

**REPUBLIC OF TURKEY
ISTANBUL MEDENİYET UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF FOREIGN LANGUAGE EDUCATION
ENGLISH LANGUAGE EDUCATION**

**EFL Learners' Use of Formulaic Sequences in Writing: A Study
across Proficiency Levels**

MASTER'S THESIS

Öyküm Tuğçe ÇETİN

September 2022

**REPUBLIC OF TURKEY
ISTANBUL MEDENİYET UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF FOREIGN LANGUAGE EDUCATION
ENGLISH LANGUAGE EDUCATION**

**EFL Learners' Use of Formulaic Sequences in Writing: A Study
across Proficiency Levels**

MASTER'S THESIS

Öyküm Tuğçe ÇETİN

SUPERVISOR

Asst. Prof. Ferdane DENKÇİ AKKAŞ

September 2022

BİLDİRİM

Hazırladığım tezin tamamen kendi çalışmam olduğunu, akademik ve etik kuralları gözeterek çalıştığımı ve her alıntıya kaynak gösterdiğimi taahhüt ederim.

Öyküm Tuğçe ÇETİN

Danışmanlığını yaptığım işbu tezin tamamen öğrencinin çalışması olduğunu, akademik ve etik kuralları gözeterek çalıştığını taahhüt ederim.

Dr. Öğr. Üyesi Ferdane DENKCI AKKAŞ

T.C.
İSTANBUL MEDENİYET ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
YABANCI DİLLER EĞİTİMİ ANA BİLİM DALI
TEZ ONAY İMZA SAYFASI

Öyküm Tuğçe ÇETİN tarafından hazırlanan “*EFL Learners’ Use of Formulaic Sequences in Writing: A Study across Proficiency Levels (Yabancı Dil Olarak İngilizce Öğrenenlerin Yazılı Üretimlerinde Kalıp Söz Kullanımları: Yeterlik Seviyeleri Üzerinden Bir Araştırma)*” başlıklı bu yüksek lisans tezi, Yabancı Diller Eğitimi Ana Bilim Dalında hazırlanmış ve jürimiz tarafından kabul edilmiştir.

JÜRİ ÜYELERİ

İMZA

Tez Danışmanı:

Dr. Öğr. Üyesi Ferdane DENKCI AKKAŞ

Kurumu: İstanbul Medeniyet Üniversitesi

Üyeler:

Prof. Dr. Selami AYDIN

Kurumu: İstanbul Medeniyet Üniversitesi

Dr. Öğr. Üyesi Esra HARMANDAOĞLU BAZ

Kurumu: Müzik ve Güzel Sanatlar Üniversitesi

Tez Savunma Tarihi: 15.09.2022

ACKNOWLEDGEMENTS

“Thankfulness is the beginning of gratitude. Gratitude is the completion of thankfulness. Thankfulness may consist merely of words. Gratitude is shown in acts.”

Henri Frederic Amiel

From the bottom of my heart, I would like to express my deep gratitude to those who made my journey pleasing, encouraging and enlightening.

To begin with, I am grateful to Asst. Prof. Ferdane DENKÇİ AKKAŞ for her generosity in time, knowledge, patience, and help from the very beginning to the end of my master’s study. She was not only a supervisor all through the way; we felt the same excitement, curiosity, disappointment, or sense of achievement together. It was truly enlightening and engrossing to write this thesis under her guidance.

Secondly, I would like to express my heartfelt gratitude to the esteemed faculty members and our instructors Prof. Dr. Selami AYDIN, Assoc. Prof. Özge CENGİZ, Asst. Prof. Elif KIR CULLEN and again Asst. Prof. Ferdane DENKÇİ AKKAŞ for equipping us with necessary information and skills to further our studies. Similarly, I am indebted to all the great teachers who have inspired and encouraged me.

Thirdly, I sincerely appreciate the financial support by Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) during my master’s study. I am also thankful to its staff for always being accessible, attentive and solution-oriented.

Last but not least, I want to say a big thank you to my family for providing me with any opportunity to run after my goals in life. And I would like to express my deepest thanks to my life partner, Tolga Çetin. Thank you very much for being in my corner in any situation, for encouraging me each time I feel overwhelmed or desperate, for helping me realize my dreams, and of course for being my ‘technical man’ throughout my thesis writing process.

ÖZET

YABANCI DİL OLARAK İNGİLİZCE ÖĞRENENLERİN YAZILI ÜRETİMLERİNDE KALIP SÖZ KULLANIMLARI: YETERLİK SEVİYELERİ ÜZERİNDEN BİR ARAŞTIRMA

ÇETİN, Öyküm Tuğçe

Yüksek Lisans, Yabancı Diller Eğitimi Ana Bilim Dalı,
İngiliz Dili Eğitimi Bilim Dalı

Tez Danışmanı: Dr. Öğr. Üyesi Ferdane DENKÇİ AKKAŞ

Eylül, 2022. 117 Sayfa

Kalıp sözler, bir bütün olarak bellekte tutulan ‘prefabrike’ dil birimleri olarak kabul edilir. Bu nedenle dil üretimi sırasında yeniden analiz edilmeleri veya yeniden yapılandırılmaları gerekmez. İnsan zihnindeki bütünsel temsillerinin, yabancı dil olarak İngilizce öğrenenler için zorlu bir görev olarak görülen akademik yazımda birçok avantaj sağladığı bilinmektedir. Mevcut çalışma, kalıp sözlerin sunduğu bu avantajlardan yola çıkarak, İngilizce öğrenenlerin akademik yazılarındaki üç ve dört kelimeli kalıp sözlerin türlerini, sıklıklarını, işlevsel ve yapısal yönlerini öğrencilerin dil yeterlik düzeyleriyle ilişkili olarak araştırmayı amaçlamaktadır.

