out about lexical bundles and most probably most of the studies were contrastive analyzis which means

that usage of lexical bundles by L1 and L2 learners are compared to each other. For instance, in his study, De Cock (2000) found out that French learners overuse the multi-word combinations 2 to 6 words compare to the L1 learners. Parallel to De Cock, Hyland (2008) informs that there is an overusing problem among L2 learners in his study. However, it can be said that both De Cock and Hyland's studies contradict with many studies carried out by other linguists' as there is a problem of underusing is dominant in those studies. In the study of Chen and Baker (2010), we can see that Chinese Learners of English used much less bundles than L1 learners. Also, Biber and Conrad (2004) figured out that L1 learners use lexical bundles more than L2 learners and the reason for that problem is that L2 learners are not exposed the language as much as L1 learners. Adel and Erman (2012) stated that texts of L1 learners contain much more bundles than L2 learners.

Although there has not been enough number of studies in a Turkish context it can be concluded that Turkish learners have problem in terms of using lexical bundles. For example, in the study of Ozturk and Kose (2016), it is easy to see that Turkish learners overuse the lexical bundles in terms of variety and frequency. Bal (2010) underlines that Turkish learners use plenty of lexical bundles but most of their preferences are not seen in L1 learners' texts and they also overused some lexical bundles. Finally, Gungor and Uysal (2016) states that L2 learners overuse some of the lexical bundles which are not preferred by L1 learners. Although, there has been ongoing problems related to use of lexical bundles by L2 learners, number of the studies are not satisfactory and this situation directs linguists to research lexical bundles in Turkish context.

Research Questions

Significance of lexical bundles especially in the field of writing is shown in current study. Howie (1991: 25) expresses that "To find the right question requires that we understand what we are asking about, and know to keep the question simple enough to be answerable, but challenging enough to be interesting". When a question is stated well then, it may be accepted that half answer is ready. Parallel to quotation above which clearly demonstrates the importance of forming a question, main research question and sub-questions for current study are stated below;

- What characterizes Turkish EFL tertiary level learners' phraseological awareness in terms of lexical bundles in their writings?
- Are there any differences between Turkish EFL learners' and native learners' lexical bundles performance?
- What are the mostly preferred lexical bundles used by Turkish EFL learners and differences from native learners' mostly preferred bundles?

- What are the significant lexical bundles performance differences in expository argumentation and academic argumentation of Turkish EFL learners?



CHAPTER ONE

1. LITERATURE REVIEW

The introduction chapter aims to establish bridges between phraseology which is in the heart of the linguistics and used by linguists to understand language and the aim of this study which is closely related to phraseology in terms of what it contains and related terminology is to investigate phraseological awareness of Turkish EFL learners. Phraseology which plays crucial role in linguistics and has contributed to perceive authentic language. Phraseological awareness of native speakers (L1) as well as second language learners were analyzed in details to find out differences. Components of this study have been introduced and discussed based on importance of the phraseology and its place in the field of linguistics which has been followed by the research aims. After that, reasons that why phraseological awareness of learners is crucial in a language were explained. Following this part research aims were stated and discussed. Finally, the outline of this thesis has been stated to show the way that this thesis has been carried out.

Phraseology is the study of ways how the language is used especially in the choice of the words and expressions. Susan Hunston (2011: 5) states that “phraseology is a general term used to describe tendency of words and a group of words, to occur more frequently in some environments than others”. In other words, it can be concluded that recurrent word combinations are often seen in a language and they are encountered in language registers. Cowie (1994) defined phraseology as “phraseology is generally perceived as the study of structure, meaning and use of the word combinations”. According to the Gledhill (2000: 202) “phraseology is a dimension of language use in which patterns of wording (lexico-grammatical patterns) encode semantic views of the world, and at a higher-level idioms and lexical phrases have rhetorical and textual roles within a specific discourse”. Each culture has some impacts on languages which means that when language users are exposed to target cultures, they encounter lexical bundles more often. Firth (1957) states that “you shall know a word by the company it keeps”. As noted by the quotation above, words and word combinations have to be known to use language in a better way. However, it may not be as easy as it is seen because of the fact that learners are needed to be exposed enough to a language so as to be efficient language users. It is not uncertain that the research on phraseology (formulaicity) goes back to middle of 20th century. Phraseology dealing with the use of word combinations which are highly important in their own context and give learners opinions regarding the meaning is used in that context. In the next part below, the features of phraseology which are significant key terms in the environment of phraseology were introduced.

1.1. Terminology in Phraseology

It is quite difficult to brand the terms related to phraseology under the same umbrella by different researchers as classifying the terms may differ from one researcher to another. One of the main reasons for this situation is that word combinations vary a lot and classifying word combinations under the umbrella of phraseology can be difficult. In an analysis of Altenberg (1998: 101), he pointed out that;

Phraseology is a fuzzy part of language. Although most of us would agree that it embraces the conventional rather than the productive or rule-governed side of language, involving various kinds of composite units and 'pre-patterned' expressions such as idioms, fixed phrases, and collocations, we find it difficult to limit the area and classify the different types involved.

As the notion of phraseology is quite broad to be defined there could be lot of definitions and features of phraseology in the literature. By drawing on the concepts of phraseology, Stefan Th. Gries (2008: 5) emphasizes that “while the notion of phraseology is very widespread, just as with other linguistic concepts, different authors have defined it differently, sometimes not providing a clear-cut definition, or conflating several terms that many scholars prefer to distinguish”. Gries (2008: 5) defines a set of parameters which are typically seen in phraseological study shown below.

1. the natural structure of the elements in a phraseologism (grammatical and lexical items);
2. the number of elements involved in a phraseologism;
3. the frequency of an expression must be observed before it is counted as a phraseologism;
4. the permissible differences between the elements included in a phraseologism (discontinuous phraseologisms, immediately adjacent elements);
5. the degree of syntactic and lexical flexibility of the elements involved (completely inflexible patterns, standardly quoted as by and large, relatively flexible patterns such as kick the bucket, which allows different tenses but, e.g., no passivization));
6. the role of semantic non-compositionality / non-predictability and semantic unity in the definition (function as a semantic unit in sentence or clause).

These parameters are supported by researchers like Schmitt and Carter who states that there is little wonder about these differences. Whilst some linguists have mainly stated that there could be one group of parameters occur in phraseology, others have highlighted that different categorizations may occur. Thus, this situation supports the idea that terms used under the phraseology umbrella may vary from each other when they are classified by different researchers.

On the other hand, in the literature phraseology can be classified as a part of formulaic language which can be defined as recurrent word combinations as well. Wray (2002: 9) states that formulaic language can be defined as; “A sequence, continuous or discontinuous, or words of other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar”.

It may be possible to conclude that studies on phraseology can also be called as “formulaicity” which is broader than phraseology figured out from the definition of formulaic language. Although there were many different terms which were equal to phraseology used by different linguists in the past years, now they are all generalized as phraseology.

Any form of word combination can be accepted under the umbrella of phraseology as they are all consisted of by interaction among language users and may differ when their positions in a sentence were taken into consideration. A general categorization that Cowie (1998) is given in Table 1 below though he emphasizes that there may be differences among terms and it may differ in someone else’s categorization.

Table 1: Terms Used For ‘ Sentence-Like ’ and ‘ Word-Like ’ Combinations

Author	General Category	Sentence-like (Pragmatic unit)	Word-like (Semantic Unit)
Cheusheva (1964)	Phraseological	Phraseological expression	--
Zgusta (1971)	Set Combinations	Set group	--
Mel’čuk (1988b)	Phraseme, or Set phrase	Pragmatic or phraseme	--
Gläser(1988a)	Phraseological	Proposition	Nomination
Cowie (1988)	Word- combination	Functional expression	Composite unit
Howarth (1996)	Word- combination	Functional expression	Composite unit

Table 1 provides an overview of the main categories of the phraseology parallel to some linguists’ observation and these categories may differ in some points from each other. Bound to the categories appear on table sub-categories may also appear in phraseology. Other terms such as *lexical phrases, collocations, lexical bundles, idioms, n-grams, skip-grams, chunks, clusters, minimal pairs* are sub-variations of the main categories thus it may be counted that there are hundreds of sub-categories for the terms of phraseology. Most of these sub-categories are not the same and they cannot be used interchangeably as as they vary. Lexical bundles which are highly applied and appear in language are my main study field in this research. While some of those sub-categories are quite different from each other, some of them are similar as word combinations are repetitive such as n-grams, lexical bundles and clusters.

1.2. Features of Phraseology

This section gives useful information about the features of phraseology which enlighten us about the scope and borders of the phraseology and phraseological variations. By looking inside of the features, it will be easier to understand my research topic. Huang (2014) categorizes features of

the phraseology into four main fields namely idiomaticity, fixedness, semantic unity, frequency of co-occurrence.

1.2.1. Idiomaticity

Idioms are probably the main basement for phraseology and its definition according to Cambridge Dictionary (2008) is that “group of words used together with a meaning that you cannot guess from the meanings of the separate words”. Sonomura (1996: 28) makes definition of idioms as “units which display phrase-like behavior in some respects but word-like behavior in others, paired with the predominance of generative grammar throughout most of the twentieth century, relegated them to the margins of linguistics”.

Idioms are metaphorical and cannot be understood without looking the meaning and context in which they stand, and they are fixed words in other words they are invariable and no words can be added to idioms thus idioms can be accepted as word clouds in which they are meaningful when they are accepted as a whole but not one by one. Just like to definition of terms of phraseology there are no clear types of idioms existing though it may vary among researchers or one reference book to other reference book. In the Oxford Dictionary of Current Idiomatic English, for example, it can be seen that the kinds of idioms are based on an understanding of idiomaticity at least to some extent. Linguists, Cowie, Mackin, McCaig (1993: XII) underlines that;

A view of idiomaticity which does full justice to the rich diversity of word combinations in English must recognize that the meaning of a combination may be related to those of its components in a variety of ways, and must take account also of the possibility of internal variation, or substitution of part for part.

To give an example, Longman’s Dictionary of English Idioms presents twelve types of idioms whilst Carter (1998) presents us mainly six types of idioms which also have sub-types namely; proverbs, idiomatic similes, catchphrases, stock phrases, allusions and discursal expressions. These categories reflect the diversity of languages and show how a language is complexed. However, another linguist Balfaqqeh (2009 5-6) criticizes that “this classification does not have defined boundaries and a structural overlap is very much expected”. It can be seen from the examples that categorization of idioms is quite hard and changeable.

These kind of multi-word occasions exist in language and also can be called also as formulaic language, lexical phrases, routinized formulas or predesigned language chunks such as “on the other hand”, “so as to” and “so on”. Linguists like Sinclair, Wray, Corrigan, Schmitt used idiomaticity just like formulaicity and they behaved them in the same way also accepted them nearly the same. Wray (2012: 234) underlines that “we have long have been saying that formulaicity shapes the language”. Wray (2002: 9) makes the definitions of the formulaicity or

idiomaticity as, "a sequence or continuous or discontinuous, of words or other meaning elements which is, or appears to be prefabricated: that is, stored and retrieved whole from memory all the time of use, rather than being subject to generation or analysis by the language grammar".

In general, therefore it seems that idioms are crucial for the language and they are significant components of a language. phraseology is in the center of a language in which nearly everything related to language is born and thrives and shapes the language we use. In other words, using of phraseological features of the language is a must for speakers to convey the message in a clear and fluent way. Thus, idiomaticity cannot be underestimated in a language by language users.

1.2.2. Fixedness

Fixedness is another important feature of phraseology which is explained and discussed by many linguists such as Sinclair, Cowie, Gries etc. Fixedness means that there is lexical flexibility and syntactic substitutability in a language and this situation caused many discussions and contradictory ideas. To say it with other words, someone can assume that fixedness shows to what extend a word combination is bound to each other lexically and syntactically. Some of the word combinations which are accepted under the umbrella of phraseology may change in a form and meaning such as collocations, while some of the word combinations cannot vary in a form or meaning like idioms. This situation shows us that fixedness can differ in word combinations in terms of meaning and form, though in general it is difficult to see differences in phraseological combinations. Cowie (1981) prepared a figure which shows us continuum of fixedness in which some clear points related to phraseology were shown. Graphic 1 below shows fixedness in categorized positions enables us to see some of the components of phraseology in fixedness aspect. As it is shown in the graphic below, it may be easier to see which of the terms in phraseology are fixed or semi-fixed or not fixed.

Graphic 1: Continuum of Fixedness in Phraseology



Idioms are strictly fixed word combinations which means that it is quite difficult to see a change in a form or meaning among words. To give an example "*break a leg*" is an idiom means "good luck" and it can be seen that it is a fixed idiom as this two words are combined with each other and different meaning occurs. There cannot be change related to lexically and syntactically. However, there are also word combinations such as "*break the window*", "*break an arm*" and they

cannot be accepted as idioms because of the fact that they don't create another meaning when they are used together.

Especially some of the collocations can be restricted or semi-fixed though they are less idiomatic. Sinclair (1991) makes the definition of collocation as "co-occurrence of two or more words within a short space of each other in a text". Collocation was defined in Cambridge Dictionary as "a word or phrase that is often used with another word or phrase, in a way that sounds correct to people who have spoken the language all their lives, but might not be expected from the meaning".

Collocations are restricted or semi-fixed group of words which are less idiomatic in terms of their semantic usage which is bound to people who use it, while idioms are independent.

Free combinations are combinations of the words that can be created without being dependent to any lexical or syntactic rules just as the name suggests. Users can create free combinations themselves in anytime throughout with their experiences in a language environment.

These findings enhance our understanding of the the fixedness, it can be seen that fixed combinations are resistant to any kind of effect, they cannot change in other words they are untouchable while semi-fixed word combinations may change both lexically and syntactically and free combinations are as the name suggests can be changed by the user.

1.2.3. Semantic Unity

Another important feature of phraseology is semantic unity. Semantic briefly can be described as study of meanings. Syntactic features of the word combinations are not taken into consideration but meaning is on the stage thus in phraseology when words combine to each other, in time the semantic unity occurs and comes to crucial place among word combinations. Sinclair (1991: 112) observes that "many uses of words and phrases show a tendency to occur in a certain semantic environment". Huang (2014: 6) states that "without semantic unity as required quality, two-word combinations may be counted as phraseological unit, as evidence from British National Corpus (BNC) shows that most occurrences of two-word combinations stands together". Although there can be examples of one-word occurrences in many examples, words are bound to each other. For example, when "*so as to*" is taken as an example, it may be stated that many sentences contain "*as to*" however when details are taken into consideration, it can be seen that example of these occurrences are actually instances of *so as to*.

It can be completed that semantic unity in phraseology holds important position as phraseological units have semantic functions and thus researchers may have chance to differ semantic unity in word combinations.

1.2.4. Frequency of Co-occurrence

Frequency of co-occurrence also plays crucial role among the features of phraseology as it is inevitable for language learners to know how often does a word combination exists. Frequency of co-occurrence can shortly be defined as; a number of times word combinations occurs in a language. Having information about frequency of co-occurrence enables learners to know which combinations are highly applicable in a language apart from enabling learners the meaning and the context they occur. Biber (2006: 173) states that “frequency patterns are not accidental, but they are also not explanatory in themselves”. Frequency may be the one of the most applicable criteria to see meaning of the combinations in context and may help learners to decide which use of the word combinations are most applicable. Frequency of co-occurrence holds the most significant position among the features of phraseology for this thesis as seeing number of the times that students used word combinations is in the center of this research. According to the Mint (1985: 172) “Frequency data from a corpus is needed to respond to problems of selections and progression. They can help to make decisions about the selection of language items for teaching and about the sequencing of these selected items which determines the line”.

While choosing the language items which are needed in a context, frequency of that language item can provide learners with best choices in a target context. Having known the frequencies of word combinations learners will be able to apply them in a suitable context that chosen in which word combinations occur. Frequency can be regarded as main guiding principle for words as it shapes language and they are chosen to be used by looking to places where they stand. Seeing and trying to understand how recurrent words are used especially in corpus relying on frequencies, can enlighten learners about usage of them in a appropriate way in a language.

1.3. Major Theories of Phraseology

Phraseology has gained great interest for more than thirty years mainly in Western Europe and then in USA though these cultures ignored it before that time. Soviet Union linguists studied phraseology much earlier than the Western Countries. Because of the great incline, phraseology has become one of the hottest topics for linguists which has resulted in many studies that enlighten the languages in use. There have been international conferences mainly dealing with phraseology. Cowie (1998: 2) underlines that;

Recognition of phraseology as an academic discipline within linguistics- the term itself, like the adjective 'phraseological', reflects Eastern European usage-is evident not only from vigorous and widespread research activity, but also from the publication of several specialized dictionaries reflecting one theoretical perspective or another and from the attention increasingly given to the subject in textbooks on lexical semantics and vocabulary in language teaching.

Classical Russian theory which was first mediated by non-Russian linguists such as Klappenbach (1968), Weinreich (1969) and Arnold (1973), has great impact on phraseology and thanks to that impact, phraseology took the attention and came to the stage for linguists. Then, Melcuk, Teliya and some of the other Russian linguists have contributed phraseology to become more active in a language and embrace the culture more. This theory mainly depends on descriptive categories such as comprehensive, systematic and soundly based though categorization in phraseology may differ among the linguists. Most of the categorizations done before categorizes phraseology syntactically. With the help of these categorization of linguists, it is possible to understand how to use word combinations effectively, which means that also linguistic analysis started too. Many linguists were affected by Classical Russian theory in west so they put the bricks up on this theory. Cowie, Howarth are the pioneer of these linguists who have been affected positively by Classical Russian Theory. Melcuk (1988) prepared his categorization under the effect of Classical Russian theory, has been shown in Table 1. European Theory had the great effect of Russian theory and improved under the impact of it. Early categorizations were based on word-like categorizations while Classical Russian categorizations are sentence-like categorizations. Many linguists carried out their researches based on the Russian theory. Russian and European theory contributed a lot to phraseology to improve and after this contribution phraseology has been main interest by many linguists.

Another theory is the cultural element which means that phraseology is *linguo-cultural*, like an inner part of culture, underlines the point that phraseology is totally together with the environment and culture. Teliya and her colleagues explains many ways that language and phraseology penetrate with each other, three of the main ways are *cultural*, *semantic* and *component*, in which word combinations occur by being totally affected by historical background of the words. *Cultural concept* which means that word combinations are being affected by abstract cultural notions and *cultural connotations* which stands for combination of these another two components both cultural semantic component and cultural concept. Cowie (1998) supports this theory with the opinion that cultural connotations arise from the interpretation of concepts or sub-concepts.

1.4. Phraseology in Language Learning

Phraseology has been dominant factor in language recently, as there has been increasing growth interest by language users and linguists. Not only because of the reason that phraseology is

frequently used in language but also the factor that why phraseology plays crucial role in language learning and teaching. Phraseology and the culture intertwine with each other thus use of phraseology should be applied more in language learning and teaching. For instance, in Classical Russian Theory, Teliya and her colleagues argued that phraseology is fruitful focus for “linguo-cultural” analysis. According to Sinclair (1998: 9) “every language, especially its figurative meanings, are concerned with the reflection and extension of the world-view shared by the community”. Expressions, words, idioms, chunks etc are all conveyed to generations which implies that phraseology contributes and reinforces cultural norms. Phraseological tendencies of communities which can differ among different communities, can affect the language and culture in some circumstances like how phraseology is active or dominant in that language.

Biber (2009: 301) states the importance of phraseology in language by explaining that;

Formulaic language is very important in both conversational and written academic discourse, but it is realized in very different ways linguistically: fixed sequences that represent clause fragments in conversation, versus formulaic frames that consist of noun phrase and prepositional phrase fragments in academic writing.

