

**NEAR EAST UNIVERSITY  
INSTITUTE OF GRADUATE STUDIES  
DEPARTMENT OF ENGLISH LANGUAGE TEACHING**

**THE EMPLOYMENT OF LEXICAL BUNDLES IN ENGLISH PHD  
DISSERTATIONS WRITTEN BY ENGLISH-SPEAKING AND KURDISH-  
SPEAKING WRITERS**

**MASTER THESIS**

**Atta AHMED**

**NICOSIA**

**June, 2021**

**NEAR EAST UNIVERSITY  
INSTITUTE OF GRADUATE STUDIES  
DEPARTMENT OF ENGLISH LANGUAGE TEACHING**

**THE EMPLOYMENT OF LEXICAL BUNDLES IN ENGLISH PHD  
DISSERTATIONS WRITTEN BY ENGLISH-SPEAKING AND KURDISH-  
SPEAKING WRITERS**

**MASTER THESIS**

**Atta AHMED**

**Supervisor  
Dr Aida ARIANNEJAD**

**NICOSIA  
June, 2021**

## Approval

We certify that we have read the thesis submitted by Atta Ahmed “**The Employment of Lexical Bundles in English PhD Dissertations Written by English-Speaking and Kurdish-Speaking Writers**” and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Educational Sciences.

Head of the Committee: Assoc. Prof. Dr. Çise ÇAVUŞOĞLU

Supervisor: Dr. Aida ARIANNEJAD

Committee Member: Asst. Prof. Dr. Yasemin ÇETEREİSİ

Head of the Department: Assoc. Prof. Dr. Çise ÇAVUŞOĞLU

Approved by the Institute of Graduate Studies

...../...../2021

.....

Prof. Dr. K. Hüsnü Can BAŞER

Director of the Institute of Graduate Studies

## **Declaration**

I, hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that I have fully cited and referenced all material and results that are not original to this work.

Atta A. Ahmed

July 13, 2021

## **Acknowledgements**

I would like to show the deepest appreciation to my supervisor, Dr Aida Ariannejad, for her invaluable advice, direction, and feedback during my thesis project. I am thankful to my parents for their support and warm regards, and always made themselves available for me. I would also like to acknowledge the assistance of my friend, Lana, for her invaluable comments and suggestions which helped enrich the thesis. Last but not least, my special gratitude would go to my wife, Shajwan, for her encouragement, support, and patience during writing my thesis.

## **Dedication**

This thesis is dedicated to my little angel, Katuna, my family, friends, teachers, and anyone who is interested in learning.

## **Abstract**

### **The Employment of Lexical Bundles in English PhD Dissertations Written by English-Speaking and Kurdish-Speaking Writers**

**Ahmed, Atta**

**MA, Department of English Language Teaching**

**Supervisor: Dr. Aida Ariannejad**

**June, 2021, 82 pages**

Lexical bundles have become the point of interest in applied linguistics studies. They frequently occur in groups of words between three to five words. They are pertained to clusters, chunks, multiword sequences, lexical phrases, fixed expressions, and prefabricated patterns. This study was aimed to explore and compare the employment of lexical bundles in English PhD dissertations by English-speaking and Kurdish-speaking writers in the disciplines of biology and linguistics. To do so, the study followed a mixed-method approach. The compiled corpora included forty dissertations written in English. The corpora included biology by English-speaking writers (71,074-word), biology by Kurdish-speaking writers (106,317-word), linguistics by English-speaking writers (133,328-word), and linguistics by Kurdish-speaking writers (152,259-word). Wordsmith 6.0 tool is used to analyze the frequency of the 3-word and 4-word bundles. The results showed that Kurdish-speaking writers used more 3-word and 4-word lexical bundles compared to English-speaking writers. The corpus of linguistics discipline included more 3-word and 4-word lexical bundles than the corpus of biology discipline. Concerning the grammatical structure, Kurdish-speaking writers used more structural patters than English-speaking writers. The cross-disciplinary analysis of the data revealed that the writers in linguistics discipline are more dominant in the usage of structural patterns. Regarding the functional categories, the Kurdish-speaking writers utilized a higher percentage of research-oriented functions than the English-speaking writers did, but the employment of text-oriented functions and participant-oriented functions of Kurdish-speaking writers were less frequent than the English-speaking writers. Concerning the cross-disciplinary analysis, writers in the biology discipline used a greater number of research-oriented and participant-oriented functions than the writers in linguistics discipline. Yet, the writers

in linguistics discipline used more text-oriented functions. This paper is considered of vital importance to both linguistics and biology writers, with the implication that it offers curriculum designers with the most commonly used lexical bundles with their grammatical structures and functions, and assists them in suggesting preferred terminologies for academic modules in linguistic and/or biology disciplines.

**Keywords:** Lexical bundles, corpus, academic discourse, academic genre, and discourse community.



## Öz

### **İngilizce Konuşan ve Kürtçe Konuşan Yazarlar Tarafından Yazılan İngilizce Doktora Tezlerinde Sözcük Demetlerinin Kullanılması**

**Ahmed, Atta**

**Yüksek Lisans, İngilizce Öğretmenliği Bölümü**

**Süpervizör: Dr. Aida Ariannejad**

**Haziran, 2021, 82 sayfa**

Sözcük demetleri uygulamalı dilbilim çalışmalarında ilgi odağı haline gelmiştir. Genellikle üç ila beş kelime arasındaki kelime gruplarında bulunurlar. Bunlar kümeler, öbekler, çok kelimeli diziler, sözcük öbekleri, sabit ifadeler ve önceden hazırlanmış kalıplarla ilgilidir. Bu çalışma, biyoloji ve dilbilim disiplinlerinde İngilizce konuşan ve Kürtçe konuşan yazarlar tarafından İngilizce doktora tezlerinde sözcük demetlerinin kullanımını araştırmayı ve karşılaştırmayı amaçlamıştır. Bunu yapmak için, çalışma karma bir yöntem yaklaşımı izledi. Derlenen külliyat İngilizce yazılmış kırk tez içeriyordu. Derlem, İngilizce konuşan yazarların biyolojisi (71.074 kelime), Kürtçe konuşan yazarların biyolojisi (106.317 kelime), İngilizce konuşan yazarların dilbilimi (133.328 kelime) ve Kürtçe konuşan yazarların dilbilimini (152.259 kelime) içeriyordu. . WordSmith 6.0 aracı, 3 kelimelik ve 4 kelimelik paketlerin sıklığını analiz etmek için kullanılır. Sonuçlar, Kürtçe konuşan yazarların İngilizce konuşan yazarlara göre daha fazla 3 ve 4 kelimelik sözcük grupları kullandığını göstermiştir. Dilbilim disiplininin külliyatı, biyoloji disiplininin külliyatından daha fazla 3 kelimelik ve 4 kelimelik sözlük demetleri içeriyordu. Dilbilgisi yapısına ilişkin olarak, Kürtçe konuşan yazarlar, İngilizce konuşan yazarlardan daha fazla yapısal kalıplar kullanmışlardır. Verilerin disiplinler arası analizi, dilbilim disiplinindeki yazarların yapısal kalıpların kullanımında daha baskın olduğunu ortaya koydu. İşlevsel kategorilerle ilgili olarak, Kürtçe konuşan yazarlar, İngilizce konuşan yazarlardan daha yüksek oranda araştırma odaklı işlevler kullanmışlardır, ancak Kürtçe konuşan yazarların metin yönelimli işlevleri ve katılımcı yönelimli işlevlerin kullanımı, Kürtçe konuşan yazarlara göre daha az sıklıkta görülmektedir. İngilizce konuşan yazarlar. Disiplinler arası analizle ilgili olarak, biyoloji disiplinindeki yazarlar, dilbilim alanındaki yazarlardan daha fazla sayıda araştırma odaklı ve katılımcı odaklı işlevler

kullanmışlardır. Ancak dilbilim alanındaki yazarlar daha çok metne yönelik işlevler kullanmışlardır. Bu makale, müfredat tasarımcılarına gramer yapıları ve işlevleriyle en sık kullanılan sözcük demetlerini sunduğu ve dilbilim ve/ veya biyoloji disiplinleri.

**Anahtar Kelimeler:** Sözcük paketleri, külliyat, akademik söylem, akademik tür ve söylem topluluğu

## Table of Contents

Approval.....	1
Declaration .....	2
Acknowledgements .....	3
Dedication .....	4
Abstract .....	5
Öz .....	7
Table of Contents .....	9
List of Appendices .....	11
List of Tables .....	12
List of Abbreviations .....	13

### CHAPTER I

Introduction.....	14
Background of the Study.....	14
Statement of the Problem.....	17
Aim of the Study .....	18
Significance of the Study .....	18
Limitations and Delimitations of the Study .....	19
Definition of the Key Terms .....	19

### CHAPTER II

Literature Review.....	20
Academic Discourse and Academic Discourse Communities.....	20
Academic Genre.....	21
Formulaic Language.....	22
Lexical Bundles.....	23
Related Studies on Lexical Bundles.....	24
Differences between native and non-native writers .....	24
Cross-disciplinary studies .....	26
Differences between expert and novice writers .....	27
Cross-linguistic studies .....	27

### CHAPTER III

Methodology .....	29
Design of the Study.....	29
Corpus Characteristics .....	30
Corpus Compilation Procedure.....	31
Data Analysis Tools.....	32
Data Analysis Method.....	32
Issues of Reliability and Validity.....	33

### CHAPTER IV

Result and Discussion .....	35
Frequency of Lexical Bundles .....	35
The variations between ES and KS writers in using lexical bundles:.....	35
The cross-disciplinary analysis of lexical bundles:.....	41
Structural Analysis of Lexical Bundles.....	46
The structural analysis of lexical bundles between the ES and KS writers: ....	46
The structural analysis of lexical bundles across the disciplines: .....	49
Functional Analysis of Lexical Bundles .....	51
The functional analysis of lexical bundles between the ES and KS writers: ...	51
The functional analysis of lexical bundles across the disciplines:.....	54

### CHAPTER V

Conclusion and Recommendations .....	58
Summary of the Main Results.....	58
Implications.....	59
Recommendations for Further Research.....	60
References .....	61
List of Appendices .....	66
Plagiarism Report.....	80

## **List of Appendices**

<b>Appendix A.</b> 3-word lexical bundles in biology corpus written by ES writers .....	65
<b>Appendix B.</b> 3-word lexical bundles in biology corpus written by KS writers .....	66
<b>Appendix C.</b> 3-word lexical bundles in linguistics corpus written by ES writers ....	68
<b>Appendix D.</b> 3-word lexical bundles in linguistics corpus written by KS writers....	72
<b>Appendix E.</b> 4-word lexical bundles in biology corpus written by ES writers.....	75
<b>Appendix F.</b> 4-word lexical bundles in biology corpus written by KS writers .....	76
<b>Appendix G.</b> 4-word lexical bundles in linguistics corpus written by ES writers ....	77
<b>Appendix H.</b> 4-word lexical bundles in linguistics corpus written by KS writers....	78

## List of Tables

<b>Table 3.1</b> Published PhD Dissertations in Biology and Linguistics Disciplines.....	30
<b>Table 4.1</b> 3-word Bundles in the Corpora of ES and KS Writers .....	36
<b>Table 4.2</b> 4-word Bundles in the Corpora of ES and KS Writers. ....	38
<b>Table 4.3</b> 3-word Bundles in the Corpora of Biology and Linguistics Disciplines ..	41
<b>Table 4.4</b> 4-word Bundles in the Corpora of Biology and Linguistics Disciplines ..	43
<b>Table 4.5</b> Main Structures of Bundles between ES and KS Writers .....	46
<b>Table 4.6</b> Main Structures of Bundles between Biology and Linguistics Disciplines .....	49
<b>Table 4.7</b> Main Functions of Bundles between ES and KS Writers .....	51
<b>Table 4.8</b> Main Functions of Bundles between Biology and Linguistics Disciplines .....	54

## **List of Abbreviations**

**EAP:** English for Academic Purposes

**EFL:** English as a Foreign Language

**ES:** English Speaking

**KS:** Kurdish Speaking

**L1:** First Language

**L2:** Second Language

**LB:** Lexical Bundle

**LGSWE:** Longman Grammar of Spoken and Written English

**NF:** Normalization Frequency

**PhD:** Doctor of Philosophy

## **CHAPTER I**

### **Introduction**

This chapter introduces the major concepts of the thesis. It comprises of six sections: the background of the study, the problem of the study, the aim of the study, the significance of the study, limitation & delimitation of the study, and the chapter ends with the definition of the key terms.

#### **Background of the Study**

Studies on discourse analysis were firstly conducted in the 1960s with the focus on the quantitative investigations of the discursive characteristics of registers and language variations (Flowerdew, 2013). Swales (1990) states that research interest in academic discourse analysis increased successively. As a discipline, discourse analysis is defined by Paltridge (2012) as the study that investigates the language models and patterns throughout the texts, and examines the contextual relationship between language, culture, and society.

With regards to the basic concepts in discourse analysis, discourse community is considered as one of the remarkable areas of the field. In Flowerdew's explanation, discourse community is the community that has a recommended use of language patterns in which it interprets particular aspects of the pragmatics of both spoken and written discourses (2013). Discourse community is distinguished by its features among the members of a certain community. The participants of the community follow some specific patterns of a language, particular kinds of vocabulary, and terminology within a special related genre in particular kinds of highly qualified instances of professionalism (Paltridge, 2012). It is significant to mention that the comprehension of community members would depend on the community disciplines, in other words, certain disciplines have certain forms and patterns of language use they communicate through as discussed by Hyland (2009). For that reason, academic discourses are linked to social participations, epistemological views, and cognitive styles.

Within discourse community, academic discourse emerges in which it explains that there are some specific types of language practices that are employed in the university settings. These types of language practices would stand for academic events, observations, concepts and they support adequate communication among



insiders of the community. The language of the community is different from the language, which is normally used by the same group of people in the working places, at homes, or any other places (Hyland, 2009). One type of academic discourse that is widely used by the scholars is academic genre. Academic genre, according to Bhatia, is the study in which the linguistic behavioral functions are positioned in professional environments and institutions where disciplinary differences do not have an important influence on it (2014). Academic discourse has two major forms: spoken and written. This study focuses on the written academic genre. The written academic genre includes texts such as academic articles, textbooks, master's theses, and PhD dissertations that have recently been the point of interest by the academic discourse analysts (Hyland, 2008a).

What makes academic genre analysis important in doctoral dissertations is that PhD dissertations are considered the prototypical sample of academic genre with rich content for analysis. Furthermore, doctoral dissertations are the publishable compilation of research articles (Dong, 1998). Thompson (2005) believes that it is impossible to indicate the exact components of PhD genre across and within disciplines. Moreover, Hyland asserts that researches of PhD dissertations have emphasized their macro-structures (2009). Different academic discursive features are investigated such as meta-discourse markers, conjunctures, formulaic sequences, and lexical bundles. One of the linguistic features that have been studied recently is lexical bundles.