Derlem incelemesi üzerine yürütülen bu çalışma, mevcut dokümanların incelenmesine odaklanan nitel bir araştırmadır. Dokümanlar, İstanbul’da bulunan iki farklı devlet üniversitesinin Yabancı Diller Yüksekokulları arşivlerinden resmi izinle elde edilmiştir. A, B ve C yeterlik seviyelerini yansıtan araştırma derlemleri, araştırma verilerinde üç ve dört kelimeli kalıp sözler belirlenmesine de yardımcı olan #LancsBox aracılığıyla oluşturulmuştur. Araştırma derleminde beş ve daha fazla sıklıkta kullanılan kalıp sözler, ana dili İngilizce olan kişilerin yazılarından oluşturulan yerli bir derlem olan “LOCNESS” ile karşılaştırılarak bu yerli derlemde karşılık bulmayan kalıp sözler kapsam dışı bırakılmıştır. Geride kalan kalıp sözler, tekrar yerli bir derlem

olan “BNC” ile karşılaştırılmış ve bu derlemde yirmi ve üzeri kullanım sıklığına sahip olmayan kalıp sözler tekrar kapsam dışı bırakılmıştır. Araştırmanın son adımında, nihaî hâle getirilen kalıp söz listeleri, Biber ve diğerleri’nde (2002, 1999) bahsedilen sınıflandırma kriterlerine göre fonksiyonel ve yapısal olarak analiz edilmiştir.

Söz konusu çalışma, üç kelimeli kalıp söz kullanımının dil yeterlik seviyesi hakkında tam bilgi vermediğini, çünkü dil yeterlik seviyesi arttıkça üç kelimeli kalıp söz türlerinde doğrusal bir artış veya azalma olmadığını ortaya koymaktadır. Dört kelimeli kalıp sözlerin çeşitliliğinin ise dil yeterlik düzeyine paralel olarak arttığı gözlemlenmiştir. Üç kelimeli kalıp sözlerin işlevsel analizi, referans içeren kalıpların tüm dil yeterlik seviyelerinde en sık meydana gelen diziler olduğu, ardından duruş/düşünce belirten ve söylem düzenleyen kalıp sözlerin geldiği sonucuna varmıştır. Dört kelimeli kalıp sözlerin işlevleri incelendiğinde, her bir araştırma derleminin mutlaka bir duruş/düşünce belirten kalıp söz içerdiği, A ve B düzeyindeki derlemlerde söylem örgütleyici kalıp sözlerin olmadığı ve A düzeyindeki derlemde ise referans içeren kalıpların olmadığı görülmüştür. Hem üç hem de dört kelimeli kalıp sözlerin yapısal analizi son olarak, fiil öbeği içeren kalıp sözlerin yapısal türlerin çoğunluğunu oluşturduğunu, ardından isim ve edat öbeği içeren kalıp sözlerin ve ‘diğer’ öbekleri içeren kalıp sözlerin en çok kullanılan yapısal türler olduğunu göstermiştir. Bağımlı yan tümce öbeklerini içeren kalıp sözlerin ise her bir araştırma derleminde oldukça sınırlı sayıda kullanıldığı gözlemlenmiştir.

Dört kelimeli kalıp sözlerin analizinin yabancı dil olarak İngilizce öğrenenlerin dil yeterlik seviyeleri açısından üç kelimeli kalıp sözlere göre nispeten daha bilgilendirici sonuçlar sağladığı göz önüne alındığında, dört veya daha fazla kelimedenden oluşan kalıp sözler üzerinde araştırma yapılmasının veya araştırma derleminin çok daha geniş tutulmasının faydalı olacağı söylenebilir.

Anahtar kelimeler: Akademik Yazı, Kalıp Söz, İşlevsel İnceleme, Yapısal İnceleme, Dil Seviyeleri

ABSTRACT

EFL LEARNERS' USE OF FORMULAIC SEQUENCES IN WRITING: A STUDY ACROSS PROFICIENCY LEVELS

ÇETİN, Öyküm Tuğçe

Master's Thesis, Department of Foreign Language Education

Supervisor: Asst. Prof. Ferdane DENKCI AKKAŞ

September, 2022. 117 Pages

Formulaic sequences are considered to be *prefabricated* linguistic units that are kept in memory as wholes, so they do not need to be reanalysed or reconstructed in the course of language production. Their wholistic representation in the human mind is known to provide EFL learners with many advantages in academic writing which is found to be a challenging task by many EFL learners for some reasons. Drawing upon the advantages that formulaic sequences offer, the present current study is conducted to investigate the types, frequencies, and functional and structural aspects of 3- and 4-word formulaic sequences in the EFL learners' academic writings in relation to the learners' proficiency levels.