Sinclair (2004: 29) underlines that “phraseological tendency occurs in language where meanings are created through word combinations”. Bolinger (1976: 1) informs us that “the phraseological nature of the language has long been recognized as language does not expect us to build everything starting with ladders, nails and blueprint and rather it provides incredibly large number of prefabs”. Researchers have accumulated extensive data for phraseological tendency. Sinclair (2004: 29) also states that “words tend to go together and make meanings by their combinations”. Combinations are bound to cultures’ use of the language which means that bound to their phraseological tendencies. There are factors that affect the position of the phraseological tendency namely, lexical priming theory, pattern grammar and idiom principle and these factors are helpful for learners to understand the importance of phraseological tendency. Phraseology supplies us with the rich features of the language which help us improve our speaking and writing skills and develop fluency in a language. Thus, phraseology in language learning holds critical position in lessons. Sinclair (2009: 9) underlines the importance of phraseology by stating that;

If a word is likely to be intricately associated with the words that occur round about it, then the consequences of studying its meaning in isolation are unpredictable.[...] Words have many meanings, and there is no way of working out in advance which one is appropriate in a text. [...] However, if we extend the viewpoint to two or three words (which is normal when lexicographers recognise a relatively fixed phrase) much of the ambiguity drops away. People use this extended viewpoint so naturally in reading and listening, and language teachers labour the importance of concentrating on the broad aspects of meaning and not the particulars of a single word.

The generalizability of this studies shows that phraseology is a must for a language and it develops learners' communication skills as language and the culture is integrated to each other and use of the phraseological word combination is highly preferred by language users.

1.4.1. Lexical Priming Theory

Some linguists like Chomsky notes that first grammar is generated and then words are dropped in opportunities. However, according to the most of the classical theories, some of the other linguists like Hoey states this in a reverse way by changing the roles of the lexis and grammar. Apart from Hoey, Firth (1957: 154) also explains that;

The traditional categories of grammatical description are survivals of a medieval scholastic instrument. They have been used to deal both with the forms and meanings of linguistic constituents in the vaguest of socio-philosophical terms, and judged by modern standards they have been found wanting in both enterprises. ... Is there any more reason to perpetuate them than medieval alchemy?

Hoey believes that lexis is complexly and systematically structured and the grammar is result of this system. According to him words are primed as a result of someone's or cultures' experience with words in other words everything related to word combinations is a result of experiences with them. Learners prime the words parallel to their interaction with that word, or similar situations that they have lived before. Lexical priming, having impact on discourse can be defined according to the Mac Millan Dictionary (2004) as;

The theory of lexical priming suggests that each time a word or phrase is heard or read, it occurs along with other words (its collocates). This leads you to expect it to appear in a similar context or with the same grammar in the future, and this 'priming' influences the way you use the word or phrase in your own speech and writing.

Words and word combinations are primed by learners regarding that how much experience with these words occur and they are used in the language in that way. Pinker (1994) states that semantic is first generated and lexis merely actualizes the semantic which means that it is the lexis that making semantic happen in real life situations. When someone first encounters word he/she experiences them and when he/she re-encounters that word repetitively then it can be remembered from those experiences and they are used in language environment. By lexical priming Hoey (2009: 34-35) means that;

whenever we encounter a word, syllable or combination of words we note subconsciously (1) the words it occurs with (its collocations), the meanings with which it is associated (its semantic associations), the grammatical patterns it is associated with (its colligations), and the interactive functions it contributes to serving (its pragmatic associations)", (2) "the genre and/or style and/or social situation it is used in", (3) "its text- linguistic characteristics: the positions in a text that it occurs in (its textual colligations), the cohesion it favours or avoids (its textual

collocations) and the textual relations it contributes to forming (its textual semantic associations)..

As it can be seen from the quotation above, interactions of learners with the environment direct them to use word combinations sub-consciously in which they transfer the word combinations from short-term memory to long-term memory. When learners encounter words in real life situations, it is significant that how often they are being exposed these words and learners start using these words when they become a habit in a time. Specific combinations are not chosen consciously by learners, these word combinations are uttered if they are recurrent and frequently seen in the language environment.

In Sage handbook of social psychology Hogg and Cooper (2007: 55) states that;

(...) factor that influences the accessibility of information in memory is priming. The activation of stored knowledge through experiences in the immediate context can make prime-relevant information more accessible in memory, and such recent construct activation can influence inferences, evaluations, and decisions on subsequent tasks (...)

Hogg and Cooper (2007: 55) also underlines that;

“the frequency with which a construct has been primed (...). Traits, attitudes, or stereotypes that have been frequently activated in past experience are more available in memory than those that have been less frequently primed. This kind of frequency of activation, if it reveals on a continuing on regular basis, can result in certain structures becoming chronically accessible, such that no external priming in the immediate context is necessary to make them highly accessible. Moreover, as people differ in the types of experiences, that would generate such routine construct activation which means that particular constructs are chronically accessible as individuals differ quite naturally from each other”.

Hoey (2005: 13) states that “Every word is primed for use in discourse as a result of the cumulative effects of an individual’s encounters with the word”. In other words, it can be figured out from this statement that, someone’s experiences in life, leads him / her to use discourse parallel to experiences. Lexical priming is closely related to discourse because learners prime words intuitively. Table 2 below illustrates how is priming used among people and what affects lexical priming in a language environment.

Table 2: Priming Hypothesis by Hoey (2005)

- | |
|--|
| <ol style="list-style-type: none">1. Every word is primed to occur with particular semantic sets; these are its semantic associations.2. Every word is primed to occur in association with particular pragmatic functions; these are its pragmatic associations.3. Every word is primed to occur in (or avoid) certain grammatical positions, and to occur in (or avoid) certain grammatical functions; these are its colligations.4. Co-hyponyms and synonyms differ with respect to their collocations, semantic associations and colligations. |
|--|

Table 2: (Continue)

- | |
|---|
| <ol style="list-style-type: none">5. When a word is polysemous, the collocations, semantic associations and colligations of one sense of the word differ from those of its other senses.6. Every word is primed for use in one or more grammatical roles; these are its grammatical categories.7. Every word is primed to participate in, or avoid, particular types of cohesive relation in a discourse; these are its textual collocations.8. Every word is primed to occur in particular semantic relations in the discourse; these are its textual semantic associations.9. Every word is primed to occur in, or avoid, certain positions within the discourse; these are its textual colligations10. Every word is primed to occur with particular other words; these are its collocates. |
|---|

These hypothesis by Hoey explains the situations in which learners use lexical priming and how does priming occur. The lexis and grammar from different point of views can be seen above. These theories are supported with the corpus data in a way that the words which were examined are taken from the real situations.

1.4.2. Pattern Grammar

Hunston and Francis (2000: 3) define pattern grammar as “corpus-driven grammar of English which describes the syntactic environments of lexical items”. Some of the patterns occur in a language frequently while some of the other patterns do not appear as much as others is a nature of the languages. Hunston and Francis (2000: 3) also state that, “patterns and the lexis are mutually dependent to each other as each pattern occurs with restricted set of lexical items and each lexical item occurs with a restricted set of patterns”. It is not always clear from the intuition that a word can be used with which pattern, in that circumstance learners may make use of the corpus and it enables researcher or speaker to see which patterns are mostly used with the word that researchers used. Word patterns could be defined as all of the words and all structures which are regularly associated with word contributes the meaning. Patterns can be association of variety of different words as well as a word can have many patterns. Susan Hunston (2000: 13) also emphasizes that “Grammar patterns, on the other hand, constitute an attempt to describe the whole of the language (or rather, all the frequently-occurring items in the language) in a principled way, and the lists of words collected in a given pattern are not random”. Patterns are essential for language learning as they are mostly language and cultural product and also enable learners to compose utterances without much effort. Sinclair (1992) states that “patterns are important not only because of they are crucial to usage but also pointer to its meaning”. Pattern structures shows whether the element or elements are object, the propositional object, complement or adjunct.

1.4.3. Idiom and Open-Choice Principles

Sinclair brings out two models namely idiom principle and open-choice principle which contribute to find out how the meanings are arising from the language texts. Sinclair introduced the distinction between open-choice principle and idiom principle that help learners to understand the subjects. Sinclair (1991: 110) “The principle of idiom is that a language user has available to him a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segment”. Sinclair’s idiom principle is mainly concentrating on the presumption that language users select naturally from idioms or pre-constructed phrases, just like the name of the model implies.

However, open-choice principle is another term which distinguishes from idiom principle in some aspects. Sinclair makes definition of open-choice principle as the open choice principle “is a way of seeing language text as the result of a very large number of complex choices”. At each point where a unit is completed—a word or a phrase or a clause ---a larger range of choice ends up, and the only restraint is grammaticalness.” Sinclair considers open-choice principle as the normal way of seeing and describing the language. These two theories are opposite to the each other and none of them provides sufficient explanation according to Erman and Warren. Erman and Warren (2000) revealed in one of the researches that %45 of all authentic text goes with open-choice model while %55 of authentic models belongs to idiom principle model. These both models show that words are not applied merely but to be co-selected instead which means that learners choose the words do with together.

1.5. Phraseology in EFL Context in Turkey

As phraseology has been popular field for linguistic recently, it has also gained importance in language teaching parallel to this situation, thus many researches and studies have been carried out by linguists and language teachers so as to enhance the learning level of EFL students. Because of the fact that phraseology is shaped by the authentic language used in daily life and academic life, it comes to the stage more rapidly than ever. Critical position of the phraseology in a language has been studied by some linguists in twentieth century. First a research in cognitive science has figured out that frequency of occurrence and frequency of experience establish words and collocations as units of learning in other words they have been used in teaching the language as well. Phraseology had always been neglected up to that time and by gaining critical position in language learning, it has become one of the utmost factors in language and language teaching.

According to Sinclair (1970: ix) “the idea of the collocations first emerged among some language teachers between the two world wars, particularly that of Harold Palmer in Japan”. Palmer (1933: i) suggested that “collocation is a succession of two or more words that must be

learned as integral whole and not pieced together from its component parts”. Later other linguists like Firth, Sinclair, Halliday & Hassan, Wray also stated that formulaic language has effects on language and especially the field of language learning. To see the effect of the phraseology in EFL environment, in that study Foreign Language Teaching in Turkey, Phraseological Researches in Turkish EFL setting, Problems in Language Learning in Turkey and Limitations of the Study in the Field of Phraseology were discussed in details to figure out the problems and possible solutions about this matter in Turkish context. As the tertiary level language awareness is the main aim in that study, it holds the most detailed position in that part.

Although there are studies which were conducted to analyse phraseology in linguistic environment, these are not enough for such a broad field as there are many branches of phraseology exist. Recent interests to phraseology show us how critical it is in language environment. There are comparatively more studies in the branches of phraseology such as collocations, idioms while there are limited number of studies about lexical bundles. In the linguistic environment, EFL learners have to use phraseological words appropriately to be active, clear and fluent in target language so researches may shed a light to this linguistic field and enable us with sufficient data which learners will make use of it during the process of communication.

1.5.1. EFL Learners in Turkish Curriculum

Foreign language teaching especially English as it is spoken throughout the world, has had important place in Turkish curriculum. Learning English starts at second grade level in primary schools and it continues until the end of the university level. As Turkey’s geographical position is convenient for tourism and trade, knowing the language contributes and affect people especially who live in the west of Turkey. Thus, it can be seen that language learning is inevitable in Turkish curriculum. Celebi (2006: 292) states the reason for learning foreign language especially English for learners as; “Knowing the foreign language improves the tourism, knowing the language fastens the development, Turkish is not a language of science, western languages especially English has become international language”. Thus, English language has been being thought in all levels of schools in Turkey also in universities. English has been being thought partly or completely. Students take more than 1000 hours lessons of English in Primary, Middle and High schools in total. Also, in some of the university students study preparation class in which they study more than 15 hours per a week and after the preparation class they have at least 4 hours lesson per a week. Apart from this some universities use the English as university language and every lesson is thought in English. Also, some universities give extra courses for English so as to improve the level of the students. When curriculum of the Ministry of the Education (2009) is glanced, it can be seen that learning English is mainly divided into 4 sections namely, speaking, writing, reading-understanding and listening-understanding. While, learning the language there are some important points contributing language learning. First, someone cannot learn the language if he isolates

himself from the language and culture. Kargi (2006: 62) points out that “No one can learn the language as being isolated from the culture and cultural values.” Also, Nunnan (2008: 104) underlines that “the language should be in the position that in which there is combinations of the culture and students are active”. Another important factor is that willingness which stimulates learners to learn the language. If one person is reluctant to learning the language, it becomes harder to learn in an expected level. Motivations to language learning should be developed to make learners use the language in expected level.

As it is my main target, information about the English lessons of tertiary level students who learn English language starting with the middle school years was given in details.

Despite the fact that nearly all of the universities in Turkey have their own special curriculum programs to teach English language, students proficiency level contradicts with the level that they have to be in schedule. According to the British Council (2015) which is accepted as one of the most popular companies in the field of English as L2, there are 175 universities and they have their English curriculum. British Council and The Economic Policy Research Foundation of Turkey (2015) states that there have been ongoing problems faced while learning L2 in tertiary level which may lead disadvantages in future. According to the report students take more than 1000 hours of English courses till the end of the high school but the learning level is not acceptable when learners enter the universities. Çetinsaya (2014) states that in Organization for Economic Co-operation and Development (OECD) countries, “there are 16 students in each class and Turkey is far behind of this number. It needs 45.000 academician to get the level of the OECD countries”. According to report of The British Council and The Economic Policy Research Foundation of Turkey (2015) universities can be divided into three categories according to their English education system. First there are universities in which teaching language is totally English, regardless of whether lesson is related to English or not, whole language is English. In every lesson students and academician communicate with each other in English and exams are in English too. Secondly, universities which have Turkish as teaching language in that situations the universities constitute preparation class which is followed by the English courses in schedule in following years. Finally, the universities present English-Turkish teaching language environment which means that students use English as a teaching language in some of their lessons. Although enough information about the education system in universities occurs, there is not specific information about how much phraseological variations are thought during their lessons or while learners are using the language. As using phraseology in language is inevitable for the speakers of that language, it will be useful to see the level of tertiary students in terms of using phraseology.

To sum up the information it can be concluded that there are enough lessons for tertiary level of language learners who are taking English courses throughout their university period and English as a foreign language holds crucial position in the policy of the government. Although classes are

more crowded than it should be for a language learning, lesson hours are acceptable for learners to get sufficient level of English.

1.5.2. Problems Faced in Teaching and Learning English

Learning a language contributes a person in several ways namely; personally, economically, culturally etc. Oktay (2015) states that;

One of the functions of Turkish Educational system is to prepare enough and qualified manpower, required by economics. In order to be able to open the ways for this qualified manpower, for economic and social improvement and change, it became more important to know even more than one foreign language.

Also Vergili (1993: 25) underlines that every individual may benefit in many different ways when they learn a foreign language. For example, they may get higher salary or position in their career, they may have better social status and have deep understanding. Culture is another feature that may affect language learning positively or negatively. Widdowson (1990) explained that semantic and syntactic systems are acquired by a child at the same time. If languages are thought without looking cultural factors, learning or teaching will be useless.

In Turkey it is commonly believed that EFL learning is not fruitful though there have been many attempts aiming to develop the level of the language learning for many years. According to common belief in public main reason for that problem is that; methods which have been being applied in language teaching are based on the classical methods like Grammar Translation Method, deficiencies of the language materials in language teaching and problems are faced while planning to curriculum for language learning. Although it is compulsory in Turkish curriculum, the level that has been expected from the learners has not been efficient for years.

Minister of Education, Yilmaz in 2017 states that “there is ongoing problem in language learning in Turkey. Despite the enough class hour level of language learning level is not acceptable”. It can be seen in that statements that there are sufficient class hours for language learning but result is not satisfactory and the problem is not a new and goes back to previous years.

Tosun (2017: 37) emphasizes that “there is constant belief in the public that we are not successful at learning and teaching foreign language.” That anxiety among the public also hinders the language learning and to teach and the learn the language it is crucial to overcome that problem too. Anxiety causes people to feel like they cannot learn the language and they will not be able to learn regardless of their efforts. This constant belief affects readiness of learners in a negative way.

Learners should believe in themselves and should concentrate so as to use the target language more effectively.

According to report of The Economic Policy Research Foundation of Turkey (2011) Turkey is in the forty third line among forty-four countries in Proficiency of English Index. It is behind of the countries like Chili, Nigeria, Saudi Arabia. Also, according to this report there is not sufficient investment per a student, less salary and less chances presented to teachers, course books are worse and there are less outside class activities than being expected in Turkey. It can be seen from this report that there are many problems which lead language learning deficiencies in language environment. Besides, there are other researches too showing the language learning problems in our curriculum.

Following principles are settled related to the foreign language programs in Official Newspaper-Resmî Gazete, (2006: n.26184).

- a) Programs at the primary and secondary education have to follow and complete each other.
- b) Educational materials and equipment are to be improved according to innovations, scientific and technological developments and to the needs of the country and environment.
- c) Programs of the compulsory and elective foreign language courses are to be put into effect after approval of the Council.
- d) Programs followed in the foreign language courses at schools and other institutions are prepared by the branch teachers and to be approved by the provincial (or sub-provincial) directorates of education.

Although many researchers in Turkey carried out studies to explain the problems and solutions, English language learning is still a dilemma. These problems result in underestimating the reasons for language learning. Many researchers have stated possible solutions to overcome the language learning problems by analyzing our foreign language curriculum.

According to Celebi (2006: 305) there is an ongoing problem considering to language teaching and learning and also some measures are to be taken to overcome these problems. Thus, learners may reach the level which has been aimed. Celebi (2006: 305) states some solutions for language learning and teaching. Table 3 below shows the possible solutions in language learning in Turkish curriculum

Table 3: Solutions for Language Learning Problems

I. Language is a skill that can be learned as the learners use it in their life thus curriculum should be designed parallel to this principle.
II. Curriculum and learning experience should be in accordance with the characteristics of the learners.
III. Language learning should be bilateral as understanding and teaching.
IV. There should be activities outside of the schools as they have to use target language in their daily life. Exposing the target language will affect cumulatively.
V. While planning the lesson both stimulants and reinforcers should be more in lessons.
VI. During the process of measurement we should avoid using multiple choice tests.
VII. Language teachers should be more planned in the process of language teaching.

It can be concluded that English holds the most important position among the foreign languages in Turkey. As a result of this situation, there is a raising interest to language environment in which there are some frustrations as the learners are not the level that they are expected. There should be more investments and effective time management to overcome language problems.

1.5.3. Limitations of The Study

In particular, it can be concluded that there has been a raising interest on phraseology among linguists recently and studies have been carried out mostly in the environment of L1 learners about the field of the phraseology. There are limited number of researches about phraseology in L2 learning. While learning and using phraseology in L2 environment, there is a common problem among the learners that they cannot use the formulaic language efficiently compared with L1 learners. Baker (2010: 34) states “L1 learners use the formulaic language more than L2 learners”. Previous studies about phraseology have not dealt with phraseological awareness of the Turkish EFL students enough thus there are no satisfactory researches which may supply us with a solution. The aim of the thesis is to provide a conceptual framework based on the tertiary level of students’ phraseological awareness and usage of phraseological language among the Turkish EFL learners are enlightened. Also, findings may help learners or teachers to concentrate on the subjects which have to be considered during the teaching and learning processes. This thesis provides an important opportunity to advance the understanding of awareness level of phraseology.

On the other hand, it is beyond the scope of this study to examine diagnostic studies which are needed to see the development of learners in a time. In the recent years trend has been raising up for qualitative research and Winter (2000: 14) explains that;

Qualitative research, arising out of the post-positivist rejection of a single, static or objective truth, has concerned itself with the meanings and personal experience of individuals, groups and sub-cultures. Reality is the main concern of the qualitative research whereas research findings are in the front line in quantitative research.

The reader should bear in mind that every source in current study is authentic and based on collection of students' writings and it was enriched by qualitative aspects. Both qualitative and quantitative methods were used in this investigation to shed a light to uncertain points. Reality is a significant factor and all of the sources in current study are authentic materials consisting of real-life situations.

1.6. Summary

Current research based on corpus-based study starts with the introductory part which enlightens the field of phraseology and its role in language learning and word combinations especially lexical bundles which is the main subject in my thesis are among the components of the phraseology. Starting with brief background of the phraseology, aims to examine phraseology in detailed way. Information was collected and presented to show the importance of phraseology.