Baker and Ellece (2011, p.68) defined lexical bundles, in "Key Terms in Discourse Analysis", as a group of three to five words frequently and naturally used in a language. Lexical bundles, in certain instances, are pertained to as clusters, chunks, multiword sequences, lexical phrases, formulas, routines, fixed expressions, and prefabricated patterns which they have the propensity to be recognized since they link two structural units, for instance, *the lack of the, I don't know if, I don't want you to*. The term lexical bundles were firstly introduced by Biber and his colleagues in the "Longman Grammar of Spoken and Written English" (LGSWE) in 1999 as a group of words that come together at a minimum of ten times per million words. Hence, Biber and Barbieri (2007) stated that there are distinctive features between multi-word idiomatic expressions and lexical bundles in four various manners: these are continuous, grammatically incomplete, meaningfully transparent, and frequently driven. The analysis of expert researchers provided a basic

understanding of which lexical bundles are discipline-specific and share features more broadly. Biber and Barbieri further differentiated lexical bundles according to genres and disciplines concerning frequency, structural type, and functional usage. The natural appearance of these bundles demonstrated the distinctions across various genres such as frequent sequences of three-or more-word bundles that follow in high frequency through texts (2007). Lexical bundles are one of the types of formulaic sequences. Lexical bundles studies are conducted from the following major forms of analyses; Cross-linguistic, cross-disciplinary, and the differences between native and non-native language users. The current study has chosen the last two areas of research to analyze.

Studies on the differences between native and non-native users of lexical bundles are determined by comparative investigations of the employment of lexical bundles in a specific genre between two languages. Researches have revealed that studies in different disciplines conducted by native and non-native researchers show variations in the employment of lexical bundles. This is validated in Cortes's study (2004) which shows that various disciplines have bundle preferences according to their fields and their languages. Byrd and Coxhead (2010), on the other hand, further solidify this claim and demonstrate the inconsistency of the functions of lexical bundles in interdisciplinary fields. Their investigation suggested that native and non-native writers in different disciplines have different uses of lexical bundles.

Cross-disciplinary studies compares the academic genres written across different fields. Research on lexical bundles in the written academic genre has demonstrated the differences among various disciplines in academia. Academic genre analysis studies as Swales (1990), in "Genre Analysis: English in Academic and Research Settings", explains in order to join academia, writers and researchers from multi-disciplinary fields should be familiar with conventions and values employed by members of the discourse community. Hyland (2008a) examined and categorized the academic genre analysis into three genres: academic research articles, master's theses, and doctoral dissertations to see how lexical bundles are disseminated across each of the aforementioned disciplines. The examination of Hyland's study indicated the centrality of the word bundles in the academic discourse community and they also provide a significant mechanism of distinguishing disciplinary variation of written texts. In the viewpoint of Biber et.al. (2004), each genre has a certain language system of communication of the jargon and

terminology it contains. To identify this reality, they compared two academic genres – classroom teaching and textbooks. Their analysis showed that the academic genres consist of various types of lexical bundles.

### **Statement of the Problem**

Kurdish language is one of the languages that belongs to the Iranian language group within Indo-European language family. It is not recognized as a standard language. From the structural analysis perspective, Kurdish language is a subject headed language in which it generally follows SOV pattern (Khalid, 2020).

In the recent decades, the Kurdistan government scholarship programs funded a group of Kurdish researchers and scholars to study in English countries for postgraduate studies in different disciplines, among those disciplines are biology and linguistics. These groups of Kurdish researchers carried out researches in English language. From the discursive point of view, there is probability that the first language of the researchers had impact on the second language. As a result, the current study attempts to explore and compare the use of lexical bundles between English-speaking and Kurdish-speaking writers.

Researches have shown that non-native writers find it difficult to write fluent and natural texts similar to the ones written by native writers and there is always influence of the first language on the second language (Durrant & Schmitt, 2009). This is due to the lack of information regarding the linguistic devices such as lexical bundles, which can improve the quality and fluency of the texts.

This study increases the awareness of non-native researchers and writers to enhance their understanding of challenges that they encounter to use lexical bundles, and raising their awareness to the way ES writers employ the lexical bundles. Studies on the employment of lexical bundles across different academic discipline is a growing body of research and there are many areas that are required to be further investigated (Hyland, 2008a). Based on the personal observation of the researcher, this study hypothesizes that KS writers, who write in English, have limited knowledge about the common uses of academic lexical bundles and their frequencies in their dissertations in both fields of biology and linguistics as Salazar (2014) explained that non-native writers face difficulties when writing effective and precise academic prose in English. This might be due to their underuse, overuse, and/or misuse of the bundles in their studies.

Another issue is the deficiency of researches carried out on the preceding topic by researchers of the Kurdish community. To the best knowledge of the researcher, no study on a similar topic has been conducted in the universities of Kurdistan, neither the current affiliation.

### **Aim of the Study**

This study aims to explore and compare the employment of lexical bundles used in English PhD dissertations written by the ES writers and KS writers in the fields of biology and linguistics. The study attempts to pursue the syntactic and functional connections between similar bundles. This comparative study demonstrates the native and non-native differences in identifying and using the lexical bundles.

The present study seeks to answer the following questions:

1. What are the most frequently used lexical bundles in English PhD dissertations written by ES and KS writers in the discipline of biology and linguistics?
2. What are the structural characteristics of lexical bundles in English PhD dissertations written by ES and KS writers in the discipline of biology and linguistics?
3. What are the functional characteristics of lexical bundles in English PhD dissertations written by ES and KS writers in the discipline of biology and linguistics?

### **Significance of the Study**

This thesis can contribute to the importance of lexical bundles in academic writing. The implication of this study can be beneficial for EFL classroom teaching by focusing on teaching and learning of these word strings. This is due to the vital role the lexical bundles play in providing a fluent linguistic production in written discourse. These bundles help Kurdish writers to improve their academic writing skills through improving their fluency and increasing the quality of their texts via using these elements of natural academic English that is used by native speakers in their fields. This study can also further enrich the native English academic researchers' writing skills since the use of lexical bundles is the focal point of interest by the writers and this is a recent topic in research field. Moreover, this thesis

serves to guide material developer – people who outline and develop teaching materials that helps learners to boost their understanding of subjects and performances – to further incorporate lexical bundles into language skills’ course books.

### **Limitations and Delimitations of the Study**

In the limitation of this study, several factors were out of the control of the researcher such as the lack of time that restrained the researcher to enlarge the corpus. The lack of access to the updated sources is another limitation; since it was nearly impossible to find studies in both fields done by Kurdish EFL researchers. Another obstacle was the shortage of technological resources such as a full functioning university website that stored researches.

The delimitation of the study is that the focus was only on two disciplines namely, biology and linguistics. Likewise, the researcher sheds light on PhD dissertations ignoring other academic studies. Furthermore, within the chosen dissertations, only the introductions and the literature reviews were concentrated on. Moreover, for future research, the corpora must be larger.

### **Definition of the Key Terms**

1. **Lexical bundles:** lexical bundles are repetitive groups of words in a text that normally occur together in bundles of three to five words (Biber et al., 2004).
2. **Corpus:** a corpus is a compilation of written texts that are gathered on a computer that is beneficial to show how language is used (Allen, 2010).
3. **Academic discourse:** Academic discourse is a specific type of language practice that is employed by the members of an academic community in academic events, concepts, and observations (Hyland, 2009).
4. **Academic genre:** academic genre is a particular approach of language use. It is used when language practitioners get things done according to their spoken and written discourses (Paltridge, 2012).
5. **Discourse community:** Discourse community is a community that has a recommended use of language for language users in which it shows particular aspects of the pragmatics of the text (Flowerdew, 2013).

## **CHAPTER II**

### **Literature Review**

This chapter aims to review the literature conducted on the current topic. It consists of five major sections: academic discourse & academic discourse communities, academic genre, formulaic languages, and lexical bundles.

#### **Academic Discourse and Academic Discourse Communities**

There has been a growing body of research on academic discourse and discourse community. Academic discourse is viewed as one of the key elements of discourse analysis. It is defined as a particular type of language used among the language users in academic environments, such as, academic debates, events, observations, presentations, and seminars (Hyland, 2009). The academic discourse used by members of engineering, marketing, literature, and science are different from each other in terms of vocabulary, utterances, formulaic languages, and expressions (Dontcheva-Navatilova, 2012).

Furthermore, Hyland (2006) stated that members of different disciplines are oftentimes characterized by their assorted topics, different approaches and methodologies, and certain types of specific subject areas. In this way, language users of the community employ various discursive features, diverse prospects, and different ways of communication in line with the variety of disciplines. For instance, to learn a discipline, members need not only to communicate within the community but also need to use the preferred jargon language of that discipline. And this is accomplished via building knowledge, educating students, and spreading ideas.

From this perspective, researchers, such as Güngör and Usyal (2016) have shown that academic discourse is formed from the social roles and relationship between the language users, which generates knowledge that sustains in the university contexts. Furthermore, other studies revealed that one of the most common features of academic discourse is composed of a set of considerably stylized phrases, and for this reason, proficient writers use genre-specific collocation (Paltridge, 2012). Hyland (2009) provides a more comprehensive description for this matter as he demonstrates the construction of language, which is based on different sets of grammatical usages, lexical words, and rhetorical preferences to a particular discourse community. In this regard, it would seem reasonable; then, that the

prevalent characteristics of multiword expressions propose that language experts use these chunks of words in their academic writings (Hyland, 2012).

The concept of a discourse community is one of the remarkable areas of academic discourse research. The participants of the community follow some specific linguistic patterns, particular vocabulary items, and terminologies within a genre (Paltridge, 2012). This justifies academic discourses being linked to social participations, epistemological views, and cognitive styles (Hyland, 2009).

Researchers, such as Esfandiari and Barbary (2017) have almost exclusively focused on the academic discourse community and demonstrated that the enhancements of phraseological and grammatical aspects of language are linked to the practice of language users. As a matter of fact, academic disciplines have certain academic discursive features that lead to language preferences by writers in academic fields. This is concisely the reason behind the writers' attempts to convey their texts in such particular ways which are expected to be acknowledged by the readers within the community (Hyland, 2009). The discussion on the prominent speaking and writing tasks is further discussed by Hyland (2006); he elaborates, to learn to speak and write within the community is to engage with other members of the group as they are called the insiders of the community (2009). The idea of discourse community grasps attention to the fact that in a community, language users communicate with other participants in the corresponding societal groups, utilize particular norms, and categorizations.

### **Academic Genre**

Genre analysis is an instructive method of discourse analysis which investigates specific socio-rhetorical linguistic features used in a particular group of texts among the members of a specific discourse community (Hyland, 2008b; Paltridge, 2012; Swales, 1990). Bhatia (1999) explained that genre practice is very nearly similar to a chess game in which both have their conventions and regulations. To set up a genre, participants as readers and writers play skillfully to manipulate the game successfully. Learning the rules of the game is comparable with learning the language which is manipulating and exploiting the rules to satisfy disciplinary and professional considerations.

Paltridge (2012) further explained genre as a specific approach of language use. Language practitioners get things done according to their spoken and written

discourses. Genres are meant to express those areas of work which it undertakes through the use of language. In addition, concerning the fact that there is certainly impact of genre on language teaching, Paltridge (1997) believed that genre plays a vital role on English language teaching either as a first language or as a second or foreign language. He further focused on the value of the discipline of teaching (EAP) English for academic purposes. It is important to mention that there are some other areas of genre studies such as linguistic anthropology, ethnography of communication, conversational analysis, rhetorical studies, literary theory, the sociology of language, and applied linguistics. Furthermore, Devitt (2004) has pinpointed that human interaction possibly relies on the genre, which is used by the community. For example, to allow somebody to talk or dishearten someone from uttering something which is not permitted in the community.

Concerning the characteristics of academic writing, Hyland (2009) indicated that the characteristics of writing research are considered as one of the essential key elements of postgraduate education. One of the most commonly discussed academic genres is PhD dissertations. PhD dissertation may differ in their length, the scope of the research, the amount of time, the energy spent, and the aims of the papers. He also demonstrated that it is challenging to study the characteristics of doctoral dissertations since there is a significant alteration in the genre (2009).

### **Formulaic Language**

Numerous researches have been conducted during the past few decades on the recurrent multi-word units (Wray, 2002). First, the research on formulaic language, as Hyland (2008b) asserts, belonged to the attempts made by Jespersen (1924). Formulaic languages are defined by Wray and Perkins as groups of words that appear continuously or intermittently in a text (2000). In other terms, formulaic language can be defined as a broad term that carries a wide range of multi-word strings. The consequences of using these strings have two significant functions; their communicative role and their abilities making comprehension easier. This leads to the point of what makes formulaic language of significant interest is that not only they have their own syntactic and semantic meaning depending on the context, but they also require less cognitive effort on the part of the writers since their usages come naturally/ spontaneously. That is why learning and appropriately using word chunks in a language is important in language acquisition (Carrol & Coklin, 2020).



Further, Ädel and Erman (2012) stated that one of the characteristics of recurrent word combinations is to provide a particular set of connection among the members of the discourse community. While using a language, certain words, phrases, and expressions that are associated with particular styles, groups, and values are chosen. Additionally, formulaic language is considered as a major component of written discourse. Researchers, like Wray and Perkins (2000), have shown that one-third to half of language components are composed of formulaic languages. It can also be a vital part of spoken discourses. In their study, Wray and Perkins (2000) demonstrated that approximately 70% of native adult speakers of the English language make use of formulaicity in their language naturally. They further stated that a vast amount of the combinations are some repetitive pieces of larger structures (e.g., *the end of the, in a, out of the*). These studies, thus, showed that the functions, forms, and dissemination of the word chunks throughout the language is constantly changing. Furthermore, these changes would also vary in different discipline and across various academic genres (Conklin & Schmitt, 2008).

### **Lexical Bundles**

The notion of lexical bundles has been introduced firstly by Biber and his colleagues in 1999 in their book “Longman Grammar of Spoken and Written English” (LGSWE), which provided a list of chunks that come together. Lexical bundles are described as continuous chunks that are grammatically lacking, meaningfully apparent, and recurrently used (Biber & Barbieri, 2007; Hyland, 2008). For instance (e.g., *I don't know if, I just want to, the lack of the, or I don't want you to*). These linguistic elements generally do not have a complete grammatical structure, neither do they perform an idiomatic meaning, but rather in both written and spoken texts, they carry out discourse functions (Baker & Ellece, 2011). The minimum frequency of such chunks in a compiled corpus was expected to be 10 per a million words. Biber and his colleagues discussed that the occurrences of three 3-word lexical bundles could reach up to 60000 times and 4-word bundles over 5000 times per million words in the academic prose. Moreover, lexical bundles should be seen for at least between five or more various texts; otherwise, they are not considered as lexical bundles.