This corpus-based research is a qualitative study focusing on examining exam documents. The documents were obtained from the archives of the Schools of Foreign Languages of two different state universities in Istanbul, Turkey with official permission. The research corpora, reflecting proficiency levels A, B and C, were formed through #LancsBox, which also helped identify 3- and 4-word formulaic sequences in the research data. The FSs with an occurrence of 5 times or more were checked against the Louvain Corpus of Native English Essays (LOCNESS) to exclude the ones which do not occur in the LOCNESS and then against the British National Corpus (BNC) to exclude the ones which have an occurrence rate of below 20 per million words in the BNC. As the last step, the finalized FS lists were functionally and structurally analysed for each proficiency level separately according to the taxonomies specified in Biber et al. (2002, 1999).

The study reveals that the use of 3-word formulaic sequences is not fully informative about the proficiency level because there was no linear increase or decrease in the types of 3-word FSs as the proficiency level increased; it rather fluctuated. On the other hand, the number of the types of the 4-word formulaic sequences was found to have increased in parallel with the proficiency level. The functional analysis of the 3-word FSs concludes that referential bundles were the most frequently occurring sequences at all proficiency levels, followed by stance and discourse organising bundles. Upon the analysis of the functions of 4-word formulaic sequences, it was seen that each research corpus includes a stance bundle, while there were no discourse organising bundles in the A- and B-level corpora and no referential expression in the A-level corpus. Lastly, the structural analysis of both the 3- and 4-word FSs reveals that the sequences incorporating verb phrase (VP) fragments comprise the majority of the structural types followed by noun phrase (NP) and prepositional phrase (PP) fragments and the “other” fragments. The formulaic sequences incorporating dependent clause fragments were found to be quite limited in each research corpus.

Considering that the analysis of the 4-word formulaic sequences provided relatively more informative results than the 3-word FSs in terms of EFL learners’ proficiency levels, it is recommended to research 4- or more word FSs or to keep the size of the research corpus much greater.

Key words: Academic Writing, Formulaic Sequence (FS), Functional Analysis, Structural Analysis, Proficiency Levels

DEDICATION

**Dedicated to the great leader Mustafa Kemal ATATÜRK
and
to my beloved nuclear family;
Defne Mila & Tolga**

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
ÖZET	vi
ABSTRACT	viii
DEDICATION	x
TABLE OF CONTENTS	xi
LIST OF TABLES	xiv
LIST OF ABBREVIATIONS	xvi
1. INTRODUCTION	17
1.1. Background of the Study	17
1.2. Statement of the Problem	19
1.3. Purpose of the Study	20
1.4. Research Questions	20
1.5. Significance of the Study	21
1.6. Limitations of the Study	22
1.7. Definitions of the Terms	23
2. THEORETICAL FRAMEWORK	24
2.1. Writing and Academic Writing	24
2.2. Formulaic Sequences	26
2.2.1. Identifying Formulaic Sequences	27
2.2.2. Functions of Formulaic Sequences	30
2.2.3. Advantages of Formulaic Sequences	30
2.2.4. Advantages of Formulaic Sequences in Academic Writing	32
2.3. Literature Review	33
3. METHODOLOGY	48
3.1. Research Design	48
3.2. Data Collection Procedure	49
3.3. Research Setting	50
3.4. Data Analysis	51
3.4.1. Identification of 3- and 4-Word Formulaic Sequences Across Proficiency Levels	51

3.4.2. Identification of Functions of 3- and 4-word Formulaic Sequences Across Proficiency Levels.....	52
3.4.3. Identification of Structural Features of the 3- and 4-Word Formulaic Sequences Across Proficiency Levels.....	54
3.5. Reliability and Validity of the Study.....	57
3.5.1. Reliability.....	57
3.5.2. Validity.....	57
4. RESULTS	59
4.1. Findings Regarding the Types and Frequencies of the Formulaic Sequences.....	59
4.1.1. 3-Word Formulaic Sequences in A-Level Writing Exam Documents in terms of Type and Frequency.....	60
4.1.2. 3-Word Formulaic Sequences in B-Level Writing Exam Documents in terms of Type and Frequency.....	62
4.1.3. 3-Word Formulaic Sequences in C-Level Writing Exam Documents in terms of Type and Frequency.....	63
4.1.4. Comparison of 3-Word Formulaic Sequences Across Language Levels....	65
4.1.5. 4-Word Formulaic Sequences in A-Level Writing Exam Documents in terms of Type and Frequency.....	66
4.1.6. 4-Word Formulaic Sequences in B-Level Writing Exam Documents in terms of Type and Frequency.....	67
4.1.7. 4-Word Formulaic Sequences in C-Level Writing Exam Documents in terms of Type and Frequency.....	68
4.1.8. Shared 4-Word Formulaic Sequences across Language Levels A, B and C	69
4.2. Functional Categorisation of the Formulaic Sequences in the Research Data.....	70
4.2.1. Functional Categorisation of 3-Word Formulaic Sequences for Level A...	70
4.2.2. Functional Categorisation of 3-Word Formulaic Sequences for Level B...	73
4.2.3. Functional Categorisation of 3-Word Formulaic Sequences for Level C...	75
4.2.4. Functional Categorisation of 4-Word Formulaic Sequences for Level A...	77
4.2.5. Functional Categorisation of 4-Word Formulaic Sequences for Level B...	78
4.2.6. Functional Categorisation of 4-Word Formulaic Sequences for Level C...	79
4.3. Structural Categorisation of the Formulaic Sequences in the Research Data.....	81
4.3.1. Structural Categorisation of 3-Word Formulaic Sequences Identified in the A-Level Corpus.....	82