Following to brief introduction of the phraseology, it continues with the features of the phraseology namely idiomaticity, fixedness, semantic unity, pattern grammar and frequency of co-occurrence so as to understand to important points of phraseology. Then it continues with highlighting phraseological tendency in a language. Especially place of the phraseological tendency in L2 learning is emphasized so as to obtain its position in L2 learning.

In the next step phraseology in Turkish concept and in language learning is widely explained so as to understand the place of the phraseology in L1 and L2 learning environments. Phraseology in language learning is discussed and it is explained that phraseology holds critical position in language environment thus learners have to be able to apply phraseology in their real life. However, in Turkish context there are very limited number of the studies have been carried out to show the importance of the phraseology for EFL learners of Turkey.

In the final part, limitations of the study are discussed to show the research more transparent by explaining that in recent years there has been an incline for qualitative research whereas this study seems to be quantitative. However authentic data is used in the process of study. Limitations of the study are characteristics of the methodology or the design which may affect findings in a negative way.

In current study, limitations may occur because of the constraints on my methodology. It is better in a research to minimize the limitations to make the study stronger. There is a process of naturalistic inquiry which seeks social phenomena in-depth to understand in natural setting. Apart from the main method which is based on quantitative authentic data, this report is qualitative as it focuses on outputs of L2 learners in natural setting.

Another limitation of the study is that, it can not be generalized to whole L2 learners in Turkish settings because there are mainly two corpora from two universities in Turkey and it these two corpora are not sufficient to be generalized to whole. However, general phraseological habits in two universities can be seen. My target Turkish EFL learners are learners studying at Karadeniz Technical Universty in Trabzon and Çukurova Universty in Adana. Although it could be better to get data from variety of universities, compiling a corpus is a process which may take years to collect authentic data. Thus, results may be generalized to EFL learners in Turkish seeting as Turkish corpora were compiled in years with many students.

1.7. Literature Review of Lexical Bundles

One of the most pragmatic components of phraseology is lexical bundles which are actively standing throughout the language. Lexical bundles dating back to 1999 in a *Longman Grammar of Spoken and Written English*, is relatively new term among the components of the phraseology as most of the other terms date back in 1950s. Biber and his friends (1999: 990) define bundles as “bundles of words that show a statistical tendency to co-occur” and “recurrent expressions regardless of their idiomaticity, and regardless of their structural status”. Lexical bundles can be regarded as extended collocations which consist of commonly used with two words together while lexical bundles consist when there are at least three words.

Lexical bundles were studied before and after late 1990s under the different names but later all of these terms were generalized as lexical bundles. This can be seen in the case of previous studies in which there are different labels such as; recurrent word combinations from the literature, n-grams, lexical bundles, prefabricated patterns, formulas, clusters, phrasal lexemes, prefabs or lexical phrases, sentence stems and formulaic sequences. Although all terms refer to the same meaning, different terms were used by different linguists. However, in recent years they are all named as lexical bundles. Huang (2014: 30) explains that;

Biber and his colleagues started their scholarly works on lexical bundles identified from both spoken and written English along with the Longman project in late 1990s. With their further phraseological studies on oral or written corpora from university discourse to academic genres, the lexical bundle approach to phraseology has been widely practiced to describe or analyze phraseological features such as distribution and function in a specific genre or context.

Upon being introduced to language society, lexical bundles have become a popular field for researchers. Cortes (2004) proposed a methodology aiming mainly to find out the usage of non-native students' lexical bundle approach. From the late 1990s to today there have been various studies carried out by many linguists in many cultures to show the power of the lexical bundles in a language.

1.8. What is Lexical Bundle?

Lexical bundles are at least three words combinations that mostly occur in a sentence or context and they became chunks in a language in time. Apart from Biber's definition of lexical bundles, there are some other definitions that explain lexical bundle and its character in terms of structure and semantics. Again, Biber and Barbieri (2007: 263) define lexical bundles as "important building blocks of discourse". Lexical bundles are not used to introduce different meaning like idiomaticity rather they stand as functional in written and spoken context. Wei and Lei (2011: 37) state that "Lexical bundles contrast with idioms, which are whole phrases with a meaning unrelated to the parts". Lexical bundles do not compose new term but they refer the first meaning of the words mostly. Lexical bundles are recurrent expressions without considering their idiomaticity and their structural status. They are sequences of word forms which go together in general in a natural discourse. This situation seems to be the biggest difference between lexical bundles and idiomaticity, clearly explaining that words come together but the first meaning does not change in lexical bundles while meaning may change in idiomaticity. For example; the word combinations "*shoot the breeze*" is completely an idiomatic word combination as it creates new meaning when combined. Thus, it cannot be argued that this combination is lexical bundle, but word combinations such as "*I would like to, I think that*" can be accepted as lexical bundles because of the fact that all words forming this combination keep their first meaning. Flexibility cannot be seen in lexical bundles compared to idiomaticity and collocations.

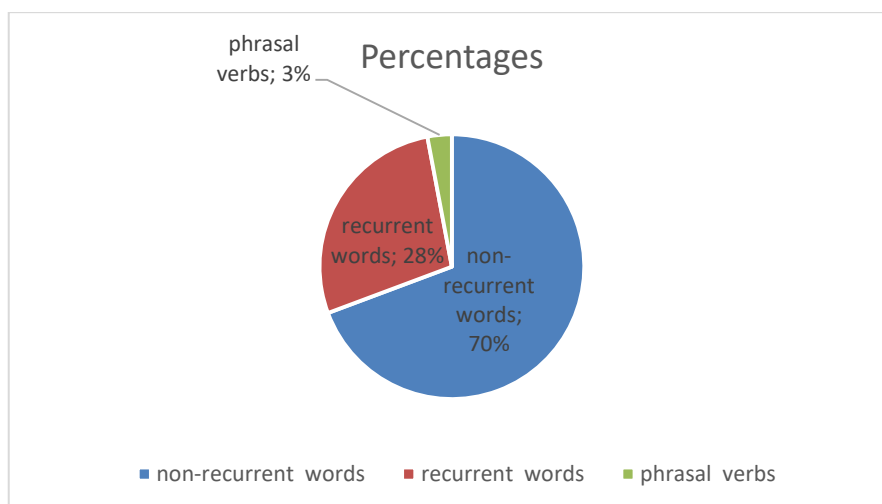
Biber (2006: 174) states that "lexical bundles are crucial for constructing a discourse in university register." Therefore, students at the level of university should use the lexical bundles as much as possible in their writing or speaking. To know large quantity of formulaic language enables students to use the language more natural way. If the use of lexical bundles or formulaic language is not enough or misused, then there could be problems for students to communicate in target language and they will have difficulty understanding discourse. Miller (2009: 13) states that "Conversely, misuse of formulaic language has been shown to be a potential source of communication difficulties". The more students have the ability to use the lexical bundles in a language, the more they are able to use the language in a fluent and natural way. It is not surprising that lexical bundles used in conversations and academic prose vary from each other which have been stated by many linguists. For example, one of the important linguists interested in bundles, Biber et al (1999) states that common lexical bundles include sequences in conversations such as: "*do you want me to, I said to him, going to be a, I don't know what*". On the other hand, quite different lexical bundles can be common in academic prose, containing sequences like: "*in the case of the, there was no significant, it should be noted that.*"

1.8.1. Lexical Bundles in Academic Prose and Conversation

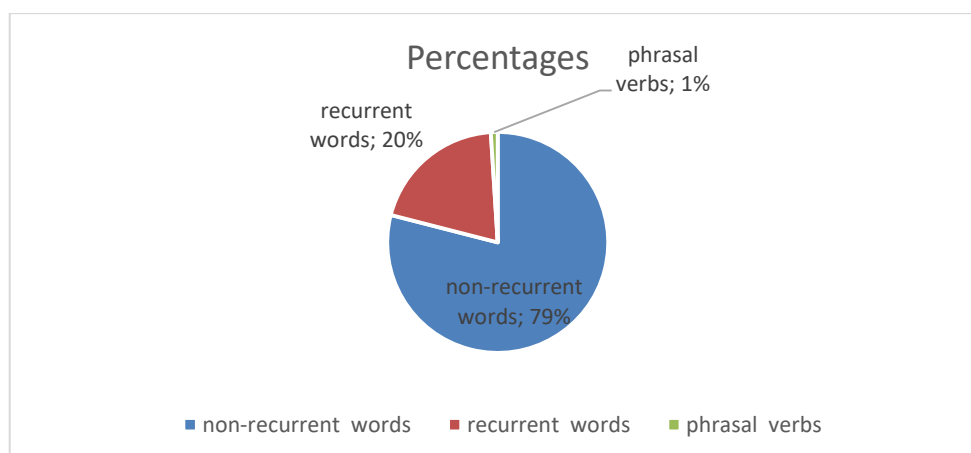
Idiomatic words are important both in academic and conversational fields. Lexical bundles are generally seen more in conversations rather than academic prose. According to Biber (1999: 989) “lexical bundles take crucial places in academic writing and conversations. Recurrent word combinations take nearly % 20 percent of the academic prose while it is % 28 percent in conversation”. It can be noted from this that large stocks of lexical bundles are common in both fields. Pawley and Sinder (1983: 191) argue that “fluent and idiomatic control of a language depends on the knowledge of lexicalized sentences that are units of clause length or longer whose grammatical form and lexical content are wholly or largely fixed.” By using large quantity of recurrent word combinations, language users are able to concentrate on other subjects of language considering that recurrent words can be applied in a language without much effort. Lexical bundles can be identified empirically as they can be found in a given register with the help of the corpus. Biber (2006: 174) explains that “lexical bundles are crucially important for the construction of discourse in all university degrees”. Therefore, academic learners need to know and use lexical bundles to create effective outputs.

A notable example of place of recurrent word combinations is given by Biber (1999) who gave the percentages of the recurrent and non-recurrent words in registers which can be main source for understanding the significance of the recurrent word combinations in a language. Graphic 2 below reveals that general percentage of recurrent word combinations in academic prose. Graphic 3 below shows the general percentage of the recurrent word combinations in conversations. These figures reveal that recurrent words are important components in conversations and prose.

Graphic 2: Percentage of Recurrent and Non- Recurrent Words in Conversation



Graphic 3: Percentages of Recurrent And Non-Recurrent Words in Academic Prose



The proportion of recurrent word combinations in both graphics indicates that recurrent word combinations occur frequently in a language though recurrent word combinations in conversations occur more than in academic prose. Bearing these graphics in mind it can be concluded that;

- Recurrent word combinations are at the average of %24 in a language regardless of looking where they are used. Thus, it can be concluded that recurrent word combinations hold considerable place in language.
- Recurrent word combinations are used more in conversations by holding a place up to %28 percent.
- Recurrent word combinations are used in academic prose %20 percentages and it can be accepted as high percentage for academic prose.
- Both in conversation and academic prose, lexical bundles may be accepted as a part of the discourse as they frequently occur.
- Besides, considering two-word circumstances in both places it can be noted that phraseology is the one of the key factors in language.

1.8.1.1. Lexical Bundles in Conversations

It is not easy to comprehend the structure of lexical bundles completely whilst they can be categorized based on their structural relations. Most common lexical bundles are three-word lexicals which occur most frequently among the lexical bundles and can also be the part of four-word lexical bundles and more. For example, one of the most common lexical bundles in prose is “one of the” which can be seen in four-word lexical bundles in the form of “one of the most” and which can be seen in 5 words in the form of “one of the most important”. In these examples three-word lexical bundle is the starter of four-word lexical bundles and five-word lexical bundles.

Lexical bundles are part of discourse in conversation and more likely to be used considering that they are recurrent combinations, which contributes the quality of speech or prose.

By way of illustration, Biber (1999: 1002) shows how bundles are divided into categories in conversations. According to him, there are mainly 14 categories of lexical bundles in conversations. It is not feasible to categorize three-word lexical bundles because of the fact that there are thousands of three-word bundles in languages. Thus, categorizations of bundles are formed without taking three-word bundles into consideration. Even though categories may be more than fourteen as Biber underlined, these fourteen categories are accepted as major ones. Table 4 shows categorization of bundles.

Table 4: Categorization of Lexical Bundles in Conversations

- | |
|---|
| <ul style="list-style-type: none">○ personal pronoun + lexical verb phrase○ pronoun / noun phrase + <i>be</i> +○ verb phrase with active verb○ <i>yes-no</i> question fragments○ <i>wh</i>-question fragments○ lexical bundles with <i>wh</i>-clauses○ lexical bundles with <i>to</i>-clauses○ verb + <i>that</i> - clause fragments○ adverbial clause fragments○ noun phrase expressions○ prepositional phrase expressions○ quantifier expressions○ other expressions○ meaningless sound bundles |
|---|

Most common lexical bundles above are written in bold and they occur more frequently in conversations. As major types of lexical bundles in conversation, these types of bundles are taken into consideration.

Kaisheng (2004: 3) states that “in addition to being fluent and effective in language, formulaic sequences have been shown to provide speakers with communicative skills”. Unless speakers improve times of using recurrent words combinations, communicative skills will not increase in such circumstances.

1.8.1.2. Lexical Bundles in Academic Prose

As it has been stated in previous parts of this research, lexical bundles hold significant place in a language by enriching the language with the use of the word combinations. It is notable that lexical bundles can be empirically showed and analyzed and as a result of this, language learners can be more effective in using the language. Lexical bundles are quite common in academic

registers and this makes lexical bundles all the more significant for academic prose. The importance of formulaicity in academic prose has been underlined by Biber (2009: 301);

formulaic language is very important in both conversational and written academic discourse, but it is realized in very different ways linguistically: fixed sequences that represent clause fragments in conversation, versus formulaic frames that consist of noun phrase and prepositional phrase fragments in academic writing.

Apart from Biber, Romer (2010: 26) underlines the importance of lexical bundles stating that “it highlights the importance of knowledge about modification of common multi-word expressions, their functions and positions in a text for non-native writers”. Hyland (2008: 60) made a research about lexical bundles by analysing doctoral and master theses, which resulted in finding out that lexical bundles were frequently used both in master and doctoral theses though they were preferred more in master theses.

Lexical bundles in academic prose are divided into twelve main categories though it could be more. As three-word combinations cannot be categorized because of the fact that they occur in a language frequently and there are thousands of them, four-word lexical bundles, five-word lexical bundles and six-word lexical bundles are categorized. Biber (1999: 1002) categorizes lexical bundles in academic prose in Table 5;

Table 5: Categorizations of Lexical Bundles in Academic Prose

- | |
|--|
| <ul style="list-style-type: none">○ noun phrase with of- phrase fragment○ noun phrase with other post modifier fragments○ prepositional phrase with embedded of- phrase fragment○ other prepositional phrase fragment○ anticipatory it + verb phrase / adjective phrase○ passive verb + prepositional phrase fragment○ copula be + noun phrase / adjective phrase○ (verb phrase +) that - clause fragment○ (verb adjective +) to- clause fragment○ adverbial clause fragment○ pronoun / noun phrase + be (+ ...)○ other expressions |
|--|

Especially *noun phrase with of- phrase fragment* is quite common in academic writing and the list is too long. Lexical bundles in this category cover a wide range of meaning and functions. The functions of them is fewer but important for discourse in register. All of those categories have also sub-categories which are recognized by looking up to meaning and function, etc. Among those categories, the first four category in the table above are used more in academic prose.

1.9. Classification of Lexical Bundles

It is noted in previous part that it is not feasible to categorize all bundles but it may be possible that lexical bundles can be divided into two main categories namely functional classification and structural classification. Hyland (2008: 13-14) states that, “lexical bundles are too broad to classify though there are two main categories standing over.” A similar statement was made by Conrad (2004: 60) who underlined that “there are mainly two categories; structural and functional classification” According to Conrad in (2004: 60)

The lexical bundles are classified in two major ways. First, we consider the structural characteristics of the bundles. Although most of the bundles are not complete structural units, they do fall into groups with certain structural associations. The second type of classification presented in this paper is a preliminary classification of the bundles by their function in a discourse context. No a priori categories were assumed. Instead, we examined each bundle in concordance listings and made interpretations of its function. We placed bundles into groups unified by similar discourse functions.

1.9.1. Structural Classification of Lexical Bundles

Great deal of the lexical words are not structural units but part of the phrases or fragments of the embedded units. Yet, structural ones have strong impact on grammatical correlations. Biber (1999: 993) states that “there are only % 15 of the lexical bundles in conversation and % 5 of the lexical bundles in academic prose as they combine two different words each other. In other words, the last word of the lexical item is the first word for the following”. Even though the percentage of the structural words are not much, structural ones are correlative. These correlative lexical words may differ in different registers. One of the other features of the structural words is that they can be the part of other lexical bundles. For instance, three word-lexical bundles can be four or five with the addition of another word to the end. Biber et al (1999: 997) underlines that “bundles in conversation are most commonly clausal, of the type *pronoun + verb + complement* (e.g., *I want you to, it's going to be*), but in academic prose, most of the lexical bundles are phrasal and can be parts of *prepositional or noun phrases*”. List of the structural lexical bundles has been given in the parts “Lexical Bundles in Conversation and Academic Prose”.

1.9.2. Functional Classification of Lexical Bundles

Biber (2004), Conrad (2004), Cortes (2002), Hyland (2008) attempted to classify lexical bundles based on their characters and nature. Later developments of these categorizations have been carried out and those frameworks revealed three categories, namely stance expressions, discourse organizers and referential expressions.

Biber (2004: 384) defines these three categories as;

Stance bundles express attitudes or assessments of certainty that frame some other proposition. Discourse organizers reflect relationships between prior and coming discourse. Referential bundles 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.

Table 6 shows the Biber's functional taxonomy (2004) which has been developed by some other linguists like Conrad, Cortes, Pang.

Table 6: Functional Classification of Lexical Bundles

Stance Expressions	Discourse organizers	Referential Bundles
Express attitudes or assessments of certainty that frame some other proposition	Reflect relationships between prior and coming discourse	Make direct reference to physical or abstract entities, or to the textual context itself
A. Epistemic stance I don't know if, I think it was, are more likely to, the fact that the	A. Topic introduction/focus what do you think, if you look at	A. Identification/focus that's one of the, of the things that
B. Attitudinal/modality stance B1) Desire if you want to, I don't want to B2) Obligation/directive you might want to, it is important to B3) Intention/prediction I'm not going to, it's going to be B4) Ability to be able to, can be used to	B. Topic elaboration/clarification I mean you know, on the other hand	B. Imprecision or something like that, and stuff like that C. Specification of attributes C1) Quantity specification there's a lot of, how many of you C2) Tangible framing attributes the size of the, in the form of C3) Intangible framing attributes the nature of the, in the case of D. Time/place/text reference D1) Place reference in the United States D2) Time reference at the same time, at the time of D3) Text deixis shown in figure N, as shown in figure D4) Multifunctional reference the end of the, the beginning of the

In academic writing, Hyland (2008: 13) developed a functional classification of lexical bundles. Functional classifications can vary from one linguist to another as there is not clear categorization accepted by linguists. Thus, every linguist, designed categories according to their studies. Hyland (2005) who is a famous linguist shaped his classifications of lexical bundles based on his research throughout the years. Table 7 below shows Hyland's classification of the lexical bundles.