Concerning lexical bundles' usages, researches have shown that LBs differ from one discipline to another, and that experts and novice writers use them

differently (Wood, 2015). For instance, Cortes (2004, 2013) noted that professionals in both disciplines of applied linguistics and literature present a clear insight into which lexical bundles are properly used for their fields. Moreover, she stated that LBs' structural patterns could also differ in an academic discipline. What is more is that Cortes analyzed texts written by experts and arrived at the conclusion that the experts recognize which lexical bundles are discipline specific. The use of bundles between two groups differed in terms of frequencies, structural types, and functional usages.

### **Related Studies on Lexical Bundles**

Earlier researches on lexical bundles in the written academic genre have demonstrated the differences between first language L1 and second language L2 (Chen & Baker, 2010), differences among cross-academic-disciplines (Cortes, 2004; Hyland, 2008), and differences between experts and novice writers (Chen & Baker, 2010), in terms of frequencies, structural & functional analyses.

#### ***Differences between native and non-native writers***

There has been a numerous body of study to investigate the use of lexical bundles between native and non-native writers. A study has been undertaken by Ädel and Erman (2012), investigated lexical bundles used by the Swedish EFL writers and native English writers. The researchers reported that the Swedish EFL writers used a smaller number of lexical bundles than the native English ones in terms of frequency. They also investigated the function of the lexical bundles in their corpora, whereas the Swedish EFL writers utilized less participant-oriented bundles and more text-oriented bundles than the English native writers.

Others like, Güngör and Uysal (2016) compared the structures and functions of lexical bundles used by native English and Turkish EFL writers in research articles. The study was aimed to point out the overuse, misuse, and underuse of the lexical bundles by the Turkish non-native writers and compared them to native English writers. The finding of the study displayed that the Turkish EFL writers prefer to use clausal or verb phrase bundles while native English ones employed more prepositional and noun phrase bundles.

Amirian et. al. (2013) researched the use of lexical bundles between Iranian EFL and English native master students in their theses in the field of applied linguistics. Their finding suggested that a smaller number of lexical bundles was found in the native English theses than the non-native ones. The prepositional phrases are considered as one of the most used phrases in the Iranian corpus more than the native English corpus '*on the other hand*'. Esfandiari and Barbary (2017) have carried out research to investigate the use of lexical bundles between Persian writers and English writers in the field of psychology in the research article genre. The structures, functions, and frequencies have been analyzed in 4-, 5- and 6-word bundles. The English writers preferred to use more 4-and 5-word lexical bundles than the Persians. In terms of structure and function, the Persian writers used fewer lexical bundles than the English ones.

Other researchers, such as Kashiha and Heng (2013), compared transcripts of 24 academic lectures of four disciplines (chemistry, computer, and engineering) from hard sciences and (politics, law, and English language teacher education) from soft sciences in classroom discussion. The finding of their study suggested that hard sciences lectures employed a larger number of lexical bundles between native English and Malaysian non-native speakers. The research consisted of the transcript of 26 discussion sessions of both native and non-native corpora. The result of the study proposed that the corpus compilation of native speakers contained a more extensive word chunk in terms of frequency. The finding of the study also certified that speakers of English as a second language performed a smaller number of lexical bundles than native ones. As for the functional analysis, the lexical bundles in the Malaysian corpus have elevated using the stance expressions comparing to the English corpus using the discourse organizers as an insignificant functional use of lexical bundles.

In another study, a more comprehensive description can be found in Ucal's (2017) research that he employed a corpus-based analysis to compare the native English and Turkish non-native writers in research articles. The research resulted that Turkish non-native writers' utilization of three-word lexical bundles is more frequent than English ones; however, the study suggested a diversity in the usage of the bundle types between both corpora. It appears that Turkish authors employed fewer valid bundles comparing to native English authors in terms of frequency. The structural uses of Turkish non-native writers consisted of prepositional phrases such

as *with respect to, in order to, of the participants*, whereas native English writers' corpus contains more noun phrase fragments, for instance, *use of the, one of the, part of the*. The functional analysis corpus showed that lexical bundles by Turkish writers have a declination instance and discourse organization in comparison with native English writers.

### ***Cross-disciplinary studies***

There has been interest in academic genre analysis research to investigate the use of lexical bundles across disciplines. Cortes (2004) have conducted a study to compare two different disciplines – biology and history journals. The result of the study showed that the writers in biology employed fewer lexical bundles than the history writers. As for the functional use of the bundles, history writers utilized less (text-oriented bundles) than the biology writers. Furthermore, Hyland, (2008a) in his study, investigated the frequency, functions, and structures of 4-word lexical bundles in research articles, master's theses, and doctoral dissertations in four different disciplines – biology, electrical engineering, applied linguistics, and business studies. Hyland used a corpus linguistics methodology to collect the data. The finding of his study revealed that electrical engineering studies contained the most lexical bundles while biology recorded the least number of cases. In terms of functional analysis, writers in soft science field – linguistics utilized more research-oriented bundles than text-oriented and participant-oriented bundles.

Byrd and Coxhead (2010), in a corpus-based study, demonstrated the inconsistency in the functions of lexical bundles in inter-disciplinary fields. Science, law, commerce, and arts have been selected. The corpus compilation, 3.6 million words, consisted of seven subject areas in each discipline of academic written English classes. The result suggested that prepositional phrases are used widely by all disciplines for instance, adverbials phrases such as, in *the case of* or *on the basis of*. A series of recent studies have indicated the differences between cross-disciplines by native and non-native writers. Kashiha and Heng (2015) compared two different disciplines of chemistry and politics to see the structural variations in using lexical bundles and their frequencies. The compiled corpus included transcripts of 8 university lectures – 4 lectures for each discipline. The result of the study illustrated that both fields used fragments of prepositional phrases and noun phrases. According to the statistics, in comparison with chemistry lectures, more lexical bundles were

used in the politics discipline, which indicated the tendency of members of the politics departments to use lexical bundles in their lectures.

### *Differences between expert and novice writers*

Differentiation between novice and expert writers is another area of interest among researchers to investigate lexical bundles. Chen and Baker (2010) compared native and non-native experts and novice writers regarding their used of lexical bundles at functional and structural levels. The result of their study showed that non-native experts and native students' writing contained unexpectedly similar uses of lexical bundles. The researchers further explained that they each utilized considerably more discourse organizers and verb-phrase-based bundles than the native experts did. Using such features were discussed to be the characteristics of unsophisticated writing. However, the native expert writers showed a larger number of referential markers and noun-phrase-based bundles in their writings.

### *Cross-linguistic studies*

Various studies have been conducted on the employment of lexical bundles in cross-linguistics analysis. Granger (2014) conducted a cross-linguistic study to compare the use of lexical bundles in French and English languages. The study showed how lexical bundles are investigated in two different genres – newspaper editorials and parliamentary debates to evaluate the influence of genres and languages on the lexical bundles. A corpus-driven approach was used to conduct the analysis of the employment of the word bundles. The result demonstrated that French language writers preferred to use more lexical bundles than did the English language writers.

Furthermore, Roldán-Riejós and Grabowski (2019) also conducted a corpus-based study to analyze the use of lexical bundles across various linguistic genres and registers in English, Spanish, Polish, and Russian languages. The research indicated that any language has a special way of communication among the members of the community. In that study the corpus-based methodology was used to detect over 250,000 lexical combinations in different genres – product descriptions, technical reports, texts, and research articles.

The review of the literature illustrates those lexical bundles are used by both English native and non-native speakers, and expert and novice writers in line with

various uses of frequency, functional, and structural analyses. There is no clear-cut rule to oblige the language users to employ lexical bundles in a specific manner and it is also difficult to imagine the exact reason behind utilizing these bundles. Each of these abovementioned groups has their preferred rhetoric and lexical bundles variously according to their language ability, cultural backgrounds, and lexical norms.

## **CHAPTER III**

### **Methodology**

This chapter includes the following sections: the design of the study, the characteristics of the corpus, the data analysis method, the corpus compilation procedure, the data analysis tool, the data analysis method, and the issues related to reliability and validity.

#### **Design of the Study**

The design of the study follows a mixed methodology. It implies the integration of the two quantitative and qualitative methods in an individual research. Moreover, applying mixed-methods offers a more comprehensive analysis of the problem of the study (Dorney, 2007; Fraenkel & Wallen, 2009).

The collected data is analyzed using both quantitative and qualitative paradigms. Reasons to support choosing quantitative and qualitative technique as Dörnyei (2007) explains are 1) it can achieve the best result for the study, 2) selecting one of quantitative or qualitative would be too content-specific, and 3) it can prevent biasness of the results. Quantitative approach is used to analyze the frequency of the lexical bundles to see the distribution of word bundles across disciplines. On the other hand, the structural and functional patterns are investigated qualitatively in order to analyze the structural patterns and functional categories to see how both groups of writers employ them.

The structural analysis follows the taxonomy of Biber et al., which categorized the structures of lexical bundles into three headings; “lexical bundles that incorporate with verb phrase fragments, lexical bundles that incorporate dependent clause fragments, and lexical bundles that incorporate noun phrase and prepositional phrase fragments” (2004, p. 381). As far as the functional analysis is concerned, Hyland’s modification of the taxonomy originated by Biber et. al. (2004) is followed. The classification provides three broad sub-categories namely: “research-oriented bundles, text-oriented bundles, and participant-oriented bundles” (2008a, p. 13).

## Corpus Characteristics

This study focuses on English PhD dissertations written by ES and KS writers in two fields of linguistics and biology. The corpus was divided into four sub-categories including the biology written by English writers (71,074 words), the biology written by Kurdish writers (106,317 words), the linguistics written by native English writers (133,238 words), and the linguistics written by Kurdish writers (152,259 words). In total, all the corpora contained (462,888 words).

**Table 3.1**

*Published English PhD Dissertations in Biology and Linguistics Disciplines*

Disciplines	Number of dissertations	Mean Length of Texts in Words	Word Count
ES Biology Corpus	10	7,107	71,074
KS Biology Corpus	10	1,631.7	106,317
ES Linguistics Corpus	10	13,332.8	133,328
KS Linguistics Corpus	10	15,225.9	152,259
Total Number of words in corpora			462,888

It is significant to mention that the only two chapters that are selected to be analyzed are the introduction and the literature review. The reason to choose those two chapters is the similarity between introduction and literature review. At first, in terms of content, in the introduction chapter, a significant number of rhetorical attempts can be seen while writers made efforts to establish a research niche to substantiate the significance of the study. Along the same line, in the literature review chapter, the researchers seek to acknowledge what is novel to the social communities, and what other somehow similar research has been conducted in the field (Hyland, 2009). Another reason of the comparability of introduction and literature review is that the researchers present the core of the text to the readers in the introduction chapter. Meanwhile, the literature review section is a continuation of the introductory part that highlights further studies conducted in the area of research, and the researchers' attempts to pinpoint the gap of the study (Hasa, 2017).

This study investigated the lexical bundles used in two disciplines of biology and linguistics. Considering the disciplinary domains, linguistics and biology falls within two different scientific areas of hard sciences and soft sciences. The difference between these two fields is that hard sciences such as biology focus on



experimental proof and hypothesis formation, acceptance, and rejection. However, studies in soft sciences such as linguistics are introspection research, which are usually based on arguments and their acceptance or rejection (Hyland, 2009).

The PhD dissertations were all published between 2010 and 2020 to keep the corpora comparable. In this regard, the PhD dissertations were all written and published in the United Kingdom universities. ES and KS postgraduate students wrote the dissertations. There are two basic criteria to determine the case of nativeness of the English writers; 1) name and surname of the researchers, and 2) the name of the affiliations.

### **Corpus Compilation Procedure**

The gathered data consist of two corpora including the introduction and literature review sections of PhD dissertations written by the ES and KS writers in two disciplines of biology and linguistics.

In the first step, corpora were selected and downloaded from the British Library e-theses online service (<https://ethos.bl.uk>). There were some criteria for gathering the initial data, the dissertations were searched through the keyword biology and linguistics, and the timeline from 2010 to 2020 was set in the search engine of the website. In order to include the dissertations in the corpora, the content of the PhD studies was checked.

Some of the dissertations were excluded from the collected corpora; however, the titles contained the key words biology or linguistics. After checking the content and the departments where the studies were conducted, they were related to the departments of medicine, chemistry, physiology, politics, translation, and literature.

The dissertations were originally downloaded as pdf files, then an online converter software was used to change the format of the files to plain text (txt) to be later used in the concordancing software. The present study follows Sinclair's (1991) clean-text policy, hence, any extra information has been removed from the dissertations, for instance, header, footer, endnotes, page numbers, names of authors, figures, diagrams, graphs, tables, interviews, and quotes. This is due to the fact that each of these aforementioned parts would have impact on the result and would be counted as tokens in the database of the software.

It is necessary to report that there were differences in the size of ES and KS corpora. To solve this issue in the corpus size, McEnery and Hardie (2012) proposed

the process of normalization to be used. The number of examples of words in the entire corpus is divided by the corpus size, and then the result is multiplied by the base of the normalization. The formula below demonstrates the procedure of normalization frequencies:  $nf = (\text{the total number of examples of the words in the entire corpus} \div \text{size of the corpus}) \times (\text{normalization base})$ . Regarding the normalization process, only the lexical bundles with higher normalization than 1.0 per 10,000 words are selected.

### **Data Analysis Tools**

The tool which is used for analyzing the data in this study is the concordancing program of Wordsmith 6.0. This software helps researchers to identify the multiword expressions and their frequencies across texts.

### **Data Analysis Method**

The analysis of this study follows a corpus linguistics approach. It is defined as the investigation of texts in a particular language, dialect, and/or other varieties of languages using a corpus (Stefanowitsch, 2020; Tognini-Bonelli, 2001). Corpus linguistics is the study to analyze language and linguistic features in a corpus. It shows the possible ways, which languages are used in particular contexts and across contexts (Crawford & Csomay, 2016, p. 5).

According to Biber et. al. (1998), corpus-based analysis has the following common features; first, corpus-based analysis is an empirical approach based on which the actual patterns of language use are analyzed. Second, corpus-based analysis uses a large body of target language texts (a corpus) as the basis for data analysis. Third, it makes use of computer software to analyze the data. Fourth, it adopts both quantitative and qualitative techniques.

The result of the study is investigated through a non-parametrical statistical test – Spearman Brown Test, which is utilized to calculate the differences in the frequency of lexical bundles in both native and non-native corpora (Fraenkel & Wallen, 2009).

## Issues of Reliability and Validity

Reliability can be determined as the consistency of the measurements across the diversity of conditions in which the outcome should be achieved (Nunnally and Berntein, 1994). Reliability provides whether or not the statistics of the data collection has congruity and accuracy (Fraenkel & Wallen, 2009). Moreover, Conklin and Schmitt (2008) have illustrated that to determine and investigate formulaic language, notably lexical bundles, the issue of internal reliability is addressed, to accommodate internal reliability; all possible measurements should be applied uniformly.