4.3.2. Structural Categorisation of 3-Word Formulaic Sequences Identified in the B-Level Corpus.....	84
4.3.3. Structural Categorisation of 3-Word Formulaic Sequences Identified in the C-Level Corpus.....	86
4.3.4. Structural Categorisation of 4-Word Formulaic Sequences Identified in the A-, B- and C-Level Corpora	89
4.3.5. Structural Categorisation of 4-Word Formulaic Sequences Identified in the A-Level Corpus	90
4.3.6. Structural Categorisation of 4-Word Formulaic Sequences Identified in the B-Level Corpus.....	90
4.3.7. Structural Categorisation of 4-Word Formulaic Sequences Identified in the C-Level Corpus.....	91
5. CONCLUSIONS, DISCUSSION AND RECOMMENDATIONS	92
5.1. Conclusions and Discussion	92
5.2. Recommendations.....	102
5.2.1. Recommendations for Practitioners and Policymakers	103
5.2.2. Recommendations for Further Research	103
REFERENCES.....	105
APPENDICES.....	115
CURRICULUM VITAE	116

LIST OF TABLES

Table 1. Structural Features of Formulaic Sequences with Samples	55
Table 2. Comparison of 3-Word Formulaic Sequences in Data of Level A with British National Corpus	60
Table 3. Comparison of 3-Word Formulaic Sequences in Data of Level B with British National Corpus	62
Table 4. Comparison of 3-Word Formulaic Sequences in Data of Level C with British National Corpus	64
Table 5. Shared Formulaic Sequences Across Levels.....	66
Table 6. Comparison of 4-Word Formulaic Sequences in Data of Level A with British National Corpus	67
Table 7. Comparison of 4-Word Formulaic Sequences in Data of Level B with British National Corpus	67
Table 8. Comparison of 4-Word Formulaic Sequences in Data of Level C with British National Corpus	69
Table 9. Functional Categorization of 3-Word Formulaic Sequences in the Data of Level A.....	71
Table 10. Functional Categorization of 3-Word Formulaic Sequences in the Data of Level B	73
Table 11. Functional Categorization of 3-Word Formulaic Sequences in the Data of Level C	75
Table 12. Functional Categorization of 4-Word Formulaic Sequences in the Data of Level A.....	78
Table 13. Functional Categorization of 4-Word Formulaic Sequences in the Data of Level B	78
Table 14. Functional Categorization of 4-Word Formulaic Sequences in the Data of Level C	79
Table 15. Structural Categorization of the 3-Word Formulaic Sequences in the Corpus of Level A	82
Table 16. Structural Categorization of the 3-Word Formulaic Sequences in the Corpus of Level B.....	85

Table 17. Structural Categorization of the 3-Word Formulaic Sequences in the Corpus of Level C.....	87
Table 18. Structural Categorization of the 4-Word Formulaic Sequences in the Corpora of Level A, B and C	89

LIST OF ABBREVIATIONS

BNC: British National Corpus

EFL: English as a Foreign Language

FS(s): Formulaic Sequence(s)

LOCNESS: Louvain Corpus of Native English Essays

NP(s): Noun Phrase(s)

PP(s): Prepositional Phrase(s)

VP(s): Verb Phrase(s)

L2: Second Language

1. INTRODUCTION

This section firstly introduces the key terms touched upon throughout the study and their importance in literature and in the field of education together with their relevance to the current study. Secondly, it provides information about the frame of the study with explanations of the purposes. In the third part of the Introduction, the research questions were presented. The fourth part provides some justifications as to the need for such a study. Then, the limitations of the study were discussed. And lastly, the terms used in the study were defined in line with their operational definition in the study.

1.1. Background of the Study

Language learners have various motives behind their effort to learn a second or foreign language. Some of those motives could be described as increasing the chance to make a better career (Gao, 2010), enabling integration with citizens of different cultures (Noprival et al., 2021; Türker & Çelik, 2022), creating positive mindset for the learners themselves (Jin et al., 2014), and engaging in different areas of work or interest like art, education, and science (Al-Oliemat, 2013). Especially in the field of education, achieving academic literacy appears to be one of the primary motivations for language learners (Badstübner & Ecke, 2009; Tahaine & Daana, 2013; Zarrabi, 2018; Davis, 2020). It is also the one which is considered to be maybe the most challenging by language learners because as Paltridge (2004) cites in his article, academic writing requires the acquisition of new sets of rules and the application of these rules, which vary across disciplines, purposes of writing, and the targeted readers. Learners of a second/foreign language might fall short in meeting requirements to be achievers because of probable L1 effect, lack of thorough linguistic competence at different language levels, especially in reading and writing, poor command of linguistic devices related to the field, lack of organisational skills, or failure to incorporate with the audience in the course of making meaning. (Casanave, 2002).