Table 7: Hyland’s Classification of Lexical Bundles in Academic Writing

Research-oriented bundles	Text-oriented bundles	Participant-oriented bundles
Help writers to structure their activities and experiences of the real world	Concerned with the organization of the text and its meaning as a message or argument	Focused on the writer or reader of the text
<p>Location</p> <p>Indicating time/place <i>at the beginning of, at the same time, in the present study</i></p> <p>Procedure bundles</p> <p><i>the use of the, the role of the, the purpose of the, the operation of the</i></p> <p>Quantification</p> <p><i>the magnitude of the, a wide range of, one of the most</i></p> <p>Description</p> <p><i>the structure of the, the size of the, the surface of the</i></p> <p>Topic</p> <p>related to the field of research <i>in the Hong Kong, the currency board system</i></p>	<p>Transition signals</p> <p>Establishing additive or contrastive links between elements <i>on the other hand, in addition to the, in contrast to the</i></p> <p>Resultative signals</p> <p>Mark inferential or causative relations between elements <i>as a result of, it was found that, these results suggest that</i></p> <p>Structuring signals</p> <p>Text-reflexive markers which organize stretches of discourse or direct the reader elsewhere in text <i>in the present study, in the next section, as shown in figure</i></p> <p>Framing signals</p> <p>Situate arguments by specifying limiting conditions <i>in the case of, with respect to the, on the basis of, in the presence of, with the exception of</i></p>	<p>Stance features</p> <p>Convey the writer’s attitudes and evaluations <i>are likely to be, may be due to, it is possible that</i></p> <p>Engagement features</p> <p>Address readers directly <i>it should be noted that, as can be seen</i></p>

Table 7 reveals that lexical bundles being various kinds of linguistic forms, can vary a lot. Lexical bundles have been studied so far by many linguists who have had great contribution in classifying lexical bundles and revealing where they differ in registers. As being part of the discourse, lexical bundles can differ from one register to other one depending on what has been the focus of study. Halliday (1994: 84) stated that “Hyland’s typology is quite easier for researcher to analyze and distinguish three groups of lexical bundles which is based on Halliday’s theory of linguistic macro-functions”.

Besides, Altenberg (1998) analyzed recurrent word combinations, where they have been defined as “three-word combinations” and have been found at least ten times more in a corpus that Altenberg analyzed. Upon analyzing the three-word combinations, Altenberg (1998) revealed his lexical bundles taxonomy which was divided into mainly three groups and many sub-groups. Table 8 shows Altenberg’s classification of lexical bundles under the names of main categories. Many sub-categories are not shown in the table to prevent misunderstanding of lexical bundles.

Table 8: Lexical Bundles Typology of Altenberg

Types and sub-types	Examples
1. Full clauses 1.a Independent clauses 1.b Dependant clauses	I don't know, thank you very much as it were, as you wish, do you know that, there is a more or less, and so on
2. Clause constituents 2.a Multiple clause constituents 2.b Single clause constituents	as long as, a great deal of
3. Incomplete phrases	

Altenberg's categorization of lexical bundles is mainly based on the spoken corpus, London-Lund Corpus of Spoken English.

1.9.3. Types of Lexical Bundles

Plenty of lexical bundles can be seen in spoken and written language and consist of nearly the %25 of the language. Lexical bundles are divided into mainly four groups namely; three-word bundles, four-word bundles, five-word bundles and six-word bundles although there are seven-word and more lexical bundles. Seven-word and more word lexical bundles are quite less than the main ones. Thus, in this research mainly three-word, four-word and five-word and six-word combinations were analyzed. Also, two-word combinations are not accepted as lexical bundles, though they have features of lexical bundles. According to Biber (1999: 993)

three-word can be considered as kind of extended collocational association, thus they are extremely common. On the other hand, four-word bundles, five-word bundles and six-word bundles are more phrasal in nature and correspondingly less common

As it is stated by Biber, it is common that three-word combinations are mostly associations and they are more common in academic prose and conversation. There is no complete structure of any types of lexical bundles in academic prose and they are mostly combination of the structures. Especially four-word and more types of lexical bundles are mostly preferred in academic prose when compared to those in conversations.

1.9.4. Target Bundles in Current Research

Current research mainly concentrates on the lexical performance of the tertiary level Turkish EFL students' in writing. Structures of the lexical bundles in conversation are not included in this study. However, based on the Biber's and Hyland's classification, lexical bundles and structures of lexical bundles in writing are taken into consideration accordingly. It was a time consuming process to analyze and compare all categorization of the lexical bundles in academic prose, as there

were plenty of lexical bundles and classifications. Thus, in this research, most common structures of lexical bundles and most common types of lexical bundles were used. Apart from the most common bundles in our reference corpora BAWE and LOCNESS, common bundles which also appeared in the studies of Biber, Salazar, and Davis were analyzed.

Biber (1999) states that structures of the lexical bundles vary a lot in academic prose and conversation by stating that over %60 of the lexical bundles used in academic prose are parts of noun phrase and prepositional phrase. Based on Biber’s analysis in current research prepositional and noun phrases were also analyzed so as to see the performance of Turkish EFL students’ in writing.

Another important point is that as most common types of the lexical bundles are three-word, four-word, five-word and six-word bundles in language and the rest is much less compared to these four types. As a result, the current research analyzes these main four types of bundles. Just like the same procedure above, in this research most common ten lexical bundles for each category are analyzed. While choosing samples of the lexical bundles, Hyland’s and Biber’s (2008) list for lexical bundles were used. Table 9 below shows the lexical bundles used in current research based on the findings above. Structural types and sub-types of lexical bundles are shown in Table 9, and some examples are given.

Table 9: Lexical Bundles Taxonomy in Current Research

Structural types and sub-types	Samples
<p>1 Noun Phrase with <i>of</i>-phrase fragment</p> <p>1.a Bundles used for physical description, including identification of place, size etc.</p> <p>1.b Bundles marking simple existence or presence</p> <p>1.c Bundles identifying variety of abstract qualities</p> <p>1.d Bundles describes processes or events lasting over a period</p> <p>2 Noun phrase with other post-modifier fragments</p> <p>2.a Bundles describe how a process occur</p> <p>2.b Bundles identify relations amon entities</p>	<p>the surface of the, the position of the, the size of the the existence of, the presence of the use of the, the nature of the, the value of the the development of an, the course of the</p> <p>the way in which, the extend to which the the relationship between the, the difference between the,</p>
<p>3 Prepositional phrase with embedded <i>of</i>-phrase fragment</p> <p>3.a Bundles mark abstract logical relations</p> <p>3.b Bundles mark temporal situations</p> <p>3.c Bundles have specialised function identifying time periods or process</p> <p>4 Other propositional fragmnets</p> <p>4.a Bundles identify particular time or period</p> <p>4.b Bundles have idiomatic meaning</p>	<p>as a result of, in the absence of, in the case of at the end of, at the time of, at the middle of in the development of, in the process of</p> <p>in the present study, in the next chapter, on the other hand, at the same time</p>

In Table 9, three-word lexical bundles do not occur as the categorization of them is not clear because of the fact that they are much common in academic prose and they are mostly collocations. The structural types and sub-types are listed in above and these structures are analyzed in current research for the reason that they are the most frequent structures that exist most in academic prose. Erman & Warren, (2000: 103) underlines that “lexical bundles were found to constitute 52.3% of the written discourse in registers thus current study again underlines the importance of lexical bundles in academic writing.”

While choosing three-word combinations, most frequent three-word combinations were chosen for the current study. These three-word combinations were chosen among the most frequent three-word combinations in academic prose after considering their frequencies and log likelihood scores. LL standing for Log Likelihood score can be shortly defined as scores that express the extend to which words co-occur compared the number of times and are the words used close to each other in corpora. They are used seperately and frequency can affect LL scores in a big portion whether it is low or high. Although it may seem difficult to work at the first sight, it is easy to calculate and find out the LL score. LL scores are calculated in the website <http://ucrel.lancs.ac.uk/llwizard.html>. LL score is explained in the next chapter in a detail.

Based on Hyland’s and Biber’s categorization; ten lexical bundles are chosen for each category among four-word and more lexical bundles so as to see and compare the use of lexical bundles among Turkish EFL students. In total, most common forty lexical bundles are analyzed in corpus among four-word and more lexical bundles, also twenty-three most common three-word lexical bundles based on the list of Davis (2012) and Salazar (2011) are chosen in current research for analysis. While choosing target lexical bundles we make use of the research of Biber in a large corpus and we use them in our study. On the other hand, we also choose target three-word bundles based on the studies carried out before as there is no a clear criteria to evaluate them otherwise. Target four-word and more lexical bundles are shown in Figure 2.8 below which is based on the study of Biber. Biber (1999: 995) studied on a large corpus and he found out mostly used four-word and more lexical bundles. In Table 10 bundles are choosen randomly.

Table 10: Target Four-Word and More Bundles of Current Research

Noun phrase with of-phrase fragment	Noun phrase with other post-modifier fragments	Prepositional phrase with embedded of-phrase fragment	Other prepositional phrase (fragment)
the beginning of the	the fact that the	as a result of	for the first time
the end of the	the relationship between the	at the end of	as in the case
one of the most	the ways in which	in terms of the	in addition to the
the total number of	such a way that	as a result of the	of the most important

Table 10: (Continue)

the development of the	the difference between the	in the area of the	in such a way that
the work of the	the way in which the	by the end of the	on the other hand
the purpose of the the aim of this study one of the most of the first part of the	an important role in the same way as such a way as to the fact that it	from the point of view of in a number of ways at the time of the in the study of	with respect to the between the two groups in the present study to the fact that

As shown in Table 10, there are ten lexical bundles chosen randomly for each category and their frequencies are not mentioned in the study of the Biber. However, these bundles are among the most common lexical bundles in academic prose in Biber's list.

Being relatively new topic for linguist, it may be difficult to choose appropriate bundles as there are schools of types and structures. Apart from this situation, three-word lexical bundles are not mentioned in detail and they cannot be classified because of the fact that they mostly have idiomatic relations in between. However, in the present study, it is a must to see three-word recurrent combination-performance of tertiary level Turkish EFL students. Thus, we choose top 10 three-word combinations from the literature for further analysis. Davis (2012: 68) had a research and listed most common three-word recurrent combinations in his study. He made a research among the graduate and undergraduate students prose via the Corpus AUGER which is pretty large corpus to find out three-word recurrent words. AUGER corpus, consisting of more than 2.3 million tokens is large enough to see the preferences of the students. When at least one million words corpus is considered as sufficient for a research, Davis' research containing three-word recurrent words can be acceptable for current research.

According to that study most common three-word bundles are shown in Table 11 below. They were revised by combining the preference of undergraduate and graduate students' academic writings. Proper nouns in this research are omitted in the current research.

Table 11: Davis' Study of Most Common Three-Word Lexical Bundles

Most common three-word bundles	Frequency
as well as	674
one of the	639
the number of	588
part of the	427
in terms of	358
the fact that	228

Table 11: (Continue)

as a result	210
some of the	228
the case of	210
on the other	210

As shown in Table, 11 first three recurrent words are much more common than the rest and this shows us that especially these three recurrent word combinations are preferred by the native students in their academic prose.

Another study done by Salazar (2011: 77) shows the most common lexical bundles in another corpus and Salazar found out most common fifty recurrent word combinations in which most of them are three-word recurrent combinations. In that study most common three-word combinations were chosen for the current research. Three-word bundles are more frequently seen than other types of lexical bundles, and three-word bundles cannot be categorized as the number of this bundles is high. Thus researchers focus on the frequency of three-word bundles. These bundles Salazar (2011: 77) emphasizes the list by saying that;

as can be expected, the list is largely composed of three-word strings, which account for %83 or 640 of 769 target bundles. They are followed by 113 four-word bundles, which equal %15 of the total. The list is rounded out by the much rarer five-word and six-word bundles, both of which represent just %1 of all target bundles with just eleven bundles respectively

It can be seen in Table 12 that most common three-word lexical bundles are given from the most frequent to the least.

Table 12: Salazar List of Most Common Three-Word Bundles In Academic Prose

Three-word bundle	Frequency
the presence of	906
data not shown	625
the absence of	481
<u>as well as</u>	307
<u>the number of</u>	273
the effect of	259
the ability of	237
as described in	227
shown in figure	216
been shown to	209
the addition of	203
is required for	194
was used to	190
in response to	189
a number of	183

As shown in Table 12, there are most common fifteen three-word combinations and the underlined three-word combinations are almost the same with Davis' list of most common three-word combinations shown in Table 1 above.

While choosing our target three-word bundles for the current research, the researcher made use of both Salazar's and Davis' list of three-word bundles by combining them. Ten recurrent combinations from the list of Davis and fifteen recurrent combination from Salazar were grouped together. In total, there were twenty-five three-word bundles, and yet two of them were the same.

So far, it can be seen that the way target bundles in this current research were chosen is given above. Most common lexical bundles were derived from the lists prepared by Biber, Hyland, Salazar and Davis. The main concern was to investigate the phraseological awareness of the tertiary level Turkish EFL students in native and non-native written corpora.

In total, there are 63 target bundles and all target bundles were determined from the various related studies on native students' writing performance. However, there are also studies showing the performance of lexical bundles and as we compare our findings with those of the native corpora result it is seen that target lexical bundles are chosen from the most common lexical bundles among the English native students' corpora. Our main interest in this study is to see the lexical bundles performance of the EFL tertiary level students in the field of writing skills and it can be found by comparing it with the native students' lexical performance in the field of writing.

1.10. Performance of Tertiary Level Students on Lexical Bundles

Being relatively a hot topic in the field of linguistic, it is not surprising to expect linguistic studies about lexical bundles which came to stage of linguistics scene and took the attention of the linguists dramatically. Especially in the few past decades, there have been studies aiming to see the position of the lexical bundles in a language and naturally in language teaching and parallel to this situation there are already existing researches on lexical bundles. However, studies on lexical bundles in Turkish context are not many. Although lexical bundles are both crucial for L1 and L2 students, the absence of research stands as a barrier for learners.

1.10.1. Earlier Research on Lexical Bundles

Following to increasing trend to the phraseology field, it is not uncommon to estimate that components of the phraseology are specifically analyzed too. Biber and his friends (1999) made great contribution to the lexical bundles area by publishing a book named Longman Spoken and Written Corpus. They distinguished lexical bundles from the other components of the phraseology and it shows us not only to what extend lexical bundles exist but also their roles in language

teaching. Lexical bundles were given different names such as recurrent word formations, n-grams, a multiword lexical chunk, formulaic sequences and lexical phrase. Biber (1999) categorized lexical bundles differently in conversation and academic prose.

Considering the Biber's categorization in 2008, Hyland also categorized lexical bundles after analyzing a large corpus. Although their categorization of lexical bundles are different, both of them strongly state that categorising the lexical bundles is pretty difficult as there is no limit for the variety of lexical bundles. Both categorizations are the fundamental categories researched by many linguists. Apart from these linguists, Altenberg (1998) studied on recurrent word combinations and categorized them in groups based on a spoken corpus. Being a new term in the literature, studies have been conducted recently about lexical bundles. Table 13 shows studies about lexical bundles. Bal (2010) lists the major studies in the table below.

Table 13: Major Studies on Lexical Bundles

Author	Year	Corpus	Corpus Size
Biber, Johansson, Leech, Conrad and Finegan	1999	LSWE Corpus	Over 40,000,000
Cortes	2002	Native Freshman compositions(311 papers)	360,704
Cortes	2004	Published writings and students writings	Published writings;1,992,531 Student;904,376
Biber, Conrada and Cortes	2004	T2K SWAL Corpus	2,009,400
Scott and 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

CHAPTER TWO

2. METHODOLOGY

This chapter mainly concentrates on a corpus analysis of lexical bundles, introduces the research method, the data and the analytical framework of the current study. The present study aims to make a corpus-based analysis of lexical bundles in argumentative and expository writings of tertiary level EFL students. It starts with how corpus based contrastive interlanguage analysis is designed and then it is followed by the introduction of the corpora used in the current study. Following this part, some general information about corpora is given in tables to show statistical measurements and tools. All corpora are introduced in a detailed way to show the framework of the current study. In the following part, the research design, which consist of three parts, namely corpus, data collection instruments and procedures are introduced.

According to Oxford Dictionary (2006) “a corpus which can be defined as collection of written or spoken language texts in electronic form, provides us with the evidence of how language is used in real situations”. In other words, corpus can be considered as a store of authentic texts written or spoken in a computer system. Corpus based analysis in the current study analysed four different corpora.

The aim of this study is to make a corpus analysis of the lexical bundles as well as investigating the usage patterns of learners of English in a tertiary level EFL setting in Turkey. Four different corpora are used in current study so as to provide evidence for the phraseological awareness of tertiary level language learners. Two of the corpora consist of expository argumentation and two of the corpora consist of academic argumentative texts. Karadeniz Technical University Corpus of Learner English (KTUCALE), British Academic Written Corpus (BAWE), Turkish International Corpus of Learner English TICLE and The Louvain Corpus of Native English Essays are the corpora analyzed in the current study.

Consisting of 819846 words, KTUCALE is a non-native corpus compiled by Karadeniz Technical University and it is the second largest one among the four corpora. It consists of academic writings of tertiary level prep students. BAWE is another corpus which is analyzed in the current study and consists of 624294 words. Compiled from the academic writings of British learners, it is the largest corpus. LOCNESS is a corpus compiled from native English learners' essays. Consisting of 361054 words, it is compiled from expository writings of learners. Finally,

Turkish International Corpus of Learner English (TICLE) which consists of 223449 words is compiled from the expository writings of non-native learners. Table 14 below shows the categorization of corpora.

Table 14: Categorization of Corpora

Corpus	Tokens	Texts	Native/ Non-Native	Expository/ Academic
KTUCALE	819846	220	Non-Native	Academic argumentation
BAWE	624294	223	Native	Academic argumentation
TICLE	223449	287	Non- Native	Expository argumentation
LOCNESS	361054	372	Native	Expository argumentation

Biber (1993: 243) states that;

The use of computer-based corpora provides a solid empirical foundation for general purpose language tools and descriptions, and enables analyses of a scope not otherwise possible. However, a corpus must be 'representative' in order to be appropriately used as the basis for generalizations concerning a language as a whole; for example, corpus-based dictionaries, grammars, and general part-of-speech taggers are applications requiring a representative basis.

These four corpora are representative of learners' general use of writing skills in different cultures. According to Tversky and Kahneman (1974) "representativeness bias is mental shortcut and is defined as the tendency to irrationally attribute one characteristic to imply another". Representativeness of these corpora help us to generalize the results to the whole population of learners. Again Shefrin (2001) defined representativeness heuristics as "relying on stereotypes that are used to form quick but irrational opinions".

2.1. Frequency Based Analysis

The advent of the computer technologies affected linguistic fields and studies have been carried out since then. Using authentic texts which were collections of tertiary level learners' essays, is the main interest of this study. Lexical bundle approach has become basically a frequency-based analysis of phraseology while intuition-based analysis was common before the advent of the computer technology. With the help of the corpora and tools, frequency-based analysis is able to analyze millions of words and show us the results. Granger & Paquot, (2008); Nesselhauf, (2005); Paquot & Granger, (2012) state that;

Two most distinct approaches typically recognized are the "phraseological approach," which focuses on establishing the semantic relationship between two (or more) words and the degree of non-compositionality of their meaning, and the "distributional" or "frequency-based" approach, which draws on quantitative evidence about word co-occurrence in corpora.

Altenberg & Granger, (2002, 4) underlines that “although the reliability of frequency studies was questioned from a relatively early stage, this did not put an end to them but, instead, merely prompted corpus linguists to gather bigger and more tightly controlled corpora”. Frequency-based analysis contributed to see how the learners have used lexical bundles in texts and enabled comparisons.

Frequency-based analysis is one of the most important components of the corpus analysis of the current study, which is based on the writings of learners. Thus, frequencies of each lexical bundle may show us the similarities and differences among those corpora. Study design of the current research, which mainly consist of four steps, is shown in Table 15 below.

Table 15: Study Design of Current Research

Step 1	Target corpora, in which there are two native learners corpora and two non-native learners corpora, were chosen.
Step 2	Data Collection: Criteria for the selected samples of target lexical bundles were stated and bundles were listed from Sketch Engine.
Step 3	Data Collection: Most common bundles in all corpora were also listed from Sketch Engine apart from the selected lexical bundles.
Step 4	Data Analysis: Findings were discussed and analyzed to show differences and similarities.