In this study, intra-rater reliability was calculated. To evaluate it, the data was analyzed twice with a 5-month time interval (January-June). The correlation between two analyses resulted as follows; the 3-word bundles scored 98.1, as far as the 4-word bundles are concerned, the result showed 97.3 as the intra-rater reliability. Moreover, Spearman's rank correlation has been used to find the correlation coefficient between the (ES and KS) variables and (biology and linguistics) variables. Spearman-Brown test has been also used for the reliability coefficient. The resulting data indicates 0.99 correlation coefficient between ES and KS for both 3-word and 4-word bundles. Concerning the correlation coefficient between the cross-disciplinary aspect (biology and linguistics), the statistical results showed that 3-word bundles hold (0.99) correlation coefficient according to Spearman-Brown test reliability. While the 4-word bundles' result is (0.98). Based on the above results, the reliability of the current study is deemed to be "*excellent*".

As far as validity is concerned, this study has a high internal validity since the compiled corpora are true representatives of the genre of PhD dissertations. Furthermore, the corpora are comparable, as it was discussed earlier, from different perspectives. First, they have, to a great extent, the same structure. Second, the same chapters (introduction and literature review) from the dissertations are chosen to be analyzed. Third, the studies are from the same publication period 2010-2020. Fourth, the studies are all written by advanced level students, and under the supervision of experienced supervisors and editors.

All the selected dissertations were individually read and the contents were checked by the researcher in order to have a valid source of data analysis. So as to find suitable dissertations, the initial search was for 60 dissertations, where each one of them was carefully investigated through their cover pages and contents. The CVs

of the researchers were also searched online via google search to make sure they were all ES and KS postgraduate students.

## **CHAPTER IV**

### **Result and Discussion**

The current chapter focuses on the presentation of the results and discussion of the study. It consists of three sections: the frequency of the lexical bundles, the structural analysis, and the functional analysis of the lexical bundles— employed by ES and KS writers in the field of biology and linguistics. Each section in this chapter aims at providing the answer to the research questions of the present study.

In order to answer the research questions, the corpus was divided into four sub-categories the biology written by the ES writers & KS writers, and the linguistics corpus written by the ES writers & KS writers. The results of the corpora are analyzed in two different aspects; differences between ES and KS writers and variances between disciplines (cross-disciplinary). For this reason, forty PhD dissertations are compared and investigated.

#### **Frequency of Lexical Bundles**

The first research question seeks to find the frequency of the usages of lexical bundles by both ES and KS writers in the introductions and the literature reviews of dissertations written in English. The results showed variation in frequency of employment of bundles by both NS and KS writers. Moreover, variations were also detected from the cross-disciplinary perspective that is from biology and linguistics disciplines. The results are discussed in the following sections.

#### ***The variations between ES and KS writers in using lexical bundles:***

The most frequently occurred 3-word lexical bundles used by the Kurdish-speaking and Kurdish-speaking writers in the selected dissertations are presented in Table 2, where the shared bundles have been highlighted in bold letters.

**Table 4.1***3-word Bundles in the Corpora of ES and KS Writers*

3-Word Bundles Used by ES Writers	Frequency (per 10,000 words)	3- Word Bundles Used by KS writers	Frequency (per 10,000 words)
<b>as well as</b>	222 (20.08)	<b>as well as</b>	147 (12.26)
the use of	118 (8.86)	as shown in	90 (5.91)
<b>in order to</b>	93 (6.98)	<b>in terms of</b>	81 (5.32)
<b>in terms of</b>	64 (4.80)	the Kurdish language	78 (5.12)
properties of the	62 (4.65)	a result of	76 (7.15)
use of the	62 (4.65)	the expression of	71 (6.68)
in this thesis	58 (4.35)	in central Kurdish	67 (4.40)
of language change	57 (4.28)	a number of	64 (4.20)
the fact that	55 (4.13)	of the Kurdish	63 (4.14)
the active centre	50 (7.03)	<b>the presence of</b>	60 (5.64)
been shown to	48 (6.75)	<b>in order to</b>	55 (3.61)
the analysis of	42 (3.15)	in the Kurdish	54 (3.55)
of the input	40 (3.00)	there is no	53 (3.48)
of this thesis	40 (3.00)	in Iraqi Kurdistan	51 (3.35)
that it is	40 (3.00)	the hair follicle	51 (4.80)
well as the	40 (3.00)	as far as	46 (3.02)
of the vowel	39 (2.93)	in this study	46 (3.02)
<b>one of the</b>	39 (2.93)	<b>one of the</b>	46 (4.33)
the number of	39 (2.93)	in the past	45 (2.96)
<b>the presence of</b>	39 (5.49)	in addition to	44 (2.89)
the properties of	39 (2.93)	in the language	43 (2.82)
referred to as	38 (2.85)	in the following	38 (2.50)
the development of	38 (2.85)	loan words in	38 (2.50)
a set of	37 (2.78)	the dermal papilla	38 (3.57)
the bone marrow	37 (5.21)	due to the	37 (3.48)
Total	1436 (122.62)	Total	1482 (110.70)

The presented data illustrates the most commonly used lexical bundles by the ES and KS writers' corpora. In total, native writers used 1436 word-bundles that is 122.62 per 10,000 words, while non-native writers employed 1482 words to be exact 110.70 per 10,000 words. The KS writers utilized more word bundles than the ES one. However, the differences are not significant.

With regards to the similarities between the two corpora, the bundle *as well as* which acts as an interactive transition marker was used 222 times (20.08 per 10,000 words) by the natives. Likewise, in the opposite column as indicated, the non-natives frequently used the same bundle 147 times, that is 12.26 per 10,000 words.

Other frequent bundles, which were recurrently used by both groups are; *in terms of*, *in order to*, *one of the*, and *in presence of*. It is important to mention that the employment of all these shared bundles in the corpus of the English-speaking writers are less than the usages in the corpus of the Kurdish-speaking writers.

Concerning, the differences between both groups of writers, a number of hedges, boosters, transition markers, and frame markers are used. For instance, *the fact that* (55 time, 4.13 per 10,000 words) and *been shown to* (48, 6.75 per 10,000 words) which work as boosters used by English-speaking writers. Other examples such as *in this thesis* is used as frame marker.

In contrast, the corpus of the Kurdish-speaking writers consists of a group of examples different from those of English corpus. These are noun phrase and prepositional phrase examples. Such as *the Kurdish language*, *in central Kurdish*, *of the Kurdish*, *in the Kurdish*, and *in Iraqi Kurdistan* with the frequency of (313, i.e., 20.56 per 10,000 words). The keywords (Kurdish or Kurdistan) are frequently used in the abovementioned examples. The reason for having this frequency of the lexical bundles containing the preceding keywords in the corpus is due to the research trends that the KS researchers conducted their dissertations on the topics related to Kurdish language features.

As a matter of fact, amongst the 10 dissertations in the field of linguistics written by Kurdish researchers, 9 of them contained the keyword (Kurdish) in the title of the dissertations. This certifies that Kurdish researchers were definitely referring to the Kurdish language when making their linguistic analyses.

A group of 3-word bundles such as *the use of*, *properties of the*, *use of the*, *the fact that*, and *the analysis of* are examples from the corpus of ES writers with the frequency of (339 i.e., 25.44 per 10,000 words). This shows that ES writers employed variety of bundles types and structures in their writings.

Furthermore, KS writers utilized example of transition markers, i.e., *as far as*, *as a result*, and *in addition* (166, 13.06 per 10,000 words). It is worth mentioning that the least used bundles are *the bone marrow* by the natives, and *due to the* by the non-native. They lie at the last row in the list with the frequency of 37 times (5.21 and 3.48 per 10,000 word) in both corpora.

The forthcoming section is the employment of 4-word lexical bundles by ES and KS writers in their dissertations.

**Table 4.2**

*4-word Bundles in the Corpora of ES and KS Writers.*

4-Word Bundles Used by ES Writers	Frequency (per 10,000 words)	4- Word Bundles Used by KS writers	Frequency (per 10,000 words)
as well as the	48 (4.13)	on the other hand	86 (6.13)
<b>has been shown to</b>	27 (3.80)	as a result of	59 (4.87)
area of the vowel	24 (1.80)	the meaning of the	44 (2.89)
extent to which the	18 (1.35)	the use of the	36 (2.36)
<b>in the case of</b>	18 (1.35)	within the nominal phrase	31 (2.04)
levels of linguistic structure	18 (1.35)	of the Kurdish language	30 (1.97)
at the same time	17 (1.28)	the past stem of	27 (1.77)
distributional learning in infancy	17 (1.28)	on the basis of	26 (1.71)
it should be noted	17 (1.28)	<b>in the case of</b>	25 (1.64)
the formation of the	16 (2.25)	it might be argued	25 (1.64)
have been shown to	14 (1.97)	might be argued that	25 (1.64)
the RNA exit channel	14 (1.97)	the representation of the	24 (1.58)
the presence of the	12 (1.69)	transitivity of the verb	24 (1.58)
the structure of the	12 (1.69)	the central Kurdish dialect	23 (1.51)
automation for the biosciences	11 (1.55)	the number of the	23 (1.51)
by the presence of	11 (1.55)	an increase in the	22 (2.07)
into the active center	11 (1.55)	meaning of the idiom	20 (1.31)
is referred to as	11 (1.55)	stem of the verb	20 (1.31)
the majority of the	11 (1.55)	the fact that the	20 (1.31)
in the nervous system	10 (1.41)	a wide range of	19 (1.79)
in the regulation of	10 (1.41)	the antibacterial activity of	19 (1.79)
of the trigger loop	10 (1.41)	the expression level of	18 (1.69)
the binding of the	10 (1.41)	in the presence of	17 (1.60)
been shown to be	8 (1.13)	is one of the	17 (1.60)
it is thought that	8 (1.13)	<b>has been shown to</b>	16 (1.50)
Total	383 (53.89)	Total	680 (49.32)

Regarding the 4-word lexical bundles, the ES and KS writers employed a variety of uses of lexical bundles. In total, English writer utilized 383-word bundles (53.89 per 10,000 words), whereas the KS writers used 680 (49.32 per 10,000 words). This shows that the Kurdish-speaking writers' corpus consists of a greater



number of lexical bundles in comparison with their English counterparts. In addition, the transition marker *as well as the* was located on the top of the list of the English corpus with the frequency of 48 times, i.e., 4.13 per 10,000 words. On the other side, the top used lexical bundles in the corpus of KS writers is *on the other hand* which was used 86 times (6.13 per 10,000 words).

The shared similarity between the two corpora is the bundle *in the case of* with the frequency of 18 times (1.35 per 10,000 words) by the English-speaking writers, and 25 times (1.64 per 10,000 words) by the KS writers. The two corpora exhibit more differences than similarities. For instance, the lexical bundle *has been shown* was used by both ES and KS, but with significant difference in terms of frequency. These differences are because of several factors such as the difference between the languages and language families of each group, for instance the Kurdish language is a branch of Indo-Iranian language family, whereas the English language is a branch of Germanic languages. The native writers utilized it 27 times (3.80 per 10,000 words) making it the second most frequent bundle used. Whilst it was employed 16 times (1.50 per 10,000 words) by the non-natives placing it at the bottom of the Kurdish corpora. Furthermore, the native writers employed several prepositional phrases such as; *at the same time, by the presence of, into the active centre, in the nervous system, in the regulation of, and in the trigger loop*. In the same way, non-native writers used a number of prepositional phrases, such as *within the nominal phrase, of the Kurdish language, and in the presence of*. In total, the uses of the prepositional phrases in the Kurdish corpus are greater than those of English corpus.

What is intriguing about the results is that although the KS corpora have recorded a higher number of using lexical bundles in their dissertations in both 3-word and 4-word bundles, yet their normalization alters the results. In terms of 3-word lexical bundles, KS used bundles 1482 times in comparison to 1436 times by the ES writers. However, the normalization rate of the English-speaking writers is higher with 122.62 per 10,000 words. While the Kurdish-speaking writers' is 110.70 per 10,000 words. This contradiction of results is due to the word count of the corpora. In other words, the usage of KS writers per 10,000 words is smaller than ES writers since the English corpus includes less word counts than the Kurdish-speaking corpus. Conversely, the 4-word bundles results show 680 times usage in total by the KS against 383 times of usage by the ES writers. Even though with this significant

difference of lexical bundles (that is to say 297 times difference) yet, the normalization rate of the natives is higher with 53.89 per 10,000 words. While the non-natives' is 49.32 per 10,000 words. This inconsistency of results is due to the same abovementioned reason.

The current study's results are in line with Ucal's (2017) findings in that, similarly, the utilization of lexical bundles by non-native Turkish writers are also more recurrent than English writers. Other researchers that further reaffirm the results is Amirian et.al (2013). They concluded that a larger number of lexical bundles are used in non-native Persian theses compared to English theses. On the contrary, Esfandiari and Barbary's (2017) findings stand opposite to the current study's results. That is English writers had the tendency to use more 4-word bundles than the non-native Persian writers.

From the current study's analysis, it is concluded that the KS writers were significantly active in utilizing lexical bundles from both frequency and structural patterns perspective in comparison to their counter group, ES writers. This can be due to the Kurdish EFLs' urge to reach the professional adequacy of native writers, and their attempts to achieve this goal through their excessive use of lexical bundles in their writings. As well as the impact of language transfer, that affects the choice of words and expressions, since Kurdish language possesses a considerable number of bundles that automatically transfer to the Kurdish EFL writers' English language expressions.

***The cross-disciplinary analysis of lexical bundles:***

The most frequently occurred 3-word lexical bundles used in the disciplines of biology and linguistics corpora are presented in Table 4, where the shared bundles have been highlighted in bold letters.