At this point, formulaic sequences (FSs) gain importance because of some of the benefits they offer in language production. One of them is *processing efficiency*. Formulaic sequences are prefabricated linguistic units that are retrieved as wholes, so they do not need to be constructed or analysed each time during production (Wray, 2000). A corpus-based study carried out by Jiang and Nekrasova (2007) reported findings strongly supporting the holistic representation and processing of FSs in learners of a second language. And this, consequently, provides language learners/users “a processing advantage over creatively generated language” (Conklin & Schmitt, 2008: 72). The whole-word representations of FSs in the human brain promote quick analysis of linguistic input and output. As Coxhead and Byrd (2007) put it, they enable fluency in language use. Secondly, Silva and Matsuda (2002) suggest that the responsibility of writers is not only to construct reality in their texts but also to align this reality with the expectations and points of view of the readers. In this respect, formulaic sequences are observed to help experienced and novice academic writers set their tone, style and voice in their work in accordance with the necessities of academic writing and the expectations of the audience. Besides, effective and accurate use of FSs facilitate register-, genre-, and discipline-specific content production, so lack of relevant or accurately-used formulaic sequences in academic writing might indicate linguistic, technical or academic incompetence, or *foreignness* to the target community (Paquot, 2008). Lastly, the use of formulaic sequences in both spoken and written language is considered to facilitate integration with the target community (Wray, 2002). Producing a language in a similar way that the target community cultivates togetherness among the members of a society.

Considering the mentioned advantages, studying the nature of formulaic sequences, understanding the tendencies of native and non-native language users in terms of FS use or avoidance, investigating the relationship between the use of FSs and the learners’ proficiency levels, and designing educational programs accordingly will evidently facilitate the acquisition of necessary academic writing skills. So far in literature, there are numerous studies comparing the use of FSs in L1 and L2 academic writing (i.e., Durrant & Schmitt, 2009; Chen & Baker, 2010; Ädel & Erman, 2012; Ahmadi et al., 2013; Pérez-Llantada, 2014; Öztürk & Köse, 2016; Pan et al. 2016; Nam, 2017; Bychkovska & Lee, 2017; Pan & Liu, 2019; Lu & Deng, 2019; Lyu &

Gee, 2020; Kim & Lee, 2020), investigating the development of FS use in academic writing longitudinally (i.e., Li & Schmitt, 2009; Crossley & Salsbury, 2011; Zheng, 2015; Shin & Kim, 2017; Öztürk, 2021; Duan & Shi, 2021), and examining use of FSs across proficiency levels (Staples et al., 2013; Ruan, 2016; Appel & Wood, 2016; Li & Volkov, 2017; Goulart, 2020; Hejazi, 2020). Like many other invaluable studies in literature, the use of formulaic sequences in academic writing has long been a compelling and important research area in the field.

1.2. Statement of the Problem

English, a world language in trade, science, politics, and education, has been taught as a foreign language (EFL) in Turkey starting in the second grade in primary school and mostly until the first year of university education, including the preparatory program. Despite this long period, the EFL learners still feel they lack the necessary skills to produce spoken and written language in English (Coşkun-Demirpolat, 2015; Suna & Durmuş-Çelebi, 2013). Especially at the tertiary level, students are expected to use English for academic purposes. Therefore, they generally feel challenged by the dynamics of academic writing since they are supposed to write in their non-native language. Academic writing requires many micro and macro skills of reading and writing along with other language skills to cooperate, making it an overwhelmingly demanding task for the writers of a second or foreign language. When it comes to the academic lexicon, writers need to deal with various “multi-word combinations, idioms, collocations, word types based on complement and valency patterns, lexical bundles, semi-fixed sequences combining semantic requirements with grammar and vocabulary, frames involving highly frequent words, and miscellaneous other subsets of recurrent lexical sequences” (Coxhead & Byrd, 2007: 135-136).

Formulaic language is invaluable for language use in general and for L2 learning in particular. As explained in detail above, the use of FSs supports L2 learners to become more productive, get more proficient and closer to native speakers and improve their grammatical knowledge in the target language. Therefore, it is no surprise that how FSs are used and acquired has become a popular research topic in second language studies. Yet, the literature on the use of FSs among L2 learners is still limited and the

present studies do not have a clear answer or a final conclusion regarding the issue. First, most of the studies have focused on comparing EFL learners from different proficiency levels or EFL corpora with native corpora to reveal the effect of proficiency on FS use. Second, almost all the studies examine FS use based on their frequencies and functions whereas there are relatively fewer studies including a structural analysis as well. Third, studies examining the writings of EFL learners from various L1 backgrounds have revealed interesting variations regarding the L1 effect on the use of FSs. Finally, there are scarcely any studies on FS use in the Turkish EFL context. In brief, considering the significance of FS use in the development of academic writing and the very limited number of studies in the EFL Turkish context, there is a need to understand how Turkish EFL learners use formulaic language in their written production.

1.3. Purpose of the Study

The present corpus-based study aims to analyse the use of 3- and 4-word formulaic sequences in the Turkish EFL learners' writing exam documents. It is designed to analyse the FSs in the research corpora concerning the learners' proficiency levels, which are levels A, B and C. The first objective is the analysis of the types and the frequencies of the 3- and 4-word FSs and to find out if there is an increase or decrease in the types and frequencies across the levels. Secondly, the study aims to investigate if the functional features of those FSs vary across levels, and the proficiency groups prefer which types of FSs in terms of functions. Lastly, the study focuses on the structural aspects of the FSs in the research corpora and investigates the changing pattern of structural features of the FSs across levels. As stated above, since this is a qualitative study, it does not aim to draw obvious conclusions or to generalise the findings to other speech communities or to all EFL learners in the Turkish context.