2.2. Corpus-Based Approaches

In the field of linguistics, corpus-based approaches are the main ones that contribute to linguistic studies in a way that more reliable and faster data has been obtained. Corpus linguistics enables researchers to make empirical investigations which have more validity and generalizability. Biber, Conrad and Reppen (1998: 4) underlines some characteristics of corpus-based analysis as;

- they are empirical, analyzing the actual patterns of use in natural texts;
- they are based on analysis of a large and principled collection of natural texts, known as a ‘corpus’; the corpus is evaluated for the extent to which it represents a target domain of language use;
- they make extensive use of computers for analysis, employing both automatic and interactive techniques;
- they depend on both quantitative and qualitative analytical techniques.

The use of computer programs provide researchers with easiness and provides the means for the detailed analysis of the words which would be difficult to analyze and categorize otherwise. With one click, it is possible to reach the list of target words among the thousands of texts. Corpus-based analysis is not merely a quantitative one. Qualitative interpretations are also needed together with quantitative data, which means that it is critical to complement quantitative data with

qualitative outcomes. Corpus-based analysis was used for grammar researchers in the beginning but it spread to include other fields in time, and it has been used in the field of phraseology more recently. Meyer (2004: 1) states that;

When the first computer corpus, the Brown Corpus, was being created in the early 1960s, generative grammar dominated linguistics, and there was little tolerance for approaches to linguistic study that did not adhere to what generative grammarians deemed acceptable linguistic practice.

Some linguists who hold the position that the only legitimate way of analyzing linguistic is “intuition” were against the idea that corpus-approach was legitimate in linguistic at first. Although this idea is still accepted by some linguists, corpus-approach is commonly used in linguistic field now. This research is mainly based on a corpus-based approach and is trying to find out EFL learners’ lexical bundles performance through contrastive interlanguage analysis.

2.3. Contrastive Interlanguage Analysis

Granger (1996: 295) defines contrastive interlanguage analysis as “a methodology which involves comparing learner data with native speaker data (L2 vs. L1) or comparing different types of learner data (L2 vs. L2)”. In parallel with most of the CIA studies, the current study aims to compare L2 learners with L1 ones in order to see the learners’ development in target norms. In CIA model, two non-native Turkish corpora (KTUCALE, TICLE) were used and compared to two native corpora (BAWE, LOCNESS) to create an awareness of EFL learners in terms of using lexical bundles. Huang (2014: 76) underlines the importance of CIA, stating that it provides some meritorious information for learner differences.

Lado was the first linguist who brought the main idea of contrastive analysis. Lado (1957: 1) stated that “in the comparison between native and foreign language lies the key to ease or difficulty in foreign language teaching”. While comparing the languages, negative and the positive sides of the learners may come out. Thus, it provides a chance to develop the language skills of learners. CIA model, which is also in the centre of the current study, shows us the deficiencies of tertiary level EFL learners’ performances in their written texts. There will not be a clear result if the compared languages and cultures are alike. Kohn (1986: 21) observes that “*transfer* is one of the major factors which shapes learners’ interlanguage performance and competence.” It can be concluded from this quotation that transfer mistakes from the mother tongue have an impact on L2 and this situation may be the reason for misuses in the target language. Also, Selinker (1979: 287) observes that language transfer studies can easily be carried out in Contrastive Analysis.

To summarize, it can be argued that Contrastive Interlanguage Analysis is a must for the current study as it aims to see the usage patterns of lexical bundles by tertiary level EFL learners

compared to the native speakers' performances. CIA contributes to linguists as it shows the deficiencies of learners. CIA can be considered as one of the methods which gives the opportunity to distinguish linguistic structures and patterns in target corpora.

2.4. Target Corpora, KTUCALE and TICLE

Target corpora in the current study are KTUCALE, which stands for Karadeniz Technical University Corpus of Academic Learner English, and TICLE, which stands for Turkish International Corpus of Learner English, and they were compiled by the writings of tertiary level EFL learners. KTUCALE includes type of academic argumentation texts of learners while TICLE includes expository argumentation texts. KTUCALE was compiled in Karadeniz Technical University while TICLE was compiled in Çukurova, Mustafa Kemal and Mersin Universities. First, the academic essays written by tertiary level learners were retyped on the computer and converted into text file documents. Following this process, written texts of L2 learners were uploaded on the Sketch Engine concordance software which is the ultimate corpus tool to search texts and it is available to users by registering online.

In KTUCALE, topics are mostly about education and in TICLE, there are many different topics and there are plenty of topics consisting of the learners' preferences which means that those texts are written by the learners based on their preferences. Linguistics, ELT and English Literature are the main categories in KTUCALE and they are all academic essays. However, the essays are not academic but expository in TICLE.

Özbay (2015: 86) states that; "Language variables can be summarized under five main titles, these being, medium of language, genre, topic, technicality and task setting. Learner variables, on the other hand, include age, sex, mother tongue, region, other foreign languages, level, learning context and practical experience."

While choosing target and reference corpora, language and learner variables were paid attention and it is the reason for choosing corpora. Also, Özbay (2015: 80) underlines that "KTUCALE was built following the criteria of design suggested in Barnbrook (1996), Hunston (2002), and Sinclair (1991)". In other words, it can be concluded that KTUCALE was compiled in collaboration with the criteria needed for corpus.

Table 16, which was prepared by Özbay, (2015) shows the design of KTUCALE in details.

Table 16: Design Criteria of KTUCALE

		Level A2-B1-B2		Sub-Fields	
Medium	Written	220	% 100		
	Spoken				
Genre	Expository argumentation				
	Academic argumentation	220	% 100		
Topic	Arts and Humanity	220	% 100	ELT	74-33%
	Life sciences			Applied L.	73-33%
	Social sciences			English Lit	73%33
Technicality	EFL Academic Essays	220	% 100		
Task setting	Timed essays	103	%46.8		
	Untimed essays (assignment)	117	%53.2		
	Reference	220	% 100		
	Exam papers	103	%46,8		

KTUCALE consists of mainly such topics as smoking, animal testing, social life, family and schools and all of the essays are academic argumentations of learners. Table 17 shows the distribution of the essays based on their fields.

Table 17: Distribution of Topics in KTUCALE

Topics	Number of Essays
Linguistics	81
Literature	68
English Language Teaching	71
Total	220

2.5. Reference Corpora: BAWE and LOCNESS

Developed at the Universities of Warrick, Oxford Brookes and Reading under the directorship of Sheena Gardner and Hilary Nesi, BAWE supplies us with the data in this study. BAWE, consisting of academic argumentation texts, is the reference corpus for KTUCALE. It was compiled in 2008 with the partnership of several universities and contains nearly 6.5m. words. However, BAWE is too large compared to KTUCALE, so the linguistic part of BAWE (BAWE-LING, which contains 624294 words), was selected as one of the reference corpora of the current study. Design of the BAWE is shown in table 18 below.

Table 18: Design of BAWE Corpus

		Level 1	Level 2	Level 3	Level 4	Level 5
Arts and Humanities (AH) Archaeology; Classics; Comparative American Studies; English; History; Linguistics / English Language Studies; Philosophy; others	Assignments	239	228	160	78	705
	Texts	255	229	160	80	724
	Words	468,353	583,617	427,942	234,206	1,714,118
	Assignments	180,188	193,206	113,120	197,205	683,719
Life Sciences (LS) Agriculture; Biological Science; Food Science; Health; Medicine; Psychology	Text words	299,370	408,070	263,668	441,283	1,412,391
Physical Sciences (PS) Architecture; Chemistry; Computer Science; Cybernetics/ Electronic Engineering; Engineering; Mathematics; Meteorology; Physics; Planning	Assignments	181	149	156	110	596
	Texts	181	154	156	133	624
	Words	300,989	314,331	426,431	339,605	1,381,356
Social Sciences (SS) Anthropology; Business; Economics; Hospitality, Leisure and Tourism; Management; Law; Politics; Publishing; Sociology	Assignments	207	197	166	207	777
	Texts	216	198	170	207	491
	Words	371,473	475,668	447,950	704,039	1,999,130
Total students						
Total assignments		333	302	235	169	1039
Total texts		840	787	625	625	2858
Total words		1,440,185	1,781,686	1,565,991	1,719,133	6,506,995

Retrieved from: <http://www.coventry.ac.uk/>

LOCNESS, which is another reference corpus that supplies data for TICLE, is a compilation of expository essays and was compiled between 1991 and 1995. It was collected from expository writings of British and American University students and A level learners. Topics in LOCNESS mostly vary; there is not a dominant specific topic but there are many different topics in an expository argumentation style. Table 19 below shows the design of LOCNESS.

Table 19: Design of LOCNESS

Topic	Field	Essays	Words	University
French intellectual tradition	Literary	39	59,568	British Universities
French society and institutions	Historical	18	18,129	British Universities
A loss of sovereignty in Britain	Social Science	33	19,019	British Universities
Transport, boxing, parliamentary system, fox hunting	Current Issues	50	60,209	British Universities
Euthanasia, capital punishment, yoga, nuclear power, values, abortion, ethics etc.	Current Issues	46	54,285	Marquette University
Money, crime, feminists	Current Issues	28	13,454	Indiana University
Divorce, welfare system, homelessness, water pollution	Current Issues	6	12,447	Presbyterian College, South Carolina
Genders' role in society	Current Issues	6	5,710	University of South Carolina
Rules and regulations, death penalty, salary caps, sex in media, US government etc.	Current Issues	17	18,630	University of South Carolina
Football, drinking age, talk shows, gun control, recycling.	Current Issues	13	15,815	University of South Carolina
Woman in combat, rules, sink or swim, curfew, book-banning etc.	Current Issues	17	12,730	University of South Carolina
Great inventions and discoveries of 20 th century and impacts on people's live.	Current Issues	43	16,502	University of Michigan
Unknown	Literature	16	18,826	Presbyterian College, South Carolina
Hamlet, Voltaire's Candidate	Literature	8	4,436	Presbyterian College, South Carolina
Aspects of social psychology, homicide, racial prejudice.	Current Issues	32	5,094	Presbyterian College, South Carolina

2.6. Corpus Tool: Sketch Engine

Sketch Engine is an effective corpus tool which helps us in exploring how language works and how language affects our culture. It has algorithms which can analyze millions of authentic texts and show what is typical in a language, or what is rare or unusual. It is frequently used by linguists, translators, learners, lexicographers etc. Sketch Engine can be accessed online partially free of charge or can be bought. Several processes were carried out in Sketch Engine software. First, lexical bundles on four corpora were analyzed, which enabled to see the frequencies of the target lexical bundles besides enabling to see the list of the most common bundles. With the help of the Sketch Engine tool, there is a chance to see the lexical bundles and how they were used in the learners' texts.

While using the Sketch Engine tool, log-likelihood value was used as a statistic to compare corpora. Log-likelihood is one of the fundamental components of the statistic, which means that by searching log-likelihood data, it can be seen whether the bundles in corpora are used in an

appropriate way compared to the reference corpora or not. The function of the log-likelihood is the density function considered as the function of 0. It can be suggested that, if the log-likelihood score is closer to zero, then it can be accepted that corpora are used in the same way for that bundle. However, if there is a big gap like more than 3,84, then it is underused or overused. Log-likelihood values were calculated at <http://ucrel.lancs.ac.uk/llwizard.html> automatically, in other words, it may be concluded that mistakes resulting from the calculation are minimized. Table 20 below shows details of the critical value details of log-likelihood.

Table 20: Values of Log Likelihood

95th percentile; 5% level; $p < 0.05$; critical value = 3.84
99th percentile; 1% level; $p < 0.01$; critical value = 6.63
99.9th percentile; 0.1% level; $p < 0.001$; critical value = 10.83
99.99th percentile; 0.01% level; $p < 0.0001$; critical value = 15.13

Retrieved from: <http://ucrel.lancs.ac.uk/llwizard.html>

According to the table, if the score is closer to zero, then the density of both corpora are acceptable and they are parallel to each other but if it is more than 3.84 or -3.84 there is a problem, which means that density is different and there is no likeness between the corpora. If the scores get bigger, there are underuses or overuses of lexical bundles compared to the reference corpora.

CHAPTER THREE

3. FINDINGS AND DISCUSSIONS

Ädel and Erman (2012: 81) state that “the degree of proficiency correlates significantly with the proportion and/or types of formulaic language used”. There have been many studies in the field of phraseology recently and lexical bundles are one of the important points in phraseology and language and a broad analysis of bundles were done in a cross-linguistic way in the current research. In the shed of four corpora, the data were collected according to values of raw frequency, normalized frequency and log-likelihood values and then, the data were listed below in several tables, in which we can see lexical bundles broadly and compare them with each other. While raw frequency means the plain number of bundles in a related corpus depending on the size, normalized frequency is a measurement which shows the related bundles in per million rate. Besides, log-likelihood value shows us how closely the bundles are used to each other in two different corpora in the present study. The first comparison was made between the two academic corpora. Log-likelihood scores of BAWE and KTUCALE were calculated in order to analyze their quantitative differences and similarities of usage proportions.

In Table 21, 100 most common bundles in BAWE and KTUCALE corpora are listed and their raw and normalized frequencies were measured and common bundles are shown bold style.

In Table 22, common lexical bundles in BAWE and KTUCALE, which were written in a bold style in table 21, are listed.

In Table 23, 100 most common lexical bundles in LOCNESS and TICLE were listed and as in Table 21, common lexical bundles in both corpora are shown in bold style.

In Table 24 most common lexical bundles in both corpora which were written in a bold style in Table 23 are shown.

Table 25 shows the most common four word lexical bundles in Biber’s corpus and the most common three word lexical bundles in the corpora of Salazar and Davis and frequencies of these bundles are shown in target corpus KTUCALE and reference corpus BAWE.

Table 26 shows: 50 most common lexical bundles in reference corpus LOCNESS.

Table 27 shows the first 50 lexical bundles in BAWE with normalized and raw frequencies and these target lexical bundles are listed also in KTUCALE.

Table 28 shows the most common four word lexical bundles in Biber's corpus and the most common three word lexical bundles in the corpora of Salazar and Davis and frequencies of these bundles are shown in target corpus TICLE and reference corpus LOCNESS.

3.1. Categorization of the Most Common Lexical Bundles in Four Corpora

Taking categorization of lexical bundles by Biber into consideration, there are some surprising results considering the use of bundles. First, L2 learners preferred bundles from the category of noun phrase with "of" phrase fragments mostly. For example, lexical bundles such as *one of the most*, *the purpose of*, *the importance of*, *the total number of*, *the development of the*, *the process of* and *the meaning of* are the most common lexical bundles used by L2 learners in KTUCALE. On the other hand, the most commonly preferred bundles in BAWE, which is also an academic argumentation, are chosen from the other prepositional phrase (fragment) such as *on the other hand* and prepositional phrase with embedded of-phrase fragment like *the in terms of*.

Another point is that L2 learners in TICLE mostly prefer bundles from the same category with L2 learners' in KTUCALE. Shortly, it can be concluded that using the habit of the lexical bundles are same in TICLE and KTUCALE. Also, in LOCNESS, L1 speakers mostly chose from the category of *noun phrase with of phrase fragments* and prepositional phrase with the embedded of-phrase fragment

3.2. The Most Common Lexical Bundles in KTUCALE and BAWE

Table 21 below shows the most common lexical bundles in KTUCALE and BAWE and they are listed in order, depending on their normalized frequency per million. Sinclair (1991: 30) stated that "anyone studying a text is likely to need to know how often each different word form occurs in it." Table 21 indicates that preference of lexical bundles of native learners differs from the Turkish counterparts in a big portion. Variety of their preferences shows us that Turkish EFL learners are using lexical bundles much more differently than native learners. 31 lexical bundles are commonly used in both corpora and 69 lexical bundles are not the same as each other in each corpus.

Allen (2010: 120) states that "by comparing learner corpora with reference corpora, language production can be investigated and when such production diverges from the target register, teaching materials can be produced specifically targeting the learner population." The main reason to compare the target corpus with reference one is like that of Allen's. It provides a chance to develop materials or make use of language environments of native speakers.

Table 21: The Most Common Lexical Bundles in KTUCALE and BAWE

Bundles	KTUCALE		BAWE		
	Raw frequency	Normalized frequency	Bundles	Raw frequency	Normalized frequency
one of the	556	678,18	the use of	374	599,08
in order to	527	642,80	in order to	325	520,59
the most important	304	370,80	be able to	231	370,02
there be a	300	365,92	there be a	186	297,94
be able to	300	365,92	the fact that	180	288,33
on the other	261	318,35	that there be	153	245,08
the other hand	259	315,91	there be no	140	224,25
on the other hand	255	311,03	that it be	137	219,45
good language learner	250	304,94	of the poem	135	216,25
a foreign language	246	300,06	one of the	134	214,64
the use of	239	291,52	way in which	133	213,04
in term of	237	289,65	the end of	129	206,63
be the most	187	288,09	seem to be	100	160,18
a second language	232	282,98	part of the	129	206,63
it can be	230	280,54	it be not	124	198,63
a lot of	222	270,78	as well as	123	197,02
the process of	217	264,68	use of the	121	193,82
of the most	213	259,80	it be a	121	193,82
it be not	208	253,71	in term of	121	193,82
in the classroom	205	250,05	a sense of	118	189,01
be one of	202	246,39	the way in	113	181,01
vocabulary learn strategy	200	243,95	the way in which	110	176,20
the target language	192	234,19	can be see	105	168,19
language and culture	191	232,97	the importance of	104	166,59
the meaning of	190	231,75	due to the	103	164,99
one of the most	188	229,31	on the other	102	163,39
the purpose of	185	225,65	the other hand	99	158,58
that there be	185	225,65	on the other hand	99	158,58
of the student	185	225,65	it be the	96	153,77
there be many	184	224,43	be use to	96	153,77
be use in	180	219,55	be see as	95	152,17
be one of the	179	218,33	a number of	95	152,17
the importance of	178	217,11	the idea of	93	148,97
culture and language	178	217,11	can not be	93	148,97
as well as	176	214,67	at the end	90	144,16
language be a	171	208,58	the role of	89	142,56

Table 21: (Continue)

Bundles	KTUCALE		BAWE		
	Raw frequency	Normalized frequency	Bundles	Raw frequency	Normalized frequency
to be a	170	207,36	as it be	89	142,56
they do not	165	201,26	heart of darkness	88	140,96
what be the	164	200,04	refer to the	86	137,76
be a good	152	185,80	look at the	86	137,76
language learn strategy	146	178,08	the meaning of	84	134,55
a new language	145	176,86	an example of	84	134,55
this study be	143	174,42	of the text	83	132,95
be very important	143	174,42	as a result	81	129,75
and it be	141	171,98	in which the	80	128,15
part of the	140	170,76	such as the	79	126,54
of this study	139	169,54	in the first	79	126,54
in addition to	139	169,54	to be a	78	124,94
be clear and	139	169,54	the beginning of	78	124,94
the relationship between	138	168,32	it have be	76	121,74
the questionnaire must	135	165,64	it can be	74	118,53
must be clear and	135	165,64	have to be	74	118,53
must be clear	135	165,64	at the end of	74	118,53
it be a	135	165,64	suggest that the	73	115,33
of the language	132	161,01	the reader to	71	113,73
as cite in	131	159,79	end of the	71	113,73
can not be	130	158,57	of the word	70	112,13
a good teacher	129	157,35	focus on the	70	112,13
accord to the	128	156,13	to the reader	69	110,53
that it be	125	152,47	be use in	68	108,92
to learn a	123	150,03	it be possible	67	107,32
of them be	123	150,03	the use of the	66	105,72
be relate to	123	150,03	be find to	66	105,72
do not have	121	147,59	the idea that	65	104,12
learn a language	120	146,37	the end of the	64	102,52
the fact that	118	143,93	this be a	63	100,92
second language acquisition	118	143,93	the effect of	63	100,92
of the most important	118	143,93	the reader be	62	99,31
a variety of	116	141,49	in the text	62	99,31
the role of	115	140,27	be not a	62	99,31
need to be	115	140,27	appear to be	62	99,31
one of the most important	112	136,61	i do not	61	97,71
language learning be	112	136,61	the form of	60	96,11
in other word	112	136,61	and it be	59	94,51
a good language	112	136,61	the nature of	58	92,91

Table 21: (Continue)

Bundles	KTUCALE		BAWE		
	Raw frequency	Normalized frequency	Bundles	Raw frequency	Normalized frequency
there be no	111	135,39	in relation to	58	92,91
learn a new	111	135,39	the concept of	57	91,30
the development of	110	134,17	in this case	57	91,30
in language learning	108	131,73	at the beginning	56	89,7
there be some	107	130,51	the relationship between	55	88,10
teacher and student	107	130,51	the speaker be	54	86,50
teacher and educator	107	130,51	the purpose of	54	86,50
the questionnaire be	106	129,29	the notion of	54	86,50
ideal language teacher	106	129,29	the fact that the	54	86,50
want to learn	105	128,07	the context of	54	86,50
listen and speak	105	128,07	in the same	54	86,50
in this study	104	126,85	fact that the	54	86,50
of language learning	103	126,85	nature of the	53	84,90
the nature of	101	125,63	it be also	53	84,90
point out that	101	125,63	be possible to	53	84,90
the second language	99	120,75	the development of	52	83,29
the effect of	99	120,75	relate to the	52	83,29
of the questionnaire	98	119,53	of the time	52	83,29
be that the	97	118,31	it be important	52	83,29
in the questionnaire	96	117,10	be base on	52	83,29
in the process	96	117,10	there is also a	51	81,69
teacher should be	95	115,88	the text be	51	81,69
it be the	94	114,64	some of the	51	81,69
it can be	94	114,64	need to be	51	81,69
language learning activities	94	114,64	be find in	51	81,69

When raw and normalized frequencies are overviewed, it can be seen that there are two bundles for each corpus that have been used much more than the other bundles; *one of the* and *in order to* in KTUCALE corpus and *the use of* and *in order to* in BAWE corpus. Although preferences of learners differ, the upmost bundles preferred by L2 learners and native speakers are the same bundles. As Schmitt and Carter (2004: 13) state “lack of rich input”, overuse and underuse of these expressions are common in L2 writing. It can be seen from table that Turkish learners overuse some of the lexical bundles and distributions are homogeneous in reference corpus, whilst in target corpus KTUCALE distribution is non-homogeneous. It is because of the fact that L2 learners overuse the bundles they learn or are exposed to. Li & Schmitt, (2009) underlines that L2 learners may overuse the bundles that they are exposed to. Despite the fact that, top bundles are generally the same in both corpora there are some points that vary from each other. For example, bundles such as *one of the*, *on the other*, *on the other hand*, *the other hand* are used in

KTUCALE to a great extent while they are not as much preferred in BAWE. On the other hand, some of the bundles preferred mostly in BAWE are not preferred in KTUCALE such as, *of the poem, way in which, the end of, seem to be*. As the BAWE is reference corpus for academic expository texts, there is a problem among Turkish learners as they do not use the bundles preferred by native learners.