**Table 4.3**

*3-word Bundles in the Corpora of Biology and Linguistics Discipline*

Biology Corpus	Frequency (per 10,000 words)	Linguistics Corpus	Frequency (per 10,000 words)
<b>as well as</b>	144 (15.97)	<b>as well as</b>	225 (16.37)
the presence of	99 (11.13)	in order to	148 (10.59)
the expression of	90 (9.35)	in terms of	142 (10.12)
been shown to	77 (9.48)	<b>the use of</b>	118 (8.86)
a number of	77 (8.03)	as shown in	90 (5.91)
in response to	77 (6.48)	the Kurdish language	78 (5.12)
the hair follicle	51 (4.80)	that it is	72 (5.10)
the active centre	50 (7.030)	in other words	69 (4.85)
it has been	50 (5.49)	in central Kurdish	67 (4.40)
<b>one of the</b>	46 (4.33)	a number of	64 (4.20)
of the cell	43 (4.93)	of the Kurdish	63 (4.14)
a result of	40 (3.76)	properties of the	62 (4.65)
the dermal papilla	38 (3.57)	use of the	62 (4.65)
the bone marrow	37 (5.21)	in this thesis	58 (4.35)
due to the	37 (3.48)	of language change	57 (4.28)
in addition to	36 (5.07)	the fact that	55 (4.13)
as a result	36 (3.39)	in the Kurdish	54 (3.55)
<b>the use of</b>	34 (3.20)	there is no	53 (3.48)
of the hair	33 (3.10)	in Iraqi Kurdistan	51 (3.335)
an increase in	32 (3.01)	as far as	46 (3.02)
has been shown	31 (4.36)	in this study	46 (3.02)
the formation of	30 (4.22)	the analysis of	42 (3.15)
the trigger loop	27 (3.80)	of the input	40 (3.00)
the majority of	25 (3.52)	of this thesis	40 (3.00)
the primer template	24 (3.38)	<b>one of the</b>	40 (3.00)
Total	1264 (140.08)	Total	1842 (130.31)

The presented data illustrates the most commonly used 3-word lexical bundles in the disciplines of biology and linguistics. In total, the biology corpus includes 1264 word-bundles that is 140.08 per 10,000 words, while the linguistics corpus includes 1842 word-bundles (130.31 per 10,000 words). The linguistics

discipline utilized more word bundles than the biology discipline. Nevertheless, the two fields of study have a considerable number of similarities and differences as discussed below.

Both disciplines show similar features in line with the employment of the lexical bundles, for instance, *as well as* is found at the top of both lists. In biology 144 times (15.97 per 10,000 words), and in linguistics with the frequency of 225 times (16.37 per 10,000 words). It is significant to mention that *as well as* in the linguistics corpus has the highest rate of frequency in usage than the rest of the lexical bundles in both disciplines. One of the distinctive features of common use of *as well as* is that both groups of writers attempt to use it to mean “*in addition to*”, “*too*”, “*also*”, and “*besides*”. Another salient feature is that almost all the uses of *as well as* are at the end of the end sentences. This certifies the fact that writers of both disciplines do not make an effort to employ other synonyms, but rather they use *as well as* most frequently.

Other frequent bundles, which were recurrently used in both fields are *the use of*, and *one of the*. The former bundle was used in biology with the frequency of 34 (3.20 per 10,000 words); however, the same bundle showed a noteworthy frequency rate of 118 (8.86 per 10,000 words) in the linguistics discipline. The latter bundle, *one of the*, show similar frequency in both fields. In biology 46 (4.33 per 10,000 words), and in linguistics 40 (3.00 per 10,000 words). Moreover, a number of 3-word bundles transition markers were used in the biology corpus such as *in addition to*, *in response to*, *due to the*, and *as a result*. In the meantime, linguistics writers used transition markers as well, for instance *in order to*, *as far as*, and *well as the*. Several boosters were used by the biology writers such as *been shown to* and *has been shown*. Whereas the corpus of the linguistics discipline included similar example of boosters for instance, *as shown in*.

Concerning the diversity in the cross-disciplinary fields, both disciplines used various noun phrases, prepositional phrases, conjunctures, and discourse markers. Albeit some lexical bundles have been commonly used in both corpora, the majority of the bundles were used differently. Groups of noun phrases were used by biology writers such as *the presence of*, *a number of*, *the dermal papilla*, *the bone marrow*, *the trigger loop*, *the majority of*, and *the primer template*. On the other hand, linguistics writers employed different noun phrases, such as *the Kurdish language*, *properties of the*, and *the analysis of*.

It is important to mention that a group 3-word lexical bundles are utilized in the corpus of biology such as *the hair follicle*, *the dermal papilla*, and *of the hair* with the frequency of (122, i.e., 11.47 per 10,000 words). The reason for having this high frequency of these bundles is in accordance with the research trends. The topics of some of the dissertations were on the dermatological issues and hair-related topics.

The forthcoming section is the employment of 4-word lexical bundles in both disciplines of biology and linguistics.

**Table 4.4**

*4-word Bundles in the Corpora of Biology and Linguistics Discipline*

Biology Corpus	Frequency (per 10,000 words)	Linguistics Corpus	Frequency (per 10,000 words)
<b>as a result of</b>	35 (3.29)	on the other hand	69 (4.53)
has been shown to	27 (3.80)	the meaning of the	44 (2.89)
an increase in the	22 (2.07)	as well as the	40 (3.00)
a wide range of	19 (1.79)	<b>the use of the</b>	36 (2.36)
the antibacterial activity of	19 (1.79)	within the nominal phrase	31 (2.04)
the expression level of	18 (1.69)	of the Kurdish language	30 (1.97)
in the presence of	17 (1.60)	the past stem of	27 (1.77)
<b>is one of the</b>	17 (1.60)	on the basis of	26 (1.71)
on the other hand	17 (1.60)	<b>in the case of</b>	25 (1.64)
the formation of the	16 (2.25)	it might be argued	25 (1.64)
<b>in the case of</b>	16 (1.50)	might be argued that	25 (1.64)
have been shown to	14 (1.97)	area of the vowel	24 (1.80)
the RNA exit channel	14 (1.97)	<b>as a result of</b>	24 (1.58)
in the expression level	13 (1.22)	the representation of the	24 (1.58)
the presence of the	12 (1.69)	transitivity of the verb	24 (1.58)
the structure of the	12 (1.69)	the central Kurdish dialect	23 (1.51)
in the absence of	12 (1.13)	the number of the	23 (1.51)
in the process of	12 (1.13)	meaning of the idiom	20 (1.31)
of the hair follicle	12 (1.13)	stem of the verb	20 (1.31)
the expression of the	12 (1.13)	the fact that the	20 (1.31)
<b>the use of the</b>	12 (1.13)	the size of the	20 (1.31)
automation for the biosciences	11 (1.55)	the structure of the	20 (1.31)
by the presence of	11 (1.55)	extent to which the	18 (1.18)
into the active center	11 (1.55)	<b>in the case of</b>	18 (1.18)
is referred to as	11 (1.55)	levels of linguistic structure	18 (1.18)
Total	392 (43.35)	Total	674 (45.37)

Table 5 shows that the 4-word lexical bundles in the cross-disciplines of biology and linguistics. In general, as it is shown, the writers in linguistics field have used a greater number of lexical bundles in comparison with their counterparts in the discipline of biology. The most commonly used lexical bundles in biology corpus is *as a result of*, 35 times (3.29 per 10,000 words). The employment of this expression as the most frequently used bundles might be due to the fact that biology writers prefer to use *as a result* more than other connecting words for concluding statements. Whereas, the most recurrently utilized lexical bundles in the discipline of linguistics is *on the other hand* with the frequency of 69 times (4.53 per 10,000 words). One of the possible reasons that writers in the field of linguistics used more transition words is related to their attempts in making preferences for the lexical bundles *on the other hand* as the most recurrently used which is used to convey contrastive view point, or to perceive the counterarguments. This is probably attributable to the background knowledge of the writers, or the tendency of the community members toward using such expressions. It is necessary to mention that the corpus of linguistics includes more word bundles than the corpus of biology.

Both disciplines share some similar lexical bundles for instance *as a result of* in biology corpus 35 times (3.29 per 10,000 words), whilst the same bundle type was used 24 (1.58 per 10,000 words) in the linguistics corpus. Another similarity across the disciplines is the employment of *the use of the* 12 times (1.13 per 10,000 words), in the biology corpus, and 36 (2.36 per 10,000 words) in the linguistics corpus. Furthermore, both groups of writers in biology and linguistics shared the word bundles *in the case of* with different frequencies. In biology, 16 time (1.50 per 10,000 words) and in linguistics 25 (1.64 per 10,000 words). Moreover, *the structure of the* bundle was also shared across both disciplines. In biology corpus, 12 times (1.69 per 10,000 words), and 20 times (1.31 per 10,000 words) in the linguistics corpus. The writers in both disciplines utilized noun phrases, *the antibacterial activity of*, *the RNA exist channel*, and *the presence of the* are examples in the biology discipline. At the same time, linguistics corpus includes a number of similar noun phrases such as *the meaning of the*, *the representation of the*, and *the central Kurdish dialect*.

Regarding the differences between the two fields, both corpora include word bundles that are used differently. For instance, biology writers used boosters such as *has been shown to* 27 times (3.80 per 10,000 words). On the contrary, linguistics

writers used interactive evidential discourse markers such as *might be argued that*, 25 times (1.64 per 10,000 words).

The cross-disciplinary analysis of lexical bundles also exhibits interesting results like the results found in analyzing native and non-native writers' use of lexical bundles. To be precise, the biology discipline has recorded a higher number of using 3-word lexical bundles with 1264 times usage in total, against 1842 times of usage in the Linguistics discipline. Even though with this significant difference of lexical bundles (that is to say 578 times difference) yet, the normalization rate of the biology discipline is higher with 140.08 per 10,000 words. While the Linguistics discipline is 130.31 per 10,000 words. This paradox of results is due to the word count of the corpora. In other words, the usage of the linguistics discipline per 10,000 words is smaller than the biology discipline since the linguistics corpus includes more word counts than the biology corpus.

Contrariwise, the 4-word bundles results show 392 times usage in total in the biology discipline against 674 times of usage in the linguistics discipline. This difference of lexical bundles (that is to say 282 times difference) along with the normalization rate of the linguistics discipline with 45.37 per 10,000 words, and biology discipline with 43.35 per 10,000 words.

It is concluded that linguistics writers use lexical bundles actively than biology ones. Since biology discipline use more technical language and biologically related terms whereas linguistics discipline uses more narrative and argumentative terms. Another reason for having the variation of the uses of the bundles between biology and linguistics is the corpus of linguistics was larger than the biology corpus. Moreover, the research trends had influence on the occurrences of the bundles, for instance the biology research trends such as *the dermal papilla*, *the hair follicle*, and *of the hair* are the examples from biology dissertations. Meanwhile, the research trends in linguistics corpus such as *the Kurdish language*, *in central Kurdish*, *of the Kurdish*, *in the Kurdish*, and *in Iraqi Kurdistan* are examples from the linguistics dissertations. This shows that the research trends in linguistics studies have a large portion of the corpus, and the Kurdish researchers were referring to Kurdish language while making linguistic analysis in their studies.

This study's results are in line with Cortes's (2004) and Hyland's (2008a) findings in that they both discovered that writers in biology discipline utilized fewer lexical bundles compared to their counter group writers. Their results reaffirm the

current study's finding, to be exact; the writers of biology discipline have tendency to use a smaller number of lexical bundles in their works.

### **Structural Analysis of Lexical Bundles**

The second research question aims to find the structural patterns of lexical bundles used by both ES and KS writers in the introductions and the literature reviews of their dissertations. The results illustrated the differences in usages of the structural patterns by both ES, and KS writers. Moreover, variations were also perceived from the cross-disciplinary perspective namely from biology and linguistics disciplines. The results are discussed in the subsequent sections.

#### ***The structural analysis of lexical bundles between the ES and KS writers:***

The most frequent structural patterns used by the ES and KS writers in the chosen dissertations are displayed in Table 6. The below table follows the taxonomy "Structural Analysis of Lexical Bundles" by Biber et.al (2004, p.381). The most common used structure patterns are highlighted and analyzed in the light of the aforementioned taxonomy.

**Table 4.5**

*Main Structures of Bundles between ES and KS Writers (%)*

Structure	ES Frequency per %	KS Frequency per %
Noun Phrase Expressions	16 (43.2)	29 (50.0)
Prepositional Phrase Expressions	18 (48.6)	18 (31.0)
Verb Phrase Fragments	2 (5.4)	8 (13.8)
Comparative Expressions	1 (2.7)	3 (5.2)
Total	37 (100)	58 (100)

The displayed data illustrates the most frequently utilized structural patterns of the lexical bundles, by ES & KS writers, of the given categories "noun phrase expressions, prepositional phrase expressions, verb phrase fragments, and comparative expressions". The total of 37 examples, in the English-speaking corpora, were detected in various structures. Whereas Kurdish-speaking writers utilized 58 examples of the same structures.

Concerning the similarities between the two corpora, both groups of writers share the same frequency of usage (18 times) of the prepositional phrase expressions.



Other similar structural patterns such as comparative expressions were employed. The ES writers utilized only 1 example (2.7%) in the corpora, while the KS writers used the same category 3 times (5.2%). It is necessary to mention that the utilization of all these structural patterns in the corpus of ES writers are less than the usages in the corpus of KS writers.

Relating to the differences between both groups of writers, a number of noun phrases have been employed by both NS writers, with the frequency percentage of (43.2 %), and KS writers (50 %). While for prepositional phrases, the natives scored a higher frequency percentage of (48.6%), the KS writers; however, scored (31%). Verb phrases and comparative expressions, on the other hand, were used more often by the non-natives; with the percentage of (19%) compared to the natives' (8.1%).

To sum up with the analysis of the structural patterns, it is deduced that the ES corpus relies more on the prepositional phrases than noun phrases. This might be by the virtue of that these examples of LBs were used in their dissertations such as *into the active centre*, *in the regulation of*, and *of the trigger loop*. The reason for having this variation might be because of the differences in the structures of both languages, such as the English language is a branch of Germanic languages whilst Kurdish language is a part of Indo-Iranian language family.

On the other hand, the KS writers used more noun phrase expressions in their writing. Their employment of this particular structural classification may result from the language transfer and the nature of Kurdish language contains more noun phrases than other categories, this might be the reason for the KS writers' preferences to noun phrases, for instance *the use of the*, *the central Kurdish dialect*, *the number of the*, and *the past stem of the*.

Below are extractions from the selected dissertations by both groups of writers employing the different structures of lexical bundles.

(1) *Metal ion A functions to promote **the formation of the** attacking hydroxyl group through the displacement of a proton from a water molecule* (ES Corpus).

(2) *In sum, negation in Central Kurdish is represented by the variants of the negation particle **the beginning of the** verbal complex in simple predicates with thematic verbs and at the beginning of light verbs in complex predicate constructions* (KS Corpus).

- (3) Again, both of these niches are characterised **by the presence of** innervated, *Gli1*-expressing stem cells that are dependent on nerve-derived SHH (ES Corpus).
- (4) Concerning the size of an offence and the required apology strategies, offences might be solved by a single or by multiple strategies, some others may require reparation **in addition to the** apology strategies (KS Corpus).
- (5) Labov asserts that (ing) **has been shown to** be the most consistently stable variable across many different communities (ES Corpus).
- (6) However, AP2 **was shown to be** expressed at a low level in adipocyte progenitor cells (KS Corpus).
- (7) Moreover, this study is not a sociolinguistic study in that it did not deal with **the relationship between the** narrative structure of the folktales and features as class, age, people, and cultural identity (ES Corpus).
- (8) In order to capture this aspect of the learning task, modelling studies have adopted clustering techniques which estimate an optimal value for **K as well as the** parameters of each cluster (KS Corpus).