1.4. Research Questions

The present study raises four research questions:

1. What are the types and frequencies of 3- and 4-word formulaic sequences used by Turkish EFL learners in academic writings across proficiency levels?

2. What functions do the 3- and 4-word formulaic sequences perform according to Biber et al.'s taxonomy (2004) in the academic writings of Turkish EFL learners across proficiency levels?
3. What are the structural features of the 3- and 4-word formulaic sequences Biber et al.'s taxonomy (1999, 2004) in the academic writings of Turkish EFL learners across proficiency levels?
4. How do the types, frequencies, functions, and structures of the 3- and 4- word formulaic sequences vary in the academic writings of Turkish EFL learners across proficiency levels?

1.5. Significance of the Study

While debates over the possible reasons for the problems regarding EFL instruction in the Turkish context (i.e., Aydın, 2008; Peaci, 2016; Özmat & Senemoğlu, 2021; Dağtan & Cabaroğlu, 2021) have been going on for several assumptions, the current study, first of all, might provide a relatively new perspective on the problem by directing attention to the acquisition and use of formulaic sequences. Second, the comparison of the FSs in the Turkish EFL learners' writings with the ones in the native writers' in the context of the current study bears practical significance. When the corpora of both spoken and written English are investigated delicately, it is seen that formulaic language is one of the defining features of a native language or well-written prose (Biber et al., 1999; Chen & Baker, 2010; Karabacak & Qin, 2013). A similar investigation on the corpora of Turkish EFL learners would reveal how parallel L2 development goes with the native corpora in terms of FS use. Third, it is important to investigate the variations in the use of formulaic sequences across proficiency levels. This may help understand and explain some difficulties EFL learners face regarding their use of formulaic language. This understanding might provide some hints about the interlanguage development as well. Fourth, this is relatively a new area of research in the Turkish context, and there are only a few studies investigating the Turkish EFL learners' use of formulaic language in relation to native corpora. Finally, the studies which analyse the functions and structures of the 3- and 4-word formulaic sequences in the EFL learners' academic writings in the light of internationally- and

academically-acknowledged specifications/taxonomies are highly limited in the Turkish context. Therefore, this study is significant in filling a gap in the related literature as well.

1.6. Limitations of the Study

This study is not without its limitations. First, to observe the changing pattern in the use of formulaic sequences and their structural and functional features across levels, examining the written productions of the same students through different language levels would reveal more reliable findings. Therefore, modular systems in preparatory programs could serve this purpose quite successfully. Before the Covid-19 outbreak, the study had been designed to select 20 students from different language levels and analyse their writing exams that they would take in successive modules. In this way, the effect of the increasing language level on the use of formulaic sequences would have been observed much better by discarding individual, curricular and environmental factors to a great extent. However, because of the shift to online education during the pandemic, it was not possible to obtain the existing exam documents in the archives of the institutions in this formidable way. Also, it would not be possible to gather accurate data if the online writing exams had been requested because of several reasons. The exams might have been simplified, the students might have inconveniently got help during the exams, or the requirements to pass an exam might have been waived. For this reason, the existing documents belonging to three proficiency levels were obtained from two different universities with a semester system. Hence the language groups were different from each other in terms of demographics, educational background and differing educational systems implemented in different universities. To make up for these shortcomings, as many documents as possible were obtained from each proficiency level. Second, the documents belonged to the very first writing exam of the semester, so some concerns about whether the exams revealed the true proficiency level arose; however, since the students were placed into their classes after a placement test, their levels were accepted to be accurate. Also, most of the A-level and some of the B-level EFL learners' written productions were short because it was their first exam, but the large number of the obtained documents was thought to compensate for it as well. Lastly, some writing

documents of the senior students were included in the C-level research corpus, for the number of the C-level EFL students in the preparatory program was limited, which made the C-level research corpus questionable in terms of the learners' actual language level. However, it was assumed that the seniors could not have improved their knowledge of English in such a short time till the very first writing exam of the semester. In brief, the study seems to have a couple of limitations, but they were recognized in advance and believed to be compensated in specified ways or conditions.

1.7. Definitions of the Terms

The operational definitions of the terms used in the current study are stated as follows:

Academic writing: It could be defined as “a means of producing, codifying, transmitting, evaluating, renovating, teaching, and learning knowledge and ideology in academic disciplines” (Fang, 2021).

Formulaic sequence: The term refers to “a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (Wray, 2002: 9).

EFL learners: Iwai (2011) defines the term as English learners “who learn English in non-English speaking countries and are therefore called English as a Foreign Language (EFL) learners” (p. 150).

Proficiency level: The term is used synonymously with “reference levels” in this study. The “reference levels” are defined to “identify the forms of a given language (words, and grammar), mastery of which corresponds to the competence defined by the CEFR” (Council of Europe Portal, n.d.).

A-, B- and C- level users: In the Common European Framework of Reference for Languages (CEFR), the A-level users are defined as “basic users”, and B- and C-level users are defined as “independent user” and “proficient user” respectively (Council of Europe, n.d.: 23).