Another important point is that, when the distribution of the lexical bundles is analyzed in BAWE, it can be noted that these lexical bundles are frequently preferred by native learners. Reppen & Biber (2016) point out that learners overuse the bundles as the learners are more exposed than L2 learners. The overuse of the bundles may cause repetitions in texts which will cause ineffectiveness.

3.3. Common Lexical Bundles in BAWE and KTUCALE

Pérez-Llantada (2014) analyzed lexical bundles in a corpus, and found that nearly %17 of the total bundles are commonly seen. Thus, the present study emphasizes the need for a cross-linguistic study on the influence of lexical bundles within the Turkish context. With the help of the cross-linguistic study, how bundles are used in Turkish EFL learners' texts and how they are used by native learners can be analyzed. Common 31 lexical bundles in BAWE and KTUCALE show us how Turkish EFL learners use bundles. Considering the normalized frequencies, some observations were described according to Table 22. It should be noted that native corpus provides significant data as it provides a chance to compare L2 learners with native speakers.

Top bundles are common in both corpora, which can be seen in the Table 22, though normalized frequencies of them differ dramatically.

Table 22: Common Lexical Bundles In BAWE and KTUCALE

Bundles	BAWE		KTUCALE	
	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency
the use of	374	599,08	239	291,52
in order to	325	520,59	527	642,80
be able to	231	370,02	300	365,92
there be a	186	297,94	300	365,92
the fact that	180	288,33	118	143,93
that there be	153	245,08	185	225,65
there be no	140	224,25	111	135,39
that it be	137	219,45	125	152,47
one of the	134	214,64	556	678,18
part of the	129	206,63	140	170,76

Table 22: (Continue)

Bundles	BAWE		KTUCALE	
	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency
it be not	124	198,63	208	253,71
as well as	123	197,02	176	214,67
it be a	121	193,82	135	165,64
in term of	121	193,82	237	289,65
on the other	102	163,39	261	318,35
the other hand	99	158,58	259	315,91
on the other hand	99	158,58	255	311,03
it be the	96	153,77	94	114,64
the importance of	104	166,59	178	217,11
can not be	93	148,97	130	158,57
the role of	89	142,56	115	140,27
to be a	78	124,94	170	207,36
it can be	74	118,53	230	280,54
the meaning of	84	134,55	190	231,75
be use in	68	108,92	180	219,55
the effect of	63	100,92	99	120,75
and it be	59	94,51	141	171,98
the nature of	58	92,91	101	125,63
the relationship between	55	88,10	138	168,32
the purpose of	54	86,50	185	225,65
the development of	52	83,29	110	134,17

Table 22 shows us the use of common lexical bundles which vary in both corpora in terms of frequencies. In general, as shown in the table, L1 speakers used lexical bundles more homogeneously than L2 learners did. However, as can be seen from some examples above, there are some bundles that L2 learners used much more than L1 speakers. This is an example of the overuse which happens when they always encounter the same bundles in the language environment. Not having been exposed to language varieties, L2 learners are accustomed to using a restricted number of bundles. The overusing of bundles is a problem in the language environment because of the fact that it hinders L2 learners' use of other bundle. Furthermore, normalized frequency values are relatively high in the table as L2 learners and native speakers tend to use the same bundles.

3.4. The Most Common Lexical Bundles in LOCNESS and TICLE

Table 23 below shows us the most commonly seen lexical bundles both in LOCNESS and TICLE corpora and they are listed in order, according to their normalized frequencies, which means the generalization of raw frequency to per million. As in Table 22, usage of the lexical bundles varies from each other in L1 and L2 corpora and contrast with each other. Although

percentages of the common bundles are higher compared to results in BAWE and KTUCALE, it is not parallel with L1 corpus. Table 23 below indicates that there are more common bundles which have been preferred by L1 and L2 learners in expository writing. As can be seen from the table, the first point is that there is a bundle named, *they do not*, and it was often used by L1 speakers. While the normalized frequency of the top bundle in L1 corpus is nearly 520 times, it is nearly 1083 in L2 corpus. The overuse of bundles is a critical problem in expository writing because of the fact that frequencies are extremely high compared to the L1 corpus.

Among the 100 most common bundles in each corpus, only 36 bundles are the same though their frequencies are different, and 64 bundles are different in each corpus. This situation shows that L2 learners use bundles in a different way compared to L1 speakers. Varieties of these bundles may cause several problems; to give an example Hyland (2008: 2) observes the following; “lexical bundles are familiar to writers and readers who regularly participate in a particular discourse, their very ‘naturalness’ signalling competent participation in a given community. Conversely, the absence of such clusters might reveal the lack of fluency”.

L2 learners cannot be native enough when they insist on using the same bundles and the fact that they are expected to use bundles naturally in language makes this finding significant in several points. Allen (2010: 107) explains that “learners rarely have competent use of such lexical bundles when they begin to study academic discourse in a second language, even if they have experience of participation in such communities in their first language”. In many of the well-known linguists’ studies, such as Granger, Hyland and Milton, Biber, it can be seen that language learners overuse, underuse or misuse particular bundles. In Table 23 below; the findings are the same as these linguists’ findings in terms of the overuse of the bundles and the underuse of them. To give an example for underusing of bundles; one of the bundles “*the fact that*” is the second most commonly used bundle in LOCNESS but it is not among the most common 100 bundles in TICLE. On the other hand, bundles like “*they can not*” and “*the most important*” are examples of two of the most common bundles in TICLE though they are not even in the list of top 100 bundles in LOCNESS.

Use of some of the bundles by EFL learners are not in the list of L1 speakers even though they are actively used in the list of L1 speakers such as *the fact that*, *the end of*, *the united states*, *the idea of*, *have to be*. These bundles are used by Turkish EFL learners in a limited number of times. Another point is that some of the lexical bundles which have been used plenty of times by L2 learners are not preferred by L1 speakers.

Table 23: The Most Common Lexical Bundles In TICLE and LOCNESS

Bundles	TICLE		LOCNESS		
	Raw frequency	Normalized frequency	Bundles	Raw frequency	Normalized frequency
they do not	242	1.083,02	be able to	188	520,70
do not have	157	702,62	the fact that	162	448,69
man and woman	148	662,34	in order to	127	351,75
it be not	147	657,87	that it be	124	343,44
there be a	136	608,64	one of the	121	335,13
a lot of	124	554,94	the united states	114	315,74
on the other	119	532,56	there be no	109	301,89
in order to	118	528,08	it be not	106	293,58
they can not	115	514,66	there be a	101	279,74
there be no	112	501,23	it be a	94	260,35
the most important	109	487,81	they do not	91	252,04
in the world	108	483,33	the end of	82	227,11
the other hand	107	478,86	the idea of	81	224,34
one of the	105	469,91	have to be	80	221,57
on the other hand	105	469,91	because of the	79	218,80
can not be	97	434,10	this be a	76	210,89
most of the	91	407,25	due to the	76	210,89
it be a	87	389,35	that they be	75	207,73
it be the	86	384,88	the right to	73	202,19
they be not	84	375,92	end of the	73	202,19
the real world	83	371,45	the death penalty	71	196,65
should not be	82	366,97	should not be	71	196,65
as a result	78	349,07	the use of	70	193,88
there be some	77	344,60	the number of	70	193,88
be able to	75	335,65	of the world	69	191,11
that it be	74	331,17	it be the	68	188,34
i do not	73	326,70	can not be	68	188,34
we can not	72	322,22	the end of the	67	185,57
that there be	71	317,75	part of the	67	185,57
people who be	71	317,75	in the united	67	185,57
in my opinion	70	313,27	be not the	67	185,57
the people who	67	299,84	do not have	65	180,03
do not know	67	299,84	in the world	64	177,26
be the most	65	290,89	for the good	64	177,26
be one of	65	290,89	to be a	63	174,49
there be many	64	286,42	as well as	63	174,49
first of all	63	281,94	a lot of	61	168,95
woman and man	61	272,99	be one of	60	166,18
be not a	61	272,99	be in the	60	166,18
i want to	60	268,52	that he be	59	163,41
we do not	58	259,57	he do not	59	163,41
for the real	58	259,57	there have be	58	160,64

Tablo 23: (Continue)

Bundles	TICLE		LOCNESS		
	Raw frequency	Normalized frequency	Bundles	Raw frequency	Normalized frequency
because of the	58	259,57	that there be	58	160,64
in our country	57	255,09	on the other	58	160,64
if there be	57	255,09	it would be	57	157,87
be one of the	57	255,09	in the united states	57	157,87
that they be	56	250,62	seem to be	56	155,10
do not want	56	250,62	at the end	56	155,10
do not prepare	54	241,67	this be not	54	149,56
university degree be	53	237,19	ethnic american literature	53	146,79
in the society	53	237,19	be not a	53	146,79
and do not	53	237,19	the people of	51	141,75
prepare student for	52	232,72	i do not	51	141,75
of the people	52	232,72	as it be	51	141,75
to have a	51	228,24	the other hand	50	138,48
of the world	51	228,24	on the other hand	50	138,48
day by day	50	223,76	be go to	50	138,48
but it be	50	223,76	this be the	49	135,71
there be not	47	210,34	there be many	49	135,71
part of the	47	210,34	invention of the	49	135,71
you do not	46	205,86	be see as	49	135,71
they have to	45	201,39	do not want	48	132,94
point of view	45	201,39	to have a	47	130,17
of the most	45	201,39	the invention of	46	127,40
be in the	44	196,91	be a very	46	127,40
for the real world	43	192,44	to be the	45	124,64
between man and	43	192,44	of the play	45	124,64
you can not	42	187,96	be to be	45	124,64
one of the most	42	187,96	be one of the	45	124,64
i believe that	42	187,96	because it be	44	121,87
between man and woman	42	187,96	be use to	44	121,87
the help of	41	183,49	as a result	44	121,87
student for the	41	183,49	the rest of	43	119,10
of the student	41	183,49	if it be	43	119,10
not prepare student	41	183,49	be allow to	43	119,10
accord to the	41	183,49	it have be	42	116,33
they want to	40	179,01	be that the	42	119,10
it be very	40	179,01	be for the	42	119,10
to be a	39	174,54	at the end of	42	119,10
the real life	39	174,54	the question of	41	113,56
student for the real	39	174,54	but it be	41	113,56
it can be	39	174,54	they be not	40	110,79
in addition to	39	174,54	the case of	40	110,79

Tablo 23: (Continue)

Bundles	TICLE		LOCNESS		
	Raw frequency	Normalized frequency	Bundles	Raw frequency	Normalized frequency
be a lot	39	174,54	the beginning of	40	110,79
to sum up	38	170,06	such as the	40	110,79
the right to	38	170,06	need to be	40	110,79
of them be	38	170,06	because they be	40	110,79
not want to	38	170,06	be try to	40	110,79
have the same	38	170,06	would not be	39	108,02
have the right	38	170,06	lead to the	39	108,02
do not prepare student	38	170,06	believe that the	39	108,02
be a lot of	38	170,06	be force to	39	108,02
not prepare student for	37	165,59	the human brain	38	105,25
do not think	37	165,59	the amount of	38	105,25
and they be	37	165,59	loss of sovereignty	38	105,25
will not be	36	161,11	in public school	38	105,25
if you have	36	161,11	aware of the	38	105,25
woman can not	35	156,64	would be a	37	102,48
there be a lot	35	156,64	the majority of	37	102,48
some of them	35	156,64	the invention of the	37	102,48

3.5. Common Bundles in TICLE and LOCNESS

If the lexical bundles which have been used in target corpus are also common in L1 reference corpus, it means that there is a tendency for EFL learners to use bundles in an appropriate way but if bundles are not in the same direction, then there is a problem in L1 usage. 36 lexical bundles are used in both of the corpora, which means that 64% of the lexical bundles which have been used by L2 learners differ from their native speakers counterparts. It can be concluded that L2 learners do not use lexical bundles like native speakers, thus, this situation may hinder fluency and semantic.

8 of the bundles among the top 10 most commonly used bundles in both corpora are the same however, there are differences in terms of frequencies. It is quite surprising to see that 8 most common bundles among 10 in TICLE are extremely overused considering the native corpus, and the other 2 bundles are used at least more than twice.

Table 24: Common bundles in TICLE and LOCNESS

Bundles	LOCNESS		TICLE	
	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency
be able to	188	520,70	75	335,65
in order to	127	351,75	118	528,08
that it be	124	343,44	74	331,17
one of the	121	335,13	105	469,91
there be no	109	301,89	112	501,23
it be not	106	293,58	147	657,87
there be a	101	279,74	136	608,64
it be a	94	260,35	39	174,54
they do not	91	252,04	242	1.083,02
because of the	79	218,80	58	259,57
that they be	75	207,73	56	250,62
the right to	73	202,19	38	170,06
should not be	71	196,65	82	366,97
of the world	69	191,11	51	228,24
it be the	68	188,34	86	384,88
part of the	67	185,57	47	210,34
do not have	65	180,03	157	702,62
in the world	64	177,26	108	483,33
to be a	63	174,49	39	174,54
a lot of	61	168,95	124	554,94
be one of	60	166,18	65	290,89
be in the	60	166,18	44	196,91
that there be	58	160,64	71	317,75
on the other	58	160,64	119	532,56
be not a	53	146,79	61	272,99
i do not	51	141,75	73	326,70
the other hand	50	138,48	107	478,86
on the other hand	50	138,48	105	469,91
there be many	49	135,71	64	286,42
do not want	48	132,94	56	250,62
to have a	47	130,17	51	228,24
be one of the	45	124,64	57	255,09
as a result	44	121,87	78	349,07
but it be	41	113,56	50	223,76
they be not	40	110,79	84	375,92

3.6. The Most Common 50 Lexical Bundles in Reference Corpus BAWE

In this section, log-likelihood values of the most common bundles in reference corpus, BAWE, were analyzed. BAWE contains academic argumentations of written texts of L1 speakers and it has the same features with target corpus KTUCALE. Log-likelihood values enable us to see

the differences in frequencies between the target and reference corpora, and it can be figured out whether there is a significant difference in use or not. If the absolute value is over 3.84, there is a significant difference, and if there is not, there is not a significant different use of bundles between L1 speakers and L2 learners. Also, when the value is $p < 0.05$, it can be concluded that usage of L1 speakers and that of L2 learners are in the same way. Briefly, it may be accepted that when the LL scores of bundles are close the zero, it is accepted that L1 speakers and L2 learners use them correspondingly. On the other hand, if the score is not close to zero, there is a problem which results in overuse and underuse of the bundles. Table 25 below shows the LL scores of 50 most common bundles in BAWE.

Table 25: The Most Common 50 Lexical Bundles in Reference Corpus BAWE

BAWE			KTUCALE			
First 50 bundles in BAWE	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency	-,+	LL score
the use of	374	599,08	239	291,52	-	78,11
in order to	325	520,59	527	642,80	+	9,07
be able to	231	370,02	300	365,92	-	0,02
there be a	186	297,94	300	365,92	+	4,92
the fact that	180	288,33	118	143,82	-	35,40
that there be	153	245,08	185	225,65	-	0,57
there be no	140	224,25	111	135,39	-	15,91
that it be	137	219,45	125	152,47	-	8,67
of the poem	135	216,25	3	3,66	-	200,93
one of the	134	214,64	556	678,18	+	175,00
way in which	133	213,04	39	47,57	-	83,09
the end of	129	206,63	64	78,06	-	43,61
part of the	129	206,63	140	170,36	-	2,42
it be not	124	198,63	208	253,71	+	12,77
as well as	123	197,02	176	214,67	+	0,56
use of the	121	193,82	45	54,90	-	59,91
it be a	121	193,82	135	165,64	-	1,69
in term of	121	193,82	237	289,65	+	13,29
a sense of	118	189,01	23	28,05	-	98,53
the way in	113	181,01	29	35,37	-	78,61
the way in which	110	176,20	26	31,71	-	81,23
can be see	105	168,19	35	42,69	-	58,25
the importance of	104	166,59	178	217,11	+	4,70
due to the	103	164,99	44	53,67	-	43,16
on the other	102	163,39	261	318,35	+	35,45
seem to be	100	160,18	53	64,65	-	30,31
the other hand	99	158,58	259	315,91	+	37,13
on the other hand	99	158,58	255	311,03	+	35,21
it be the	96	153,77	94	114,64	-	4,08
be use to	96	153,77	21	25,61	-	74,67
be see as	95	152,17	52	63,43	-	27,70
a number of	95	152,17	83	101,24	-	7,37
the idea of	93	148,97	49	59,77	-	28,48
can not be	93	148,97	130	158,57	+	0,21

Table 25: (Continue)

BAWE			KTUCALE			
First 50 bundles in BAWE	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency	-,+	LL score
at the end	90	144,16	41	50,01	-	34,56
the role of	89	142,56	115	140,27	-	0,01
as it be	89	142,56	58	70,74	-	17,75
refer to the	86	137,76	41	50,01	-	30,91
look at the	86	137,76	35	42,69	-	38,72
the meaning of	84	134,55	190	231,75	+	18,28
an example of	84	134,55	20	24,39	-	61,71
of the text	83	132,95	67	81,70	-	8,85
as a result	81	129,75	89	108,56	-	1,84
in which the	80	128,15	65	79,70	-	8,33
such as the	79	126,54	24	29,27	-	47,85
in the first	79	126,54	30	36,59	-	38,21
to be a	78	124,94	170	207,36	+	14,48
the beginning of	78	124,94	46	56,11	-	19,78
it have be	76	121,74	81	98,80	-	1,70
it can be	74	118,53	230	280,54	+	47,12

When the use of bundles in L2 learners corpus was analyzed, L1 speakers corpus is accepted as reference to see the similar and different uses of bundles between L1 speakers and L2 learners. Ghonsooly, Ahmadi and Fatemi (2013: 22) underline the importance of lexical bundles in L1 speakers corpus by stating that;

The more frequently a learner is exposed to particular bundles, the higher the chances are for that learner to produce the bundles in speech or writing. This linear relationship between exposure and output is oftentimes affected by L1 background. If a bundle shares an equivalent in the L1, the probability of its use will increase.