***The structural analysis of lexical bundles across the disciplines:***

The most frequently occurred structural patterns used in the disciplines of biology and linguistics corpora are presented in Table 7. Moreover, the abovementioned taxonomy is also followed in the analysis of cross-disciplinary sections.

**Table 4.6**

*Main Structures of Bundles between Biology and Linguistics Disciplines (%)*

Structure	Biology Frequency per %	Linguistics Frequency per %
Noun Phrase Expressions	17 (48.6)	29 (48.3)
Prepositional Phrase Expressions	10 (28.6)	25 (41.7)
Verb Phrase Fragments	7 (14.3)	3 (5.0)
Comparative Expressions	1 (2.9)	3 (5.0)
Total	35 (100)	60 (100)

The displayed data illustrates the most frequently utilized structural patterns of the lexical bundles, in the disciplines of biology and linguistics, of the given categories “noun phrase expressions, prepositional phrase expressions, verb phrase fragments, and comparative expressions”. The total of 35 examples, in biology discipline, were detected in various structures. Whereas linguistics discipline writers utilized 60 examples of the same structures.

Concerning the similarities between the two disciplines, writers of both fields of study have the similar percentage of structural usage of the “*noun phrase expressions*”. Biology scored (48.6%), and Linguistics has the percentage (48.3%). It should be noted that the noun phrase expressions are the only similarity both disciplines share.

As far as differences between both areas of study is concerned, a fair number of variances is detected. Regarding the prepositional phrases, writers in the field of linguistics were more active in using prepositional phrases with the percentage of (41.7%) compared to (28.6%) in biology discipline. Conversely, verb phrases and comparative expressions were used more often by the biology writers; with the percentage of (17.2%), while linguistics writers scored the percentage of (10%).

The variation in the cross-disciplinary fields might result from the writers’ tendency toward using particular structure in specific disciplines, such as the

linguistics writers prefer to use certain structural patterns. However, both groups of writers used noun phrases almost equally, but their difference are shown in prepositional phrases and verb phrases to report the problem of the studies or introducing their viewpoints. Meanwhile biology writers have the preferences to use less prepositional phrases and more verb phrases compared to linguistics writers.

Below are extractions from the selected dissertations in both disciplines using different structural patterns of lexical bundles:

- (9) *In vitro*, lower doses of genistein (up to 1  $\mu$ M) were reported to inhibit lipogenesis, and the effect appeared to be estrogen receptor-dependent since it was reversed **in the presence of the** estrogen receptor antagonist ICI 164,382 (Biology dissertation).
- (10) *In contrast*, borrowing could be seen as strengthening the Kurdish language by filling lexical gaps and enabling it to function in **a wide range of** modern contexts actively promote this approach (Linguistics dissertation).
- (11) *The stratum granulosum is the outermost nucleated layer and is identified by the presence of* cytoplasmic granules called keratohyalin granules (Biology dissertation).
- (12) *Hence*, when researchers measure complexity, they have to be aware of that and define what they actually mean **in terms of the type** of complexity they measure (Linguistics dissertation).
- (13) **It might be argued that** these social factors are overlapping to a certain extent (Biology dissertation).
- (14) *This is because writing* **has been shown to** contain longer and more complex noun phrases than spoken language (Linguistics dissertation).
- (15) *This aim of drought experiments must be related to the relationship between the effects of different water potentials and the mechanism of plant response to water deficit including “drought-tolerant” plants* (Biology dissertation).

## Functional Analysis of Lexical Bundles

This section attempts to investigate the functional characteristics of the lexical bundles. The framework for functional analysis of lexical bundles in this study follows Hyland's taxonomy. The classification provides three broad sub-categories: "research-oriented bundles, text-oriented bundles, and participant-oriented bundles" (2008a, p. 13). The three broad functional classifications incorporate with sub-classifications: Research-oriented bundles consist of (location, procedure, quantification, description, and topic). Text-oriented bundles include (transition signals, resultative signals, structuring signals, and framing signals). Finally, Participant-oriented bundles incorporate with (stance features and engagement features).

### *The functional analysis of lexical bundles between the ES and KS writers:*

The functional analysis of lexical bundles used by both ES and KS writers in the introductions and the literature reviews of their dissertations are investigated. The results demonstrated the variations in usages of the functional categories by both ES and KS writers. Moreover, differences were noted from the cross-disciplinary perspective, biology and linguistics disciplines. The findings are discussed in the following sections. The most commonly occurred functional categories employed by the ES and KS writers in the selected dissertations are shown in Table 8.

**Table 4.7**

*Main Functions of Bundles between ES and KS Writers (%)*

Function	English-speaking Corpus	Kurdish-speaking Corpus
Research-oriented	19 (51.4%)	37 (63.8%)
Text-oriented	10 (27.0%)	14 (24.2%)
Participant-oriented	8 (21.6%)	7 (12.0%)
Total (%)	37 (100%)	58 (100%)

The data presents the most recurrently used functional categories of LBs by ES and KS writers. The total of 37 examples in the corpora of ES, and 58 examples of KS writers were identified in various categories. In total, 95 examples of functional classifications of LBs are detected. It is noteworthy to mention that the

employment of two of the categories “text-oriented bundles and participant-oriented bundles” by English-speaking writers are more than the Kurdish-speaking writers in the corpora.

Regarding the similarities between both corpora, the two groups of writers employed a comparable frequency of functional category – “text-oriented bundles”, ES writers utilized (10 times, i.e., 27%), and KS writers (14 times, i.e., 24.2%). Out of 95 instances, 24 of them are text-oriented bundles, and this is considered the least employment of functional uses.

Concerning the differences of functional analysis of LBs between ES and KS writers, research-oriented bundles have been utilized with the frequency percentage of (51.4%) by ES writers, and (63.8%) by KS writers. Of the 95 different examples of bundles, 56 are the research-oriented bundles. This shows that Kurdish-speaking writers used a higher range of research-oriented bundles than English-speaking writers. Moreover, research-oriented bundles are found to be predominant comparing the other types of classifications. In terms of the usages of participant-oriented bundles, ES writers employed a greater number with the frequency of (21.6%) compared to KS writers with the frequency of (12%). As a whole, of the 95 instances, 15 are participant-oriented bundles.

One of the reasons that the utilization of research-oriented bundles (topic) related to the field of the study might be due to the research trend, or the research topics conducted by Kurdish researchers on the language or referring to the Kurdish language when making linguistic analysis. Another important reason is that KS writers are interested in employing word chunks which stand for field of study (topic), such as *in the Kurdistan region*, *of the Kurdish language*, *of English loan words*, and *the Central Kurdish dialect*. Using this type of functional category may be due impersonal standpoint of the writers in academic discourse comparing to the spoken discourse.

From the analysis of the current study, it is important to mention that the functional categories that are incorporated with research-oriented bundles (procedure) are interconnected strongly with structural patterns with *noun phrases + of*. A great number of research procedures are used by both groups of writers to determine the methods of how the researches and experiments were conducted, for instance *the use of the*, *the choice of the*, *the majority of the*, and *the presence of the*. Structurally, this type of *noun phrase + of* is considered as the most dominant pattern

in the current study. The purpose of having a great share of research-oriented bundles is that when the writers represent their real-world perspectives and the sense of laboratory-focused topics via writing in the academic discourse communities.

Another link between functional and structural classifications underlines the strong relationship between research-oriented functions (description) and the structural analysis of *noun phrase + of* to provide the description to the research context or objects, such as *the structure of the*, *the fact that the*, and *the formation of the*.

ES and KS writers use text-oriented bundles differently. Their role is to convey a message through organizing the text and providing their meanings to make arguments in academic written discourse. Transition signals to show contrastive cases (*on the other hand*), resultative signals to indicate the reason for doing something (*as a result*), structuring signals to engage the readers through organizing the discourse markers such as (*as shown in the*) to direct the readers elsewhere in the text, and framing signals to establish particular arguments (*with regard to the*).

The results of the current study are in line with the study of Amirian et. al. (2013) in which Iranian writers used research-oriented bundles in master theses. Furthermore, the finding of this study supports Hyland's study (2008a) which displays major differences in the frequency of the functions in the academic prose.

Below are extractions from the selected dissertations by both groups of writers employing the different functional categories of lexical bundles.

- (1) *Furthermore, it has been shown that there is a portion of the H4 tail that has a propensity to interact with an acidic patch **on the surface of the H2A/H2B dimer of an adjoining nucleosome to stabilise condensed chromatin** (ES Corpus).*
- (2) *In contrast, borrowing could be seen as strengthening the Kurdish language by filling lexical gaps and enabling it to function in **a wide range of modern contexts actively promote this approach** (KS Corpus).*
- (3) *During transcription, nascent RNA is fed through **the RNA exit channel** and out of RNAP (ES Corpus).*
- (4) *They studied apology speech act **on the basis of major strategies that were found as responses to certain situations of apology-requiring offences** (KS Corpus).*

- (5) *Frames **on the other hand** are high-level objects that are recognized as conventional ways of seeing things (ES Corpus).*
- (6) *Thus, **it might be argued that** the addressee might be described as an act sensor in terms of politeness and impoliteness (KS Corpus).*
- (7) *These types of effects have been demonstrated **with regard to the** segmentation and recognition of word forms (ES Corpus).*
- (8) *Since the topic of the apology speech act is absent from the field of research **with regard to the** investigation of the Kurdish language, the researcher has found it of linguistic and cultural significance to investigate apology strategies in Kurdish interactions (KS Corpus).*

***The functional analysis of lexical bundles across the disciplines:***

The most recurrently utilized functional classifications of LBs are used in the disciplines of biology and linguistics corpora are presented in Table 9. In addition, the abovementioned taxonomy is also adopted in the analysis of cross-disciplinary parts.

**Table 4.8**

*Main Functions of Bundles between Biology and Linguistics Disciplines (%)*

Function	Biology Corpus	Linguistics Corpus
Research-oriented	22 (62.9%)	34 (56.7%)
Text-oriented	7 (20.0%)	21 (35.0%)
Participant-oriented	6 (17.1%)	5 (8.3%)
Total (%)	35 (100%)	60 (100%)

The shown data presents the most frequently occurred functional classifications of LBs, in the disciplines of Biology and Linguistics, of the provided categories “*research-oriented bundles, text-oriented bundles, and participant-oriented bundles*”. The total of 35 instances, in biology field, were identified in various functional categories, whereas 60 examples in the discipline of linguistics were detected. On the whole, the corpora consist of 95 examples by both groups of writers.



Regarding the similarities between two fields, writers of both disciplines of study have a partial similarity in using research-oriented bundles. Biology writers have the percentage of (62.9%), and linguistics writers scored (56.7%). It is noteworthy to mention that research-oriented bundles are the sole similarity that is shared in the functional analysis between disciplines.

As far as differences between the two disciplines of the study is concerned, a reasonable number of variances is identified. Concerning the usages of the text-oriented bundles, biology discipline scores (20%) of functional uses per the texts, whilst the field of linguistics has the percentage of (35%) of functional classifications. This shows that linguistics writers utilized more text-oriented bundles than biology writers. Relating to the participant-oriented bundles, a diversity can be noticed between the two disciplines. Biology writers employed participant-oriented bundles with the percentage of (17.1%), whereas linguistics writers used (8.3%), and this is a lower usage than biology discipline.

The employment of research-oriented function (topic) results from the field of the study, which might be due to the research trends of biological topics such as *of the hair follicle* and *the antibacterial activity of*. It is significant to note that the functional research-oriented bundles are largely employed in the disciplinary fields comparing to text-oriented and participant-oriented bundles. Similarly, this is consistent with the structural patterns of the use of *noun phrase + of* as the most used structure in the current study. The methods that the researchers focused on while conducting their studies play a significant role in delivering the messages and the empirical basis of the research in hard sciences through using research-oriented bundles (procedures) such as *the use of the*, *the presence of the*, and *the expression of the*. Moreover, writers in the fields of soft sciences such as linguistics use research-oriented bundles with a high frequency because the constituents of this type are noun phrase and prepositional phrase expressions rather than verb phrase expressions. And they have a great share of usages of the academic writing (Biber, 2004). The reason for employing this huge number of research-oriented bundles might be a result of the trend which has formed of the expanded usages of the phrasal and nominal structure patterns in academic written discourse community (Biber et. as., 1999).

The use of text-oriented function (framing structure) paves the way to the writer to refer to the element/information which was previously displayed by framing

the suggestion such as *with regard to the*. As for participant-oriented bundles, the writers attempt to convey their evaluations and attitudes through using stance features and engagement features, for instance *might be argued that, has been shown to, was shown to be, and it might be argued*. Furthermore, both groups of writers' usages of participant-oriented bundles attempt to help the readers and writers to focus on the text via stance features, to express the writers' investigations and positions in the text. For instance, *might be argued that* means something which is already observed from an opposed perspective. The purpose for using engagement features is to address the readers straightforwardly through the text, for example *as shown in the* to assume seeing something.

The finding of the current research is parallel to Hyland's (2008a) study in which the writers in soft science discipline – linguistics was making far greater usage of research-oriented bundles, particularly the bundles beginning with prepositional phrases. In the same way, the thesis's result is consistent with Byrd & Coxhead's (2010) study, writers utilized more research-oriented bundles than other functional classifications in the cross-disciplinary fields.

Below are extractions from the selected dissertations in both disciplines using different structural patterns of lexical bundles:

- (9) *At the **beginning of the** NAC, elongation complexes are in a post-translocated state* (Biology Corpus).
- (10) *Traditionally, idioms have been regarded as non-decomposable, that is, the idiomatic meaning is not **the sum of the** literal meaning of its constituents* (Linguistics Corpus).
- (11) *The SG is connected to the proximal portion **of the hair follicle** (HF) via a keratinised duct through which sebum is released* (Biology Corpus).
- (12) *The claims made in Swingley (2009) cannot be assessed **on the basis of** these studies since these analyses focussed on the relative discriminability of IDS and ADS* (Linguistics Corpus).
- (13) *Dutasteride, which is a both type 1 and type 2 5 $\alpha$ -reductase inhibitor, **has been shown to** have a more rapid effect on scalp hair growth in men with androgenetic alopecia than the finasteride* (Biology Corpus).

(14) *Moreover, it is not just emotional experiences that provide a mechanism to abstract, but the sensory and motor systems also **play a role in abstraction** (Linguistics Corpus).*

## CHAPTER V

### Conclusion and Recommendations

This chapter concludes with the summary of the main results, implications for practice, and recommendations for further research.

#### Summary of the Main Results

This thesis was concerned with lexical bundles that have been growing promptly in applied linguistics studies. The present study was aimed to investigate and compare the employment of lexical bundles in PhD dissertations written by English-speaking and Kurdish-speaking in English in the disciplines of biology and linguistics. This study focused on the two chapters “introduction” and “literature review”. The corpora consisted of 462,888 words in total.