2. THEORETICAL FRAMEWORK

This part of the study aims to define some concepts such as writing, academic writing, and formulaic sequences through a concise and compact examination and synthesis of related studies in the literature. The second objective is to present some guidelines to identify formulaic sequences, by referring to prominent researchers and their estimable works as much as possible. After a touch to the framework by Biber et al. (2004) for functional analysis of formulaic sequences, the theoretical framework part of the study tries to demonstrate some of the abounding benefits of the sequences in language use in general and also specifically in academic writing.

2.1. Writing and Academic Writing

Of all the four language skills, writing is known to be the latest developing one because of its complex nature. As Nunan (1989) states, it requires a great deal of cognitive effort both at a sentence and text level. The writer has to deal with not only word and sentence formation, content and format but also the coherent and cohesive formation of the whole text. Since it is a productive skill which necessitates synthesizing different components of a language, such as morphology, syntax, semantics, and pragmatics, it also requires a distinct combination of various subskills at different levels. Chenoweth and Hayes (2001) mention three levels of writing in their work. At the resource level, language learners resort to different types of memory to make use of their internal background knowledge. At the process level, learners are in complete control of the whole writing process from the production and organisation of the ideas to the transcription of them and correction of the written language, which necessitates internal and external processes go hand in hand. The last stage is the control level, which signifies the “task schema” (p. 84) that encompasses the objectives of the task and a group of productions managing the relations among series of actions throughout levels.

Usó-Juan et al. (2006) state that writing is an important skill to gain communicative competence in a language and mention 5 components of writing. The first, being crucially important, is *discourse competence*, which enables writers to produce cohesive and coherent text. The second is *linguistic competence*, which encompasses the knowledge of spelling, words, meanings, and rules. *Pragmatic competence* is the third component of writing skill, and it requires writers to align their production with the characteristics and expectations of the audience by paying attention to “politeness issues such as degrees of formality.” (p. 392) The fourth component of writing is explained as the necessity to comply with the standards and principles of the target community during text production, which is named as *intercultural competence*. Lastly, writers need to devise strategies to acquire the skill of writing and interact with the audience efficiently through writing.

As seen, acquiring writing skills is an arduous work affected by individual factors to a great extent, and motivation could be said to be one of the most important determinants. If not done for pleasure or as a requisite for any educational activity such as passing a course, and writing to take part in academia is prevalent motivation for language users. Not surprisingly, they feel challenged by the dynamics of academic writing, especially if they are supposed to write in their non-native language. Cherry Campbell in Kroll (1994) points out that academic writing involves the integration of the highlights of previously-read academic papers in related research areas with writers’ own work. She adds that it is a skill of learning, comprehending, making comments and revisions, and adapting, which develops as writers progress. Similarly, Coxhead and Byrd (2007) state that academic writing is closely bound up with reading academic works. Therefore, writers need to be academically knowledgeable about different genres and the language those genres necessitate. The language academic prose requires comprehension and production of relatively longer and more complicated content which is made up of a higher-level academic lexicon. Again, Coxhead and Byrd (2007) summarise that academic texts grammatically consist of prepositional phrases preceded by long noun phrases, repeatedly-used small extent of verbs, structures often constructed in passive voice, many adverbial phrases to pinpoint the location of relevant information, and data, and various long word units which are not typical of everyday conversation.

As can be seen, academic writing requires many micro and macro skills of reading and writing along with other language skills to cooperate, making it an overwhelmingly demanding task for the writers of a second or foreign language. When it comes to the academic lexicon in both reading and writing, writers need to deal with various “multi-word combinations, idioms, collocations, word types based on complement and valency patterns, lexical bundles, semi-fixed sequences combining semantic requirements with grammar and vocabulary, frames involving highly frequent words, and miscellaneous other subsets of recurrent lexical sequences” (Coxhead & Byrd, 2007: 135-136).

2.2. Formulaic Sequences

Formulaic sequences (FSs) have been in the scope of academia for a long time because of their facilitative role throughout the language learning process. A great number of studies have shown that use of FSs helps learners diminish language mistakes at all levels of the target language promoting native-like language use. Besides, since formulaic sequences are *prefabricated* language units ready to use, they are invaluable language “assistants” for lower-level language learners in that they expedite language improvement, boost lexical repertoire with relatively smaller effort in a shorter time period, and facilitate the internalisation of the rules and dynamics of the target language. Just as the abundance of their benefits which have been scientifically suggested, boundless attempts to define the term or name it accordingly have been made so far. Wray (2000) defines formulaic sequences as seemingly pre-constructed language units, which are kept in memory as wholes so they do not require reconstruction or analysis in the course of production. According to Biber and Barbieri (2007), they are usually structurally incomplete, oft-used words, which are usually transparent in meaning, fulfilling significant functions in various written and spoken discourse. Another characterisation of the FSs lies in their frequency of use in different types of texts. A large amount of examination of written and spoken discourse has shown that formulaic sequences constitute the highest proportion of words that occur in the context, so some of the researchers carry out their studies analysing not the formulaic sequences for their own sake but the most-occurring ones in the text under

examination (see Biber et al., 2004; Chen & Baker, 2010). Conrad and Biber (2005), on the other hand, highlight that what makes a sequence *lexical bundle* or *formulaic sequence* as named in this study, is not only its prevalence in a context, but also its being fixed, that is, whether it allows for slots that can be occupied by various function or content words or not (see Staples et al. 2013), and its being used by numerous users of the language to be able to discriminate if they reflect the style of the persons or if they could actually be accepted as *prefabricated* language. The last mostly-adopted view of the formulaic sequences is put forward by Biber (2009) who defines formulaic sequences as “a kind of pragmatic ‘head’ for larger phrases and clauses” (p. 284). In this way, he explains, the sequences construct an explanatory basis for the rest of the expression, which means they play an undeniably significant role in building discourse.