It can be figured out from the table above that, both corpora contradict with each other dramatically, which means that there are problems while L2 learners are using bundles. There are only 9 bundles used by L1 speakers in accordance with L2, in other words, 18% of the bundles are compatible, and 82% are incoherent with each other. Among the lexical bundles, “*the role of*” with the LL score 0,01 is the most coherent one in both corpora surprisingly, as frequency of it is at the end of the list. “*it have to be, as a result, can not be, it be a, as well as, part of the, be able to and that there be* are the rest of the bundles which are coherent with reference corpus. Common feature in these bundles is that they are all consisting of three words and surprisingly, there are no four-word combinations in the list. It can also be interpreted that L1 speakers use all the four-word bundles in an incompatible way considering the reference corpus.

Another important point which can be extracted from the table is that there are lots of underuse and overuse problems which are made by L2 learners when they use lexical bundles and this situation underlines the importance of inefficiency of L2 learners while they use bundles. 29 of

the bundles, which equals to 58% of whole bundles, are underused by L2 learners, which means that L2 learners underused most of the bundles in the reference corpus. A possible reason for this might be that L2 learners use less diverse and more restricted number of bundles in their writings than native learners do because of the fact that they are frequently exposed to the same bundles. Adel and Erman (2012) summarize this situation as L2 learners are inclined to use a fewer and more limited number of bundles compared to native speakers. Nesselhauf's (2005) states that, L2 learners are more liable to underuse the bundles. Also, Elis (2012) explains that L2 learners tend to use the bundles that they are confident in, and that's why they underuse lexical bundles in the reference corpus. L2 learners are required to use underused bundles more to be fluent and effective in their writings; however, it is difficult for them to apply. When they go into detail, it can be seen that some of the bundles which were frequently preferred by L1 speakers are excessively underused by the L2 learners. If the LL values are over 15.13, then these bundles are excessively underused by L2 learners, and when the table is analyzed there are 24 bundles which were excessively underused by L2 learners. Utmost underused bundles are *of the poem*, *way in which*, *use of the*, *a sense of*, *the way in which*, *be use to* and *an example of*. Although they are used mostly by L1 speakers, they are not preferred by L2 learners.

Overusing of lexical bundles is another critical problem observed in the corpus of L1 speakers and the analysis in the current study showed that problem of overusing is less critical than underusing. Salazar (2006: 134) explains that; "Further examination of the overused bundles indicates the non-native writers' excessive reliance on a handful of highly frequent bundles, to the detriment of less common bundles with similar meanings". L2 learners do not think about diversity in their texts, they use the bundles they know best in contrast and it leads to the problem of overusing.

There are 12 lexical bundles in the table above which were overused by L2 learners and among them the bundle *one of the* is used much more than the other overused bundles with the LL score 175. *It can be* is the second bundle which has been used mostly by L2 learners with the LL score 47,12. To sum up, it can be concluded that the most commonly preferred bundle is at least three times more frequent than the other bundles. Besides, 6 bundles of 12 namely; *it can be*, *the meaning of*, *the other hand*, *one of the*, *be see as* and *on the other hand* are extremely overused by the L2 learners. Although these bundles have alternatives to be chosen, lack of knowledge by L2 learners led them to overuse bundles and the problem of overusing occurred. On the other hand, there may be many other reasons for L2 learners for not having chosen alternative bundles such as not having confidence, lack of knowledge, not being trained enough. Paquot (2013: 402) argue that "the first language may...prompt learners to use lexical bundles that display untypical...patterns in English". In other words, it can be concluded that language transfers from native language may be the reason for overusing problem too. Granger (2014: 69) also noted that "the lack of salience that

characterizes many lexical bundles constitutes a challenge for learners and trainees who may be led to produce awkward-sounding phrases, often directly transferred from their mother tongue”.

These findings suggest that in general, underuse and overuse of bundles are important problems for Turkish learners and there are many reasons for these problems, and underuse of bundles is more problematic than overusing in academic argumentation of Turkish EFL learners. These findings are the similar to the many other studies which were carried out to note differences between L1 speakers and L2 learners.

3.7. 50 Most Common Lexical Bundles in a Reference Corpus LOCNESS

In this section analysis of the log-likelihood values are taken into consideration to show whether there is a significant difference between the reference corpus and the target corpus. Log-likelihood values hold a critical position as they enable us to see how parallel lexical bundles are used in both corpora. In the previous section, the results of corpora which consist of academic argumentations were analyzed, and in this section, a different perspective that analyzes corpora, consisting of expository argumentations is used. Table 26 below shows us log-likelihood values of two expository corpora, TICLE a non-native corpus and LOCNESS, a native corpus.

Table 26: The Most Common 50 Lexical Bundles in a Reference Corpus LOCNESS

Bundles	LOCNESS		TICLE			
	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency	LL score	
					- ,+	score
be able to	188	520,70	75	335,65	-	10,94
the fact that	162	448,69	32	143,21	-	43,88
in order to	127	351,75	118	528,08	+	9,98
that it be	124	343,44	74	331,17	-	0,06
one of the	121	335,13	105	469,91	+	6,34
the united states	114	315,74	6	26,85	-	73,73
there be no	109	301,89	112	501,23	+	14,08
it be not	106	293,58	147	657,87	+	40,77
there be a	101	279,74	136	608,64	+	35,50
it be a	94	260,35	87	389,35	+	7,23
they do not	91	252,04	242	1.083,02	+	162,48
the end of	82	227,11	24	107,41	-	11,76
the idea of	81	224,34	19	85,03		18,98
have to be	80	221,57	28	125,31	-	7,71
because of the	79	218,80	58	259,57	+	0,97
this be a	76	210,89	34	152,16	-	2,57
due to the	76	210,89	7	31,33	-	38,67
that they be	75	207,73	56	250,62	+	1,12
the right to	73	202,19	38	170,06	-	0,67
end of the	73	202,19	12	53,70	-	24,71
the death penalty	71	196,65	6	26,85	-	37,80
should not be	71	196,65	82	366,92	+	14,79

Table 26: (Continue)

Bundles	LOCNESS		TICLE			
	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency	LL score	
					- ,+	score
the use of	70	193,88	22	98,46	-	8,54
the number of	70	193,88	30	124,36	-	2,96
of the world	69	191,11	51	228,24	+	0,92
it be the	68	188,34	86	384,88	+	19,63
can not be	68	188,34	97	434,10	+	28,45
the end of the	67	185,57	11	49,23	-	22,24
part of the	67	185,57	47	210,34	+	0,43
in the united	67	185,57	4	17,90	-	41,46
be not the	67	185,57	34	152,16	-	0,91
do not have	65	180,03	157	702,62	+	96,10
in the world	64	177,26	108	483,33	+	42,30
for the good	64	177,26	2	8,95	-	47,68
to be a	63	174,49	39	174,54	+	0,00
as well as	63	174,49	20	89,51	-	7,50
a lot of	61	168,95	124	554,94	+	62,67
be one of	60	166,18	65	290,89	+	9,73
be in the	60	166,18	44	196,91	+	0,62
that he be	59	163,41	9	40,28	-	21,00
he do not	59	163,41	23	102,93	-	3,76
there have be	58	160,64	23	102,93	-	3,46
that there be	58	160,64	71	317,75	+	14,91
on the other	58	160,64	119	532,56	+	60,82
it would be	57	157,87	6	26,85	-	26,83
in the united states	57	157,87	4	17,90	-	33,08
seem to be	56	155,10	23	102,93	-	2,89
at the end	56	155,10	24	107,41	-	2,37
this be not	54	149,56	33	147,68	-	0,00
be not a	53	146,79	61	272,99	+	10,90

It can be seen from the table that, use of the bundles in expository argumentations are more moderate than those in academic argumentations although there are still problems related to overuse and underuse of bundles. 32% of the bundles are in accordance with each other according to their LL scores and some of them are used in the same way as each other. There is no difference in LL scores of bundles; *there be not* and *to be a*, which means that L1 speakers use these bundles the same as L2 learners. 17 of the lexical bundles, which equals to 34%, are overused by L1 speakers and 17 of the lexical bundles, which equals to %34 are underused by L1 speakers too. In other words, it can be concluded that L1 speakers underuse and overuse the lexical bundles equally. Distribution of the lexical bundles is relatively on balance in expository texts of L1 speakers and L2 learners in terms of academic argumentations; however, the rate of underusing of lexical bundles is higher than the rate in academic texts.

The most commonly underused lexical bundles are *for the good*, *in the united*, *the fact that*, *the united states* and *due to the* among L1 speakers, which means that they have to develop their

writing skills in a way that L2 learners use these bundles. On the other hand, it can be extracted from the table that L1 speakers misuse the bundles but LL scores of these bundles are less than the scores of L1 speakers' academic argumentations. As it was stated in the previous section, there may be many reasons affecting the use of bundles by L1 speakers. According to Krishnamurthy (2002) one of the most important reasons for this situation is that L1 speakers are exposed to underused bundles in lectures. One of the other interesting analysis is that the bundle *the united states* which was preferred by L2 learners mostly is excessively underused by L1 speakers.

Overusing is another problem that we come across mostly in L1 speakers writing texts and there are some bundles which were underused in expository writings of L1 speakers in target corpus though problem of underusing is not as much as in the texts of academic argumentation of L1 speakers. Bundles *on the other*, *a lot of*, *do not have* and *they do not* are extremely overused and their LL scores are higher than extremely underused bundles. There is a significant point that the lexical bundle *they do not* with the LL score of 162,48 and normalized frequency of 1083,02 per million is on the top of the list but its normalized frequency is just 252,04 times per million in reference corpus, LOCNESS.

In the final part of this section, common values are focused on non-native corpora, KTUCALE and TICLE. Bundles *in order to*, *one of the*, *it be a*, *there be no*, *on the other* and *be not a* are the lexical bundles which have been overused by L1 speakers in both corpora. Moreover, the lexical bundles; *the fact that*, *the end of*, *due to the* and *the use of* are underused by L1 speakers in both corpora. There are not enough underused or overused lexical bundles between TICLE and KTUCALE and it may be accepted that they are not parallel to each other.

3.8. Common Bundles in Biber's, Salazar's and Davis' Corpora and Their Values in BAWE and KTUCALE

In previous sections, a broad analysis of the lexical bundles in L1 and L2 corpora was conducted so as to see the phraseological awareness of tertiary level Turkish EFL learners in written corpora. Also, in this section, some common bundles which were extracted from native reference corpora were analyzed to reinforce the analysis. Log-likelihood and normalized frequency values were the main interest in previous sections, and in this part, the most common bundles are searched in KTUCALE and BAWE to see how students used these bundles when they are compared with each other. LL scores and normalized frequency values can be seen in Table 27.

Table 27: Common Bundles in Biber's Salazar's and Davis' Corpora

COMMON BUNDLES IN BIBER'S SALAZAR'S AND DAVIS' CORPORA	KTUCALE		BAWE		+,-	LL score
	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency		
one of the	562	685,50	136	217,85	+	176
on the other hand	167	376,19	99	158,58	+	3,97
on the other	261	318,35	102	163,39	+	35,45
in terms of	232	282,98	121	193,82	+	11,79
as well as	213	259,80	150	240,27	+	0,54
one of the most	159	224,02	18	28,83	+	93,84
part of the	140	170,76	129	206,63	-	2,43
the fact that	123	150,03	182	291,53	-	33,21
of the most important	97	136,67	3	4,81	+	97,82
as a result of	88	123,20	37	59,27	+	9,10
the effect of	99	120,75	64	102,52	-	0,66
as a result	89	108,56	83	132,95	-	1,76
the number of	70	85,38	45	72,08	+	0,79
some of the	59	71,96	52	83,29	-	0,59
to the fact that	33	46,50	34	54,46	-	1,53
the end of the	29	40,86	64	102,50	-	24,76
at the end of	23	32,42	74	118,53	-	43,90
an important role in	20	28,18	17	27,13	-	0,11
the ability of	17	20,74	18	28,80	-	0,95
the case of	13	15,86	50	80,69	-	34,44
as a result of the	11	15,50	11	17,62	-	0,41
the development of the	10	14,09	10	16,02	-	0,37
the absence of	11	13,42	43	68,88	-	29,99
the fact that the	7	9,86	54	86,50	-	55,03
the presence of	8	9,76	27	43,25	-	16,72
been shown to	8	9,76	2	3,20	+	2,40
the beginning of the	6	8,45	41	65,70	-	45,33
was used to	6	7,32	5	8,01	-	0,02
for the first time	5	7,04	9	14,42	-	2,51
in addition to the	5	7,04	5	8,01	-	0,19
in terms of the	4	5,64	15	24,03	-	10,33
the purpose of the	4	5,64	9	14,42	-	4,55
the fact that it	4	5,64	11	17,62	-	5,58
in response to	4	4,88	15	24,03	-	10,13
the difference between the	2	2,82	10	16,02	-	8,22
as described in	2	2,44	3	4,81	-	0,57
shown in figure	2	2,44	0	0	+	2,26
is required for	2	2,44	3	4,81	-	0,57
the relationship between the	1	1,41	9	14,42	-	9,73
the work of the	1	1,41	6	9,61	-	5,45
the total number of	1	1,41	5	8,01	-	4,11
in the area of the	1	1,41	5	8,01	-	4,11

Table 27: (Continue)

COMMON BUNDLES IN BIBER'S SALAZAR'S AND DAVIS' CORPORA	KTUCALE		BAWE		+,-	LL score
	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency		
by the end of the	1	1,41	2	3,20	-	0,67
one of the most of	1	1,41	0	0	+	1,13
the same way as	1	1,41	9	14,42	-	9,73
with respect to the	0	0	12	19,22	-	20,33
as in the case	0	0	6	9,61	-	10,06
the ways in which	0	0	28	44,85	-	46,96
from the point of view of	0	0	2	3,20	-	3,35
such a way that	0	0	4	6,41	-	6,71
in such a way that	0	0	4	6,41	-	6,71
the way in which the	0	0	24	38,44	-	40,26
the aim of this study	0	0	1	1,60	-	1,68
the first part of the	0	0	2	3,20	-	3,35
such a way as to	0	0	1	1,60	-	1,68
in a number of ways	0	0	2	3,20	-	3,35
at the time of the	0	0	5	8,01	-	8,39
in the study of	0	0	6	9,61	-	10,36
between the two groups	0	0	0	0	+	0,00
in the present study	0	0	0	0	+	0,00
data not shown	0	0	0	0	+	0,00
the addition of	0	0	9	14,42	-	15,10

It can be seen in the table above that, there is an underusing problem faced by L2 learners, despite minor problems related to overusing of bundles. Also, it can be seen that normalized frequencies of the most bundles in L2 corpus are less than L1 corpus although there are some bundles which have been used more by L2 learners. Most of the bundles are underused by L2 learners compared to L1 speakers and it is significant that L2 learners extremely overused some bundles. Bundles such as *one of the*, *one of the most*, *on the other* and *of the most important* are dramatically overused by L1 speakers while the bundles such as *at the end of*, *the fact that*, *the ways in which* and *the ways in which the* are dramatically overused. 28 bundles are underused, and 6 bundles are overused by learners. Besides, three bundles; *data not shown*, *in the present study* and *between the two groups* are parallel to each other in both corpora as both L1 and L2 learners did not prefer them in their texts. The rest of the bundles also differ from each other, but the difference between overused and underused percentages of these bundles are not significant.

Normalized frequencies of the bundles were also compared to adopt a different perspective. 17 bundles are not seen in L2 corpus while there are just 5 bundles which are not preferred in L1 corpus and the bundle which has the highest normalized frequency value in L2 corpus is "*one of the*" with the scores of 685,50 per million and highest bundle is *the fact that* with the scores of

291,53 times per million. Most of the bundles are preferred by L1 speakers while they are not in L2 learners. L1 speakers use a variety of bundles but L2 learners prefer using a restricted number of bundles.

3.9. Common Bundles in Biber's, Salazar's and Davis' Corpora and Their Values in LOCCNESS and TICLE

There are lots of problems of L2 learners while they are using lexical bundles and these deficiencies make L2 learners inefficient in target language. These target bundles are generally underestimated by L2 learners, some of them are not even in the list. Some of the lexical bundles are overused extremely by L2 learners as it is the same in academic argumentations of learners and normalized frequencies of them are relatively high. Also, the most common bundles preferred by learners are the same and their distributions differ dramatically from the rest.

Table 28: Common Bundles in Biber's, Salazar's and Davis' Corpora and Their Values in LOCNESS and TICLE

COMMON BUNDLES IN BIBER'S SALAZAR'S AND DAVIS' CORPORA	TICLE		LOCNESS		+,-	LL score
	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency		
on the other	119	532,56	18	160,64	+	139,61
one of the	106	474,38	123	340,67	+	6,16
on the other hand	105	469,42	50	138,46	+	55,18
as a result	79	353,55	45	124,64	+	32,83
part of the	47	210,34	67	185,57	+	0,67
one of the most	14	196,94	31	85,86	-	0,99
as a result of	32	143,21	30	83,09	+	4,56
the fact that	32	143,21	163	451,46	-	44,49
the number of	30	134,26	70	193,88	-	2,96
some of the	23	102,93	35	96,94	+	0,35
as well as	20	89,51	76	210,49	-	13,43
in terms of	18	80,56	19	52,62	+	1,66
the effect of	14	62,65	34	94,17	-	1,73
at the end of	12	53,70	42	116,33	-	6,34
the end of the	11	49,43	67	185,57	-	22,24
the beginning of the	9	40,28	24	66,47	-	1,76
an important role in	7	31,33	5	13,85	+	1,98
the case of	6	26,85	40	110,49	-	14,45
for the first time	4	17,90	5	13,85	+	0,14
in addition to the	4	17,90	1	2,77	+	3,65
with respect to the	4	17,90	1	2,77	+	3,65
the ability of	4	17,9	2	5,54	+	1,98
the fact that the	3	13,43	28	77,55	-	13,03

Table 28: (Continue)

COMMON BUNDLES IN BIBER'S SALAZAR'S AND DAVIS' CORPORA	TICLE		LOCNESS		+,-	LL score
	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency		
to the fact that	3	13,43	24	66,47	-	10,06
the presence of	3	13,43	16	44,31	-	4,61
the relationship between the	2	8,95	1	2,77	-	0,99
as a result of the	2	8,95	11	30,47	-	3,28
of the most important	2	8,95	10	27,70	-	2,67
the development of the	2	8,95	3	8,31	+	0,01
as in the case	1	4,48	4	11,08	-	0,77
the ways in which	1	4,48	2	5,54	-	0,03
in terms of the	1	4,48	3	8,31	-	0,31
the work of the	1	4,48	0	0	+	1,92
the purpose of the	1	4,48	1	2,77	+	0,11
from the point of view of	1	4,48	0	0	+	1,92
the absence of	1	4,48	3	8,31	-	0,31
the total number of	0	0	1	2,77	-	0,96
such a way that	0	0	5	13,85	-	4,82
the difference between the	0	0	0	0	+	0,00
in the area of the	0	0	1	2,77	-	0,96
in such a way that	0	0	1	2,77	-	0,96
the way in which the	0	0	4	11,08	-	3,85
by the end of the	0	0	9	24,93	-	8,67
the aim of this study	0	0	0	0	+	0,00
one of the most of	0	0	0	0	+	0,00
the first part of the	0	0	1	2,77	-	0,96
the same way as	0	0	2	2,54	-	1,93
such a way as to	0	0	1	2,77	-	0,96
the fact that it	0	0	9	24,93	-	8,67
in a number of ways	0	0	3	8,31	-	2,89
at the time of the	0	0	1	2,77	-	0,96
in the study of	0	0	1	2,77	-	0,96
between the two groups	0	0	0	0	+	0,00
in the present study	0	0	0	0	+	0,00
data not shown	0	0	0	0	+	0,00
as described in	0	0	2	5,54	-	1,93
shown in figure	0	0	0	0	+	0,00
been shown to	0	0	2	5,54	-	1,93
the addition of	0	0	1	2,77	-	0,96
is required for	0	0	0	0	-	0,00
was used to	0	0	0	0	-	0,00
in response to	0	0	7	19,39	-	6,74

Distributions of the lexical bundles, including LL score and normalized frequencies above, vary from each other in minor details although distributions vary dramatically in academic argumentation corpora in the previous analysis. While 26 bundles are not used by EFL learners in TICLE, 10 bundles are not in the list of LOCNESS. This shows that EFL learners do not use variety of bundles in their writings while native learners prefer them in their writings. The most common lexical bundles from this table that L2 learners preferred are overused by them and they are not preferred much by the native learners. Compared to L2 learners' corpus, distribution of the lexical bundles in TICLE is more homogeneous and normalized frequencies are closer to each other.