The findings are summarized as follows, from the ES and KS writers’ standpoint; in terms of frequency, it is concluded that KS writers utilized more 3-word and 4-word lexical bundles than ES writers in their dissertations. Regarding the disciplinary standpoint, writers in the linguistics discipline used more 3-word and 4-word lexical bundles than the writers of biology discipline.

Another considerable difference between ES and KS writers is the analysis of structural taxonomy. The results showed that the corpus of KS writers included more examples of structural patterns in comparison to ES writers. Furthermore, KS writers used more noun phrase expressions whereas ES writers utilized more prepositional phrase expressions. From the cross-disciplinary perspective, writers of the linguistics dissertations used a greater number of structural patterns than writers of the biology dissertations; however, the usages of noun phrase expressions were similar in both disciplines.

As far as the functional analysis is concerned, KS writers used more functional categories in their dissertations in comparison with the ES writers. To be exact, the KS writers utilized more research-oriented bundles than the ES writers. However, the usages of text-oriented bundles and participant-oriented bundles were used with a higher frequency by the ES writers than KS writers.

With regards to the hypothesis constructed in the introduction of the study, from the results of the thesis it is concluded that Kurdish writers are aware of the usages of LBs in terms of frequency, structural patterns, and functional

classifications. Both groups of writers exhibited similar results; however, their corpora comprised some differences.

### **Implications**

Consequently, the results of the study provide clear practical implications for pedagogical purposes in these possible ways: first, the frequency of lexical bundles plays a role in selecting, structuring, and sequencing these bundles in teaching curricula such as syntax, comprehension, and vocabulary modules. Second, curriculum designers can benefit from the findings of this study to layout preferred English terminologies for academic purposes. For instance, biology curriculum designers can take advantage of the current study's result in preparing syllabuses such as *Introduction to Biology and its Key Terms*, as a module for freshmen students.

The finding of the current study shows that both ES and KS writers employed *as well as* with the most frequent ratio in their dissertations. This certifies that both groups of writers prefer to use such rhetoric to indicate the meanings of *too*, *also*, *besides*, and *along with*. Thus, teachers and curriculum designers must raise the awareness of the writers of the diversity of possible conjunctions to improve the quality of the text.

Regarding the 4-word LBs, KS writers employed *on the other hand* as the most recurrently utilized bundles. This suggests that for counterarguments and different viewpoints, teachers should teach more connecting words. Teachers and curriculum designers should propose more synonyms for *on the other hand* in the material they design or develop.

As for the functional analysis, the most functional pattern used by the writers is research-oriented bundles. The findings of this study suggest that teaching more functional classifications such as text-oriented bundles is necessary to be taught. The academic writers, English language teachers, and curriculum designers can teach text-oriented bundles to systemize the instruction via transitional signals, resultative signals, structuring signals, and framing signals. They may begin with teaching the transition signals to display the contrastive cases (*on the other hand*), resultative signals to express the reason for doing something (*as a result of*), and framing signals to build a particular argument (*with regard to the*). Moreover, participant-oriented bundles which aims to assist readers and writers to focus on the text through stance

features and engagement features. For instance, stance features convey the writers' assessments and attitudes of the text such as *might be argued that* to mean something which is seen from an opposing viewpoint. Yet, engagement features should be taught to tackle the readers straightaway, for instance *as shown in the* to presuppose seeing something.

### **Recommendations for Further Research**

The current study found similarities between 3-word and 4-word bundles between ES and KS writers, and between biology and linguistics. The result of this study does not represent the whole KS writers they write in English. There are other areas of research to be investigated on lexical bundles. future researchers should examine the reasons for having these similarities.

Furthermore, the study showed the similarities between the categories of the structural analysis. Further studies can be on what are the reasons these similarities occurred. Further investigations should be used for the use of lexical bundles in terms of frequency, structural, and functional analyses with selecting various disciplines since there are no studies on the abovementioned topic by Kurdish researchers.

## References

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92.  
<https://doi.org/10.1016/j.esp.2011.08.004>
- Allen, D. (2010). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*, 32, 105-127. <https://doi.org/10.1017/S0267190512000098>
- Amirian, Z., Ketabi, S., & Eshaghi, H. (2013). The use of lexical bundles in native and non-native post-graduate writing: The case of applied linguistics MA theses. *Journal of English Language Teaching and Learning*, 11, 1-29.  
[https://elt.tabrizu.ac.ir/article\\_589.html](https://elt.tabrizu.ac.ir/article_589.html)
- Baker, P., & Ellece, S. (2011). *Key terms in discourse analysis*. Continuum International Publishing.
- Bhatia, V. (1999). Integrating products, processes, purposes and participants in professional writing. In C. Candlin, K. Hyland, C. Candlin, & K. Hyland (Eds.), *Writing: Texts, processes and practices* (pp. 21-39). Routledge.
- Bhatia, V. (2014). *Worlds of written discourse a genre-based view*. Continuum International Publishing.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–286.  
<https://doi.org/10.1016/j.esp.2006.08.003>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. 25(3), *Applied Linguistics*, 371-405.  
<https://doi.org/10.1093/applin/25.3.371>
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics investigating language structure and use*. Cambridge University Press.
- Biber, D., Johansson, S., Leech, C., Conrad, S., & Finegan, E. (1998). *Longman grammar of spoken and written English*. Pearson Education Limited.
- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6(1), 97–113.  
<https://doi.org/10.1177/1468794106058877>

- Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, 5 (5), 31-64.
- Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, 63(1), 95-122.  
<https://doi.org/10.1177/0023830918823230>
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1) 72–89. <https://doi.org/10.1093/applin/amm022>
- Connor, U. (2002). New directions in contrastive rhetoric. *Tesol Quarterly*, 6(4), 493-510. <https://doi.org/10.2307/3588238>
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397–423. <https://doi.org/10.1016/j.esp.2003.12.001>
- Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12(1), 33–43. <http://dx.doi.org/10.1016/j.jeap.2012.11.002>
- Crawford, W., & Csomay, E. (2016). *Doing corpus linguistics*. Routledge.
- Devitt, A. (2004). *Writing genres*. Southern Illinois University.
- Dong, Y. (1998). Non-native graduate students' thesis/dissertation writing in science: self-reports by students and their advisors from two U.S. institutions. *English for Specific Purposes*, 17(4), 369–390. [https://doi.org/10.1016/S0889-4906\(97\)00054-9](https://doi.org/10.1016/S0889-4906(97)00054-9)
- Dontcheva-Navratilova, O. (2012). Lexical bundles in academic texts by non-native speakers. *Brno Studies in English*, 38(2), 37-58. <http://doi:10.5817/BSE2012-2-3>
- Dornyei, Z. (2007). *Research methods in applied linguistics quantitative, qualitative, and mixed methodologies*. Oxford University Press.
- Durrant, P. (2015). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 38(2), 165–193.  
<https://doi.org/10.1093/applin/amv011>



- Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes*, 29(1), 21-42.  
<https://doi.org/10.1016/j.jeap.2017.09.002>
- Flowerdew, J. (1999). Problems in writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing*, 8(3), 243-264.  
[https://doi.org/10.1016/S1060-3743\(99\)80116-7](https://doi.org/10.1016/S1060-3743(99)80116-7)
- Flowerdew, J. (2013). Introduction: Approaches to the analysis of academic discourse in English. In J. Flowerdew, *Academic Discourse* (pp. 1-18). Routledge.
- Fraenkel, J., & Wallen, N. (2009). *How to design and evaluate research in education*. McGraw-Hill.
- Granger, S. (2014). A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast*, 14(1), 58-72.  
<https://doi.org/10.1075/lic.14.1.04gra>
- Güngör, F., & Uysal, H. (2016). A comparative analysis of lexical bundles used by native and on-native scholars. *English Language Teaching*, 9(6), 176-188.  
<http://doi:10.5539/elt.v9n6p176>
- Hasa. (2017, January 23). *Difference between introduction and literature review*. Retrieved May 30, 2021, from Pediaa: <https://pediaa.com/difference-between-introduction-and-literature-review/>
- Hyland, K. (2006). *English for academic purposes an advanced resource book*. Routledge.
- Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.  
<https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62. <https://doi.org/10.1111/j.1473-4192.2008.00178.x>
- Hyland, K. (2009). *Academic discourse English in a global context*. Continuum International Publishing.
- Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150–169. <https://doi.org/10.1017/S0267190512000037>

- Kashiha, H., & Heng, C. (2015). A little bit about: Differences in native and nonnative. *Australian Journal of Linguistics*, 35(4), 1-14.  
<https://doi.org/10.1080/07268602.2015.1067132>
- Kashiha, H., & Heng, C. (2013). An exploration of lexical bundles in academic lectures: Examples from hard and soft sciences. *The Journal of Asia TEFL*, 10(4), 133-161. <http://psasir.upm.edu.my/id/eprint/27994>
- Khalid, H. (2020). Kurdish language, its family and dialects. *International Journal of Kurdiname*(2), 133-146.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. McGraw Hill.
- Paltridge, B. (1997). *Genre, frames and writing in research settings*. John Benjamins.
- Paltridge, B. (2012). *Discourse analysis: An introduction*. Bloomsbury.
- Roldán-Riejos, A., & Grabowski, Ł. (2019). Towards a cross-linguistic study of phraseology across specialized genres. *Euophras*, 140-144.
- Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing*. University of Oxford.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press.
- Swales, J. (1990). *Genre analysis English in academic and research settings*. Cambridge University Press.
- Thompson, P. (2005). Points of focus and position: Intertextual reference in PhD theses. *Journal of English for Academic Purposes*, 4(4), 307–323.  
<https://doi.org/10.1016/j.jeap.2005.07.006>
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins.
- Ucal, S. (2017). A Corpus-based study on the use of three-word lexical bundles in the academic writing by native English and Turkish non-native writers. *English Language Teaching*, 10(12), 28-36. <http://doi: 10.5539/elt.v10n12p28>
- Wood, D. (2002). Formulaic language in acquisition and production: Implications for teaching. *TESL Canada Journal*, 20(1), 1-15.  
<https://doi.org/10.18806/tesl.v20i1.935>

- Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. Bloomsbury.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Wray, A., & Perkins, M. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, 20(1), 1-28.  
[https://doi.org/10.1016/S0271-5309\(99\)00015-4](https://doi.org/10.1016/S0271-5309(99)00015-4)

## List of Appendices

### Appendix A

#### 3-word Lexical Bundles with Normalization above 1.0/10,000 words in Biology Corpus Written by ES Writers

Rank	Biology corpus written by English-speaking writers	Frequency per 10,000
1	as well as	52 (3.41)
2	the active centre	50 (3.28)
3	been shown to	48 (3.15)
4	the presence of	39 (2.56)
5	the bone marrow	37 (2.43)
6	in addition to	36 (2.36)
7	has been shown	31 (2.03)
8	the formation of	30 (1.97)
9	the trigger loop	27 (1.77)
10	the majority of	25 (1.64)
11	in response to	24 (1.57)
12	the primer template	24 (1.57)
13	structure of the	21 (1.37)
14	formation of the	19 (1.24)
15	of the cell	19 (1.24)
16	of the enzyme	19 (1.24)
17	the cell cycle	19 (1.24)
18	the expression of	19 (1.24)
19	the regulation of	19 (1.24)
20	referred to as	18 (1.18)
21	the structure of	18 (1.18)
22	a number of	17 (1.11)
23	has also been	17 (1.11)
24	it has been	17 (1.11)
25	of the RNA	17 (1.11)
26	the basal layer	17 (1.11)
27	there is a	17 (1.11)
28	are able to	16 (1.05)
29	the removal of	16 (1.05)
	Total	728 (48)

## Appendix B

### 3-word Lexical Bundles with Normalization above 1.0/10,000 Words in Biology Corpus Written by KS Writers

Rank	Biology corpus written by Kurdish-speaking writers	Frequency per 10,000 words
1	as well as	per 10,000)
2	the expression of	71 (4.66
3	the presence of	60 (3.94)
4	the hair follicle	51 (3.34)
5	one of the	46 (3.02)
6	a result of	40 (2.62)
7	the dermal papilla	38 (2.49)
8	due to the	37 (2.43)
9	as a result	36 (2.36)
10	the use of	34 (2.23)
11	in response to	33 (2.16)
12	it has been	33 (2.16)
13	of the hair	33 (2.16)
14	an increase in	32 (2.10)
15	shown to be	32 (2.10)
16	the process of	31 (2.03)
17	a number of	30 (1.97)
18	increase in the	30 (1.97)
19	was found to	30 (1.97)
20	which is a	30 (1.97)
21	been shown to	29 (1.90)
22	the effect of	29 (1.90)
23	the growth of	29 (1.90)
24	the development of	26 (1.70)
25	role in the	25 (1.64)
26	was shown to	25 (1.64)
27	appears to be	24 (1.57)
28	in order to	24 (1.57)
29	of the cell	24 (1.57)
30	part of the	24 (1.57)
31	such as the	24 (1.57)
32	the hair fibre	24 (1.57)
33	antibacterial activity of	23 (1.51)
34	the cell wall	23 (1.51)
35	the effects of	23 (1.51)
36	the level of	23 (1.51)
37	wide range of	23 (1.51)
38	expression of the	22 (1.44)

39	found to be	22 (1.44)
40	in a study	22 (1.44)
41	showed that the	22 (1.44)
42	the antibacterial activity	22 (1.44)
43	activity of the	21 (1.37)
44	has been reported	21 (1.37)
45	in addition to	21 (1.37)
46	involved in the	21 (1.37)
47	because of the	20 (1.31)
48	has been shown	20 (1.31)
49	in the expression	20 (1.31)
50	some of the	20 (1.31)
51	the production of	20 (1.31)
52	a role in	19 (1.24)
53	a study by	19 (1.24)
54	a wide range	19 (1.24)
55	cells of the	19 (1.24)
56	due to their	19 (1.24)
57	expression level of	19 (1.24)
58	shelf life of	19 (1.24)
59	the expression level	19 (1.24)
60	the role of	19 (1.24)
61	to investigate the	19 (1.24)
62	cell wall synthesis	18 (1.18)
63	have been used	18 (1.18)
64	is one of	18 (1.18)
65	on the other	18 (1.18)
66	structure of the	18 (1.18)
67	the ability of	18 (1.18)
68	the formation of	18 (1.18)
69	been reported to	17 (1.11)
70	binding to the	17 (1.11)
71	effects of genistein	17 (1.11)
72	has been used	17 (1.11)
73	in the presence	17 (1.11)
74	the amount of	17 (1.11)
75	the cell membrane	17 (1.11)
76	the majority of	17 (1.11)
77	the other hand	17 (1.11)
78	cell wall turnover	16 (1.05)
79	related to the	16 (1.05)
80	suggested that the	16 (1.05)
81	the absence of	16 (1.05)
82	the expression and	16 (1.05)
Total		2084 (137)

## Appendix C

### 3-word Lexical Bundles with Normalization above 1.0/10,000 words in Linguistics Corpus Written by ES Writers

Rank	Linguistics corpus written by English-speaking writers	Frequency per 10,000)
1	as well as	170 (11.16)
2	the use of	118 (7.74)
3	in order to	93 (9.10)
4	in terms of	64 (4.20)
5	properties of the	62 (4.07)
6	use of the	62 (4.07)
7	in this thesis	58 (3.80)
8	of language change	57 (3.74)
9	the fact that	55 (3.61)
10	the analysis of	42 (2.75)
11	of the input	40 (2.62)
12	of this thesis	40 (2.62)
13	that it is	40 (2.62)
14	well as the	40 (2.62)
15	of the vowel	39 (2.56)
16	one of the	39 (2.56)
17	the number of	39 (2.56)
18	the properties of	39 (2.56)
19	variation and change	39 (2.56)
20	the development of	38 (2.49)
21	a set of	37 (2.43)
22	the study of	37 (2.43)
23	extent to which	36 (2.36)
24	the extent to	36 (2.36)
25	the nature of	36 (2.36)
26	to refer to	36 (2.36)
27	in other words	34 (2.23)
28	nature of the	33 (2.16)
29	of the first	33 (2.16)
30	on the other	33 (2.16)
31	the basis of	33 (2.16)
32	due to the	32 (2.10)
33	the identification of	32 (2.10)
34	based on the	31 (2.03)
35	this thesis is	31 (2.03)
36	the question of	30 (1.97)
37	at the time	29 (1.90)
38	in the input	29 (1.90)

39	of distributional learning	29 (1.90)
40	the relationship between	29 (1.90)
41	a number of	28 (1.83)
42	a series of	28 (1.83)
43	in the same	28 (1.83)
44	of the linguistic	28 (1.83)
45	on the basis	28 (1.83)
46	the area of	28 (1.83)
47	the existence of	28 (1.83)
48	analysis of knowledge	27 (1.77)
49	area of the	27 (1.77)
50	the field of	27 (1.77)
51	of linguistic structure	26 (1.70)
52	of the thesis	26 (1.70)
53	the idea that	26 (1.70)
54	discussion of the	25 (1.64)
55	in the field	25 (1.64)
56	it is not	25 (1.64)
57	levels of linguistic	25 (1.64)
58	between the two	24 (1.57)
59	degree of overlap	24 (1.57)
60	knowledge of the	24 (1.57)
61	language variation and	24 (1.57)
62	number of categories	24 (1.57)
63	the case of	24 (1.57)
64	there is a	24 (1.57)
65	in this section	23 (1.51)
66	such as the	23 (1.51)
67	the other hand	23 (1.51)
68	this is a	23 (1.51)
69	as in the	22 (1.44)
70	distributional properties of	22 (1.44)
71	native language distinctions	22 (1.44)
72	the degree of	22 (1.44)
73	the hyper articulation hypothesis	22 (1.44)
74	for example the	21 (1.37)
75	in the case	21 (1.37)
76	it should be	21 (1.37)
77	language internal accounts	21 (1.37)
78	of native language	21 (1.37)
79	of the distributional	21 (1.37)
80	the distributional properties	21 (1.37)
81	the frequency of	21 (1.37)
82	to be the	21 (1.37)
83	and negative concord	20 (1.31)



84	found to be	20 (1.31)
85	in the first	20 (1.31)
86	is that the	20 (1.31)
87	of American English	20 (1.31)
88	of linguistic variation	20 (1.31)
89	the results of	20 (1.31)
90	the work of	20 (1.31)
91	to which the	20 (1.31)
92	analysis of the	19 (1.24)
93	categories in the	19 (1.24)
94	development of the	19 (1.24)
95	in addition to	19 (1.24)
96	in this case	19 (1.24)
97	learning in infancy	19 (1.24)
98	non quantificational never	19 (1.24)
99	of epistemic policy	19 (1.24)
100	that there is	19 (1.24)
101	the emergence of	19 (1.24)
102	the level of	19 (1.24)
103	at the same	18 (1.18)
104	distributional learning in	18 (1.18)
105	emergent knowledge of	18 (1.18)
106	of linguistic change	18 (1.18)
107	of the same	18 (1.18)
108	other levels of	18 (1.18)
109	part of the	18 (1.18)
110	the notion of	18 (1.18)
111	the same time	18 (1.18)
112	the vowel space	18 (1.18)
113	first year of	17 (1.11)
114	likely to be	17 (1.11)
115	of the community	17 (1.11)
116	should be noted	17 (1.11)
117	the distinction between	17 (1.11)
118	the realization of	17 (1.11)
119	theories of perceptual	17 (1.11)
120	use of distributional	17 (1.11)
121	a lack of	16 (1.05)
122	all of the	16 (1.05)
123	be discussed in	16 (1.05)
124	be noted that	16 (1.05)
125	focus on the	16 (1.05)
126	formal linguistic theory	16 (1.05)
127	in contrast to	16 (1.05)
128	in the following	16 (1.05)

129	in the literature	16 (1.05)
130	in this domain	16 (1.05)
131	level of the	16 (1.05)
132	of the literature	16 (1.05)
133	that have been	16 (1.05)
134	the first year	16 (1.05)
135	the idea of	16 (1.05)
136	the importance of	16 (1.05)
137	the quality of	16 (1.05)
138	the speech of	16 (1.05)
139	used to refer	16 (1.05)
Total		3853 (253)

## Appendix D

### 3-word Lexical Bundles with Normalization above 1.0/10,000 words in Linguistics Corpus Written by KS Writers

Rank	Linguistics corpus written by Kurdish-speaking writers	Frequency Per 10,000 words
1	as shown in	90 (5.91)
2	in terms of	81 (5.31)
3	the Kurdish language	78 (5.12)
4	in central Kurdish	67 (4.40)
5	a number of	64 (4.20)
6	of the Kurdish	63 (4.13)
7	as well as	55 (3.61)
8	in order to	55 (3.61)
9	in the Kurdish	54 (3.54)
10	there is no	53 (3.48)
11	in Iraqi Kurdistan	51 (3.34)
12	as far as	46 (3.02)
13	in this study	46 (3.02)
14	in the past	45 (2.95)
15	in addition to	44 (2.88)
16	in the language	43 (2.84)
17	in the following	38 (2.49)
18	loan words in	38 (2.49)
19	in relation to	37 (2.43)
20	it is not	37 (2.43)
21	there is a	37 (2.43)
22	in other words	35 (2.29)
23	of the language	35 (2.29)
24	it might be	34 (2.23)
25	a result of	32 (2.10)
26	English loan words	32 (2.10)
27	for example the	32 (2.10)
28	that it is	32 (2.10)
29	to be a	32 (2.10)
30	in the present	31 (2.03)
31	as a result	30 (1.97)
32	in contrast to	30 (1.97)
33	is used to	28 (1.83)
34	apology strategies in	27 (1.77)
35	be argued that	27 (1.77)
36	is based on	27 (1.77)
37	it is a	27 (1.77)
38	the language of	27 (1.77)

39	in which the	26 (1.70)
40	used in the	26 (1.70)
41	as in the	25 (1.64)
42	between the two	25 (1.64)
43	is not a	25 (1.64)
44	might be argued	25 (1.64)
45	the case of	24 (1.57)
46	in the case	23 (1.51)
47	it does not	22 (1.44)
48	the central Kurdish	22 (1.44)
49	the use of	22 (1.44)
50	brown and Levinson	21 (1.37)
51	central Kurdish dialect	21 (1.37)
52	followed by a	21 (1.37)
53	for example in	21 (1.37)
54	I argue that	21 (1.37)
55	in the examples	21 (1.37)
56	in the word	21 (1.37)
57	polio and shea	21 (1.37)
58	English and Kurdish	20 (1.31)
59	in the central	20 (1.31)
60	is regarded as	20 (1.31)
61	it can be	20 (1.31)
62	it is the	20 (1.31)
63	of a language	20 (1.31)
64	of English loan	20 (1.31)
65	relationship between the	20 (1.31)
66	shown in the	20 (1.31)
67	that there is	20 (1.31)
68	this is because	20 (1.31)
69	when it is	20 (1.31)
70	be regarded as	19 (1.24)
71	in English and	19 (1.24)
72	in the current	19 (1.24)
73	in the field	19 (1.24)
74	in the literature	19 (1.24)
75	the recipient language	19 (1.24)
76	the relationship between	19 (1.24)
77	words in Kurdish	19 (1.24)
78	account for the	18 (1.18)
79	do not have	18 (1.18)
80	is followed by	18 (1.18)
81	language of the	18 (1.18)
82	learners of English	18 (1.18)
83	Olshan and Cohen	18 (1.18)

84	the difference between	18 (1.18)
85	this is why	18 (1.18)
86	to be the	18 (1.18)
87	accuracy and complexity	17 (1.11)
88	as it is	17 (1.11)
89	be explained in	17 (1.11)
90	central Kurdish is	17 (1.11)
91	for instance the	17 (1.11)
92	Iraqi Kurdish women	17 (1.11)
93	is related to	17 (1.11)
94	is that the	17 (1.11)
95	it has been	17 (1.11)
96	rather than a	17 (1.11)
97	that is the	17 (1.11)
98	the examples in	17 (1.11)
99	the Kurdish culture	17 (1.11)
100	the Kurdistan region	17 (1.11)
101	the study is	17 (1.11)
102	to account for	17 (1.11)
103	words in the	17 (1.11)
104	accuracy and fluency	16 (1.05)
105	and social status	16 (1.05)
106	factors such as	16 (1.05)
107	in general and	16 (1.05)
108	in Iraq and	16 (1.05)
109	in the Kurdistan	16 (1.05)
110	is represented by	16 (1.05)
111	is used in	16 (1.05)
112	one of the	16 (1.05)
113	that there are	16 (1.05)
114	the Iraqi Kurdish	16 (1.05)
115	the process of	16 (1.05)
Total		3108 (204)

## Appendix E

### 4-word Lexical Bundles with Normalization above 1.0/10,000 words in Biology Corpus Written by ES Writers

Rank	Biology corpus written by native English-speaking writers	Frequency per 10,000 words
1	has been shown to	27 (3.80)
2	the formation of the	16 (2.25)
3	the RNA exit channel	14 (1.97)
4	the structure of the	12 (1.69)
5	the presence of the	12 (1.69)
6	the majority of the	11 (1.55)
7	into the active center	11 (1.55)
8	by the presence of	11 (1.55)
9	automation for the biosciences	11 (1.55)
10	is referred to as	11 (1.55)
11	the binding of the	10 (1.41)
12	of the trigger loop	10 (1.41)
13	in the regulation of	10 (1.41)
14	in the nervous system	10 (1.41)
Total		165 (23.22)

## Appendix F

### 4-word Lexical Bundles with Normalization above 1.0/10,000 words in Biology Corpus Written by KS Writers

Rank	Biology corpus written by Kurdish-speaking writers	Frequency per 10,000 words
1	as a result of	35 (3.29)
2	an increase in the	22 (2.07)
3	a wide range of	19 (1.79)
4	the antibacterial activity of	19 (1.79)
5	the expression level of	18 (1.69)
6	in the presence of	17 (1.60)
7	is one of the	17 (1.60)
8	on the other hand	17 (1.60)
9	has been shown to	16 (1.50)
10	in the expression level	13 (1.22)
11	in the absence of	13 (1.22)
12	in the process of	13 (1.22)
13	of the hair follicle	13 (1.22)
14	the expression of the	13 (1.22)
15	the presence of the	13 (1.22)
16	increase in the expression	11 (1.03)
17	play a role in	11 (1.03)
18	the inner root sheath	11 (1.03)
19	was shown to be	11 (1.03)
	total	297 (27.94)

## Appendix G

### 4-word Lexical Bundles with Normalization above 1.0/10,000 words in Linguistics Corpus Written by ES Writers

Rank	Linguistics corpus written by English-speaking writers	Frequency per 10,000 words
1	as well as the	40 (3.00)
2	area of the vowel	24 (1.80)
3	extent to which the	18 (1.35)
4	in the case of	18 (1.35)
5	levels of linguistic structure	18 (1.35)
6	at the same time	17 (1.28)
7	distributional learning in infancy	17 (1.28)
8	it should be noted	17 (1.28)
9	of distributional learning in	14 (1.05)
	Total	183 (13.73)



## Appendix H

### 4-word Lexical Bundles with Normalization above 1.0/10,000 words in Linguistics Corpus Written by KS Writers

Rank	Linguistics corpus written by Kurdish-speaking writers	Frequency per 10,000 words
1	on the other hand	69 (4.53)
2	the meaning of the	44 (2.89)
3	the use of the	36 (2.36)
4	within the nominal phrase	31 (2.04)
5	of the Kurdish language	30 (1.97)
6	the past stem of	27 (1.77)
7	on the basis of	26 (1.71)
8	in the case of	25 (1.64)
9	it might be argued	25 (1.64)
10	might be argued that	25 (1.64)
11	as a result of	24 (1.58)
12	the representation of the	24 (1.58)
13	transitivity of the verb	24 (1.58)
14	the central Kurdish dialect	23 (1.51)
15	the number of the	23 (1.51)
16	meaning of the idiom	20 (1.31)
17	stem of the verb	20 (1.31)
18	the fact that the	20 (1.31)
19	the size of the	20 (1.31)
20	the structure of the	20 (1.31)
21	with regard to the	20 (1.31)
22	of English loan words	19 (1.25)
23	the transitivity of the	19 (1.25)
24	the weight of the	19 (1.25)
25	as shown in the	18 (1.18)
26	size of the offence	18 (1.18)
27	in terms of the	16 (1.05)
28	in the present tense	16 (1.05)
29	the beginning of the	16 (1.05)
30	the choice of the	16 (1.05)
31	the relationship between the	16 (1.05)
32	the sum of the	16 (1.05)
Total		765 (50.24)

## Plagiarism Report

### The Employment of Lexical Bundles in English PhD Dissertations Written by English-Speaking and Kurdish- Speaking Writers

#### ORIGINALITY REPORT

15%

SIMILARITY INDEX

13%

INTERNET SOURCES

11%

PUBLICATIONS

%

STUDENT PAPERS

#### PRIMARY SOURCES

1

[hdl.handle.net](http://hdl.handle.net)

Internet Source

1%

2

[e.bangor.ac.uk](http://e.bangor.ac.uk)

Internet Source

1%

3

[docs.neu.edu.tr](http://docs.neu.edu.tr)

Internet Source

1%

4

[ccsenet.org](http://ccsenet.org)

Internet Source

1%

5

[mafiadoc.com](http://mafiadoc.com)

Internet Source

<1%

6

[milicavukovic.yolasite.com](http://milicavukovic.yolasite.com)

Internet Source

<1%

7

[eprints.soton.ac.uk](http://eprints.soton.ac.uk)

Internet Source

<1%

8

[www.ppk.elte.hu](http://www.ppk.elte.hu)

Internet Source

<1%

[docplayer.net](http://docplayer.net)