CONCLUSION AND SUGGESTIONS FOR FURTHER RESEARCH

The purpose of the current study was to determine the phraseological awareness of Turkish EFL learners in their writing outputs within a crosslinguistic corpus-based analysis which gives opportunities to seek answers for research questions. The current study focused on comparing the use of bundles in L2 learner environment with L1 environment in which bundles are used homogeneously in their academic and expository texts. Increasing trend to use of phraseological words and lack of studies about phraseology inspired me to concentrate on this subject. Although there are limited number of studies on phraseology, it has been neglected for a long time. There is a need to overcome these challenges so as to produce effective outputs in writing lessons. Therefore, this study concentrates on the use of lexical bundles and this is stated by Altenberg (1998) who said that 80% of the words we use in a language consist of recurrent sequences. Pawley and Syder (1983), Sinclair (1991), Cowie (1998) and Wray and Perkins (2000) also underlined that languages consist of repetitive words and combinations. Cross-linguistic study, with the help of the corpus, contributed to this study a lot in terms of seeing the positive and negative sides of the L2 learners' lexical bundle performances. Instructors can use the findings to motivate learners to use lexical bundles and emphasize the importance of bundles. Howarth and Granger (1998), Erman (2009) state that L1 speakers use wide range of lexical bundles while L2 learners use restricted numbers. It can be seen in the current study that Turkish learners often use lexical bundles. However, the diversity of bundles are restricted and L2 Turkish learners focus on some bundles used frequently in their writings.

As an answer to the first research question, it can be concluded that Turkish EFL learners avoid using many bundles and they mostly just apply the same bundles and the reason for this situation is called "avoidance". According to Granger (1998), De Cock (2000) and Foster (2001) EFL learners use some of the formulaic word combinations repetitively because these word combinations are accepted as reliable safety nets by L2 learners. The use of same preferences enhances learners' confidence and when they use the language, the same bundles are preferred frequently. Thus this repetitive routine may be causing overuse problems. Laufer (2000) concludes that EFL learners use this strategy to overcome some of the language problems which came out because of some incongruences between native and target languages. Because of the lack of being exposed to same bundles in general, L2 learners start using them in language they use, and it becomes a habit in time. Also, when this behaviour becomes a habit, other alternatives are not sought for. Cortes (2004) also supports these findings about L2 learners stating that L2 learners prefer favoured repetition of the same bundles. As they feel more conscious while using the same

bundles, they do not pay attention to other bundles in a language. With respect to the analysis in this dissertation in previous parts, Turkish EFL learners have problems while they are using lexical bundles. They cannot keep the balance while using the lexical bundles and this situation results in overusing and underusing problems which affect writing texts negatively. Instructors may pay attention to a more balanced use of the variety of lexical bundles in writing lessons and when they encounter lexical bundles more in their life, they may have a chance to use more bundles which are particularly common in L1 speakers' texts. They have to be encouraged by their instructors to overcome language problems. Hyland (2008) explains that one of the critical components of the fluent linguistic production is lexical bundles. Therefore, it is inevitable for L2 language instructors to develop the skills of L2 learners so as to use lexical bundles appropriately and accurately in the process of enhancing academic writing skills. Many linguists such as Hyland (2008) and Chen and Baker (2010) emphasize the importance of integrating bundles in writing curriculums. When L2 learners are exposed to lexical bundles, it can be much easier for them to use bundles in a language environment.

Another research question of the current study was to determine whether L2 learners use the same bundles or not in general compared to L1 speakers' use of bundles. The most important point is that L2 learners mostly use a restricted number of bundles and they underestimate most of the important bundles found in L1 registers. This is mainly because L2 learners are not exposed to a variety of lexical bundles in a language environment and the texts they are exposed to are written texts which have been formed by experts in L2 language environments. More authentic texts may overcome this problem and learners may be encouraged to use more bundles actively. Cortes (2004) recommends noticing activities to improve awareness of functions, structures and context of lexical bundles in a given register. Although L2 learners use many lexical bundles in their written texts, the accumulation of some bundles causes repetitions; however, there is not a accumulation of lexical bundles in the products of L1 speakers despite a few exceptions. There is also one significant point that Turkish EFL learners do not use many bundles in the list of the most common bundles in L1 as efficiently as native learners. Furthermore, lack of using bundles in Turkish setting cause them to use same recurrent expressions and this causes overusing of some bundles.

The most commonly preferred bundles in Turkish setting also appear in L1 speakers' list to some extent and considering their normalized frequencies, surprisingly, some of the bundles are not used as much as by L1 speakers. Pang (2010: 11) states that "it will be advantageous to explicitly raise student awareness of these most common lexical bundles and have them practice using them in communicative writing activities". In other words, they may be able to improve the skills of using lexical bundles more efficiently with the help of these writing activities. The most preferred bundles are relatively the same in KTUCALE and TICLE although they are not so common in BAWE and LOCNESS. Besides, the most common bundles in our reference corpora BAWE and

LOCNESS are underestimated by Turkish EFL learners and this situation holds critical clues implying the problems of L2 learners.

In the next phase, the researcher concentrated on the other research question which is about the differences between L2 corpora KTUCALE and TICLE, in which the use of bundles varies to some extent. It can be seen that there is nearly no difference between them in terms of the preference of most common bundles but they differ from each other when rest of the bundles are analyzed. Although there are some common bundles not preferred much by the learners in both corpora, there are some examples which show us that there is a big difference in terms of their log-likelihood values and normalized frequencies. It can also be extracted from the tables that more bundles in expository corpus do not exist in the reference corpus. There is a problem of underusing in L2 corpus KTUCALE and a balanced distribution of overusing in TICLE compared to the reference corpora. Also, when the most common bundles in KTUCALE and TICLE were analyzed and compared with their reference corpora, it can be seen that the number of the common bundles in expository corpus TICLE is more than KTUCALE. Learners have used more bundles existing in the reference corpus.

To deal with those problems stated above, one of the possible choices is that lexical bundles can be integrated into lessons more effectively. When Turkish EFL learners are exposed to lexical bundles chosen from the native corpora, they may be able to use lexical bundles parallel to the L1 speakers in time. Lexical bundles can be integrated into coursebooks, reading texts, listening activities etc. Particularly, there is a need for the most common lexical bundles seen in L1 speakers' outputs to be integrated into writing curriculum in order for Turkish EFL learners to be able to use them more appropriately.

REFERENCES

- Ädel, A. and 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, 84-91.
- Allen, David. (2010). Lexical Bundles in Learner Writing: An Analysis of Formulaic Language in the ALESS Learner Corpus. **Komaba Journal of English Education**. 1. 105-127.
- Altenberg, Bengt, (1998), On the phraseology of spoken English: The evidence of recurrent word-combinations. In: A. P. Cowie: *Phraseology*. Oxford University Press, Oxford 1998, 101–122.
- Bal, Betül, (2010), **Analysis of Four-word Lexical Bundles in Published Research Articles Written by Turkish Scholars**, Thesis Georgia State University, p.10
http://scholarworks.gsu.edu/alesl_theses
- Belfaqeeh, N. (2009). **Strategies for translating idioms and culturally-bound expressions within the human development genre**. M.A. dissertation, University of Birmingham.http://www.cels.bham.ac.uk/resources/essays/Noor_Balfaqeeh_822875Diss.pdf (accessed 12/02/2011) [4] Bentivogli, L. & Pianta,
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). **Longman grammar of spoken and written English**. Harlow: Longman.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. **Applied Linguistics**, 25(3), 371-405. <http://dx.doi.org/10.1093/applin/25.3.371>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. **English for specific purposes**, 26(3), 263-286.
- Biber, Douglas, (1993a). An Analytical Framework for Register Studies In D. Biber and E. Finegan (eds), **Sociolinguistic Perspectives on Register**. Oxford University Press, New York. In press.
- _____ (2006), **University Language**, A corpus based study of spoken and written registers. John Benjamin Publishing Company, Amsterdam/Philadelphia

- _____ (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. **International Journal of Corpus Linguistics**, 14(3), 275-311.
<http://dx.doi.org/10.1075/ijcl.14.3.08bib>
- British Council, (2015), Türkiye’de yüksek öğretim kurumlarında İngilizce eğitimi, Bir durum analizi,
- Celebi, M. D. (2006). Türkiye’de Anadili Eğitimi ve Yabancı Dil Öğretimi. Sosyal Bilimler Enstitüsü Dergisi, 21(2), 285-307
- Çetinsaya, G. (2014). Büyüme, kalite, uluslararasılaşma: Türkiye yükseköğretimi için bir yol haritası. Ankara: Yükseköğretim Kurulu.
- Corder, P.S, (1968) **Advanced Study and the Experienced Teacher**. In Perren, G.E. (ed). Teacher of English as Second Language. Cambridge: Cambridge University Press.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. **English for Specific Purposes**, 23(4), 397-423.
<http://dx.doi.org/10.1016/j.esp.2003.12.001>
- Cowie, A.P. (1981), The treatment of collocations and idioms in learners’ dictionaries, **Applied Linguistics** 2(3): 223–235.
- _____ (1994). **Phraseology, Theory, Analysis and Applications**, Oxford University Press
- _____ (1998) **PHRASEOLOGY, Theory, Analysis and Applications**, Clarendon Press, Oxford p.101
- Cowie, A.P, and Mackin, R, (1997), **Oxford Dictionary of Current Idioms**, Volume 2, Oxford University Press, Oxford
- Davies, A. (1999), **An Introduction to Applied Linguistics: From Practice to Theory**. Edinburgh: Edinburgh University Press.
- Davis, Tracy S, (2012), **Undergraduate vs Graduate Academic English: A Corpus Based Analysis**, The Pennsylvania State University The Graduate School Department of Applied Linguistic
- Erman, B., & Warren, B. (2000), **The idiom principle and the open choice principle**. Text, 20(1), 29-62.
- Firth, J. (1957). Papers in linguistics, 1934-1951. London: Oxford University Press.
- Gledhill, C. (2000), **Collocations in Science Writing**. Tübingen: Gunter Narr Verlag. p.202

- Granger, Sylviane & Bestgen, Yves. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *IRAL - International Review of Applied Linguistics in Language Teaching*, 52. 229-252. 10.1515/iral-2014-0011.
- Granger, S. (2014). A lexical bundle approach to comparing languages: Stems in English and French. **Languages in Contrast**, 14(1), 58-72
- Gries, S. (2008), **Phraseology and linguistic theory**. (A brief survey). In S. Granger, & F. Meunier (eds.), *Phraseology: an interdisciplinary perspective*, 3-25. Amsterdam & Philadelphia: John Benjamins
- Gungor, Fatih (2016) *A Comparative Analysis of Lexical Bundles Used by Native and Non-native Scholars*, Vol 9, Canadian Center of Science and Education
- Halliday, M.A.K, (1966), **“Lexis as a Linguistic Level”**. In (ed) Bazzel, C.E. (ed). *In Memory of J. R. Firth*. London: Longman.
- _____ (1994), **Introduction Functional Grammar**, Edition, (2nd Edition) Publisher: Routledge Publishing
- Hill, J. (2000) *Revising priorities: From grammatical failure to collocational success*. In M. Lewis (ed.) *Teaching Collocation. Further Developments in the Lexical Approach* 47–69. Hove: LTP.
- Hoey, M. P, (2000), **Patterns of Lexis in Text**. London: Oxford University Press. p.202
- _____ (2009), **Corpus-driven approaches to grammar: the search for common ground**. U. Römer and R. Schulze (Ed.), *In Exploring the lexis-grammar interface*, (pp. 33- 47). Amsterdam/Philadelphia: John Benjamins Publishers Company.
- _____ (2005). *Lexical priming: a new theory of words and language*. London: Routledge, p.13
- Hogg, Micheal A, and Cooper, John (2007), **The Sage Handbook of Social Psychology**, The concise student edition, Sage Publications,
- Howie JGR, (1991),**Refining questions and hypotheses**. In: Norton PG, Stewart M, Tudiver F, Bass MJ, Dunn EV (ed). *Primary Care Research: Vol 1. Traditional and Innovative Approaches*. Newbury Park: Sage Publications; 13-25
- Huang, Kaisheng, (2014) **A corpus study of Chinese EFL majors' phraseological performance**, Hong Kong,
- Hunston, Susan, (2011) *Corpus Approaches to Evaluation. Phraseology and Evaluative Language*. Routledge, London/New York p.5

- Hunston, S., and Francis, G. (2000). **Pattern grammar: A corpus-driven approach to the lexical grammar of English**. Amsterdam: John Benjamins.
- Hyland, Ken (2008). **English for Specific Purposes**, An International Journal Volume 27, Issue 1, (4-21)
- _____ (2009). *Academic discourse: English in a global context*. London: Continuum.
- İlköğretim ve Eğitim Kanunu (1961), **T. C. Resmi Gazete**, (2006 n.26184)
- Kargı, Birkan, (2006), “Edebiyat Kılavuzunda Sinema Yapıtlarının Başlangıç Düzeyi Sonrası Yabancı Dil Öğretimine Katkısı”. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi* 22, 60-71.
- Kilimci, A., Can, C. (2008). TICLE: Uluslararası Türk Öğrenci İngilizcesi Derlemi. In M. Sarıca, N. Sarıca, A. Karaca, (Eds.), *XXIInd Ulusal Dilbilim Kurultayı Bildirileri*. Paper presented at the XXIInd Ulusal Dilbilim Kurultayı, Yüzüncü Yıl Üniversitesi, Van, 8-9 May (pp 1-11). Ankara: Yüzüncü Yıl Üniversitesi
- Kjellmer, G. (1991) **A mint of phrases**. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics* (pp. 111–127). London: Longman.
- Kohn, Kurt, (1986), **The Analysis of Transfer**, In E.Kellerman & M. Sharwood Smith (eds). *Crosslinguistic in second language acquisition and performance*, Oxford: Pergamon, 21-34
- Lado, R. (1957/1971). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press
- Lang, Peter, (1998), *The Lexical Item*. In: Weigand, Edda (ed.): **Contrastive Lexical Semantics**, 1–24. Amsterdam, Philadelphia: Benjamins.
- Langacker, Ronald W. (1987) **Foundations of Cognitive Grammar**, Volume 1, Stanford University Press
- Li, J. & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18, 85-102.
- Long, Thomas Hill, (1979), **Longman’s Dictionary Of English Idioms**, Longman.
- Mel’cuk, Igor, (1988), “Semantic description of lexical units in an Explanatory Combinatorial Dictionary: basic principles and heuristic criteria”, **International Journal of Lexicography**, 1, 165-88.
- _____ (1996), ‘**Lexical Functions: a tool for the description of lexical relations in a lexicon**’. In L. Wanner (ed.) (1996), 37-102.

- Meyer, Charles F, (2004), **English Corpus Linguistic**, Cambridge University Press, University of Massachusetts, p.1
- Miller, N. (2009) Assessing the processing demands of learner collocation errors. Poster presented at Corpus Linguistics 2009, Liverpool, U.K.
- Nesselhauf, N. (2005). Collocations in a learner corpus. Amsterdam: John Benjamins.
- Oktay, Aygul, (2015). Foreign Language Teaching: A Problem in Turkish Education
- Ozbay, Ali Şükrü, (2015), **Corpus Analysis of The Support Verb Construction Development and Use by EFL Learners**, Karadeniz Technical University, January
- Paquot, Magali, and Granger, Sylviane, (2012), **Formulaic Language in Learner Corpora**, March
- Pawley, A. and Syder, F. H. (1983), “**Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency**”. in Richards, J. C. and Schmidt, R. W. (eds.) (1983): 191-226.
- Pawley Andrew, (2001) International Journal of Lexicography*, Volume 14, Issue 2, June, Pages 122–134.
- Pinker, Steven, (1994), **Lingua**, How could a child use verb syntax to learn verb semantics, Science Direct, Vol 92, 377-410.
- Reppen, R. (2009), Exploring L1 and L2 writing development through collocations: A corpus-based look. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language* (pp. 49-59). Basingstoke: Palgrave Macmillan.
- Romer, U. (2010). Establishing the phraseological profile of a text type: the construction of meaning in academic book reviews. *English Text Construction*,3,
- Salazar, D. J. L. (2011), **Lexical bundles in scientific English: A corpus-based study of native and non-native writing**. Ph.D. dissertation, Universitat de Barcelona. Available; http://www.tdx.cat/bitstream/handle/10803/52083/DJLS_DISSERTATION.pdf?sequence=1
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.). **Formulaic sequences** (pp. 1-22). Amsterdam: John Benjamins.
- Selinker, L,(1979). “**Contrastive analysis, errors analysis, Interlanguage: earliest experimental record**”. In D. Nehls, ed., *Studies in contrastive linguistics and error analysis*. Heidelberg: Groos. pp 287
- Shefrin, H. (2001), Behavioral corporate finance. *Journal of Applied Corporate Finance* 14, 113–124

- Sinclair, John, (1991), **Corpus, Concordance, Collocation**, Oxford University Press, Oxford
- _____ (1998), The Lexical Item. In: Weigand, Edda (ed.): **Contrastive Lexical Semantics**, 1–24. Amsterdam, Philadelphia: Benjamins.
- Sinclair, J., S. Jones & R. Daley (1970). **English collocation studies: The OSTI Report**. (edited by Ramesh Krishnamurthy 2004) London: Continuum.
- Sonomura, Marion O, (1996), Idiomaticity in the Basic Writing of American English. Formulas and Idioms in the Writing of Multilingual and Creole-Speaking Community College Students in Hawaii. New York:
- Tench Ralph, (2003) "Public relations writing – a genre-based model", **Corporate Communications: An International Journal**, Vol. 8 Issue: 2, pp.139 146.
- Tversky, A. and Kahnman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185, 1124-1131.
- Wei, Y. & Lei, L. (2011), Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC Journal*, 42(2), 155-166.
- Winter, G. (2000). A comparative discussion of the notion of validity in qualitative and quantitative research. *The Qualitative Report*, 4(3&4). Retrieved February 25, 1998, from <http://www.nova.edu/ssss/QR/QR4-3/winter.html>

CURRICULUM VITAE

Hakan TAŞKAYA was born in 1990 in Tarsus/Mersin. He finished primary and secondary schools in Sarihamzalı Secondary School in Adana in 2001 and 2004 respectively. He finished 19 Mayıs High School in Adana in 2008. He graduated from Department of English Language Teaching in Akdeniz University in 2013. He has been working as an English teacher since 2013. He worked in Artvin and now works in Rize. He started his master's degree in Karadeniz Technical University Applied Linguistic programme in 2015.

He is married and has one-year-old of a daughter. He can speak English fluently.