In brief, it does not seem possible to label formulaic sequence with a single definition since it is a multifaceted construct with numerous structural, frequency-related and functional features. Although researchers might choose different operational definitions for formulaic sequences in their papers, what they share in common is the assumed benefits of FSs in the process of language learning, and relevant to this research, in the construction of written texts.

2.2.1. Identifying Formulaic Sequences

Identification of a formulaic sequence is at least as toilsome as defining the term since there are various methods to acknowledge a word unit as a formulaic sequence. One of the guides to determine target word units, though a controversial one, is *intuition*. Far from being scientific or statistical, it is the way how a sequence sounds idiomatic, natural, grammatical, or acceptable to the native speakers of the language or to the language community that determines FSs. Sinclair (1991) opposes the practice of intuition when identifying FSs claiming that linguistic intuitions do not always inform practical use, and they are specific. As they might be case-specific, they might not be acknowledged by the language community. That results in an important deficit in an

academic study because it might cause the reliability of the study to become open to debate.

There is, however, a method that could compensate for the shortcomings of intuition in identifying FSs, which is *corpus check*. Corpus is a compilation of written or spoken language data, most often categorised, to study a language and its use. It basically shows the distribution of word units, their pattern and frequency of use, and their occurrence with other units across texts of various kinds and lengths. The logic behind corpus check is, in a sense, to see how frequently a word occurs with another one or ones, which is one of the most reliable ways to identify the formulaic language. Regarding the reliability of corpus check, Sinclair and Renouf (1998: 151) highlight that “the statistical evidence is helpful, too, because it distinguishes the commoner patterns of usage, which occur very frequently indeed, from the less common usage, which occurs very infrequently”. Nonetheless, corpus check has its own drawbacks as well. The very first one is that unless a corpus is a large one, it might not be a good representative of all possible word combinations. In other words, a small corpus might fail to represent genuine but infrequently-used formulaic sequences. Secondly, as Wray (2002) points out, a small corpus might ignore or overrepresent some formulaic sequences which are common in narrow contexts for a short period of time, for instance, the ones that have become temporarily popular because of TV programmes. Lastly, tools used in corpus check might again fail to show boundaries between formulaic sequences (Wray, 2002). While the expression “I miss you” can count as an FS on its own, it occurs very frequently with other parts of speech as well; “I miss you badly”, “I miss you so bad”, “I miss you so much”, and “I miss you”. In this case, it becomes difficult to determine which exactly could be accepted as formulaic only through corpus check.

Frequency-based identification method appears here as a supporting actor, which again necessitates a corpus check. It is carried out to see how frequently linguistic units occur, alone or together with other units, in the target corpus. Researchers specify some parameters before analysing the corpus: the cut-off point for frequency of word combinations, the length of combinations to be examined and dispersion of the sequence across texts. For instance, Biber et al. (2004) adopt a frequency-driven

approach to determine what counts as a formulaic sequence and expect a sequence to occur 40 times in a million words in a minimum of five distinct texts. The same approach is taken in Chen and Baker's (2010) study. They work on the sequences with a cut-off point of 20-40 per million words appearing in 3 to 5 different texts. Though frequency-based measurement is one of the most practical and decent methods for the identification of FSs, it is not thoroughly reliable when adopted alone, just as the other methods are not. The first two reasons are related to the size of the corpus and their representativeness, mentioned above among the drawbacks of corpus check. Another issue is that it is not informative of the psycholinguistic validity of the sequences (Wood, 2015). That is, it is not possible to decide if the sequence is processed as a single word unit because of the mutual information (MI) of its constituents, the coherence between the constituents, or a frequency of the sequence. Finally, as Wood (2015) points out, frequency-based search for formulaic sequences might not exclude word combinations that bear no meaning together. Or it might not include combinations which are formulaic in some other contexts but do not seem relevant in the immediate context. Researchers do follow a few more measurement methods (see Wood, 2015 for detailed explanation) to identify the formulaic language in target texts. They could be listed as psycholinguistic measurements during which eye-tracking or response latency is measured, use of criteria checklists that comprise the characteristics affiliated to FSs, exercising native speaker judgments, and measurement of phonological characteristics (mostly in spoken context) during which features like "alignment with pauses and intonation units, resistance to internal dysfluency, no internal hesitations, fast speech rhythm, and stress placement restrictions" (Wood, 2015: 23) are investigated.

The aforementioned first three methods could be said to be the most frequently-employed methods during the identification of formulaic language in related literature with all their strengths and weaknesses. However, it is worth highlighting that scholars most often use triangulation of different methods in their research to eliminate the drawbacks of each method.

2.2.2. Functions of Formulaic Sequences

Studies on formulaic sequences usually focus on the frequency of use of sequences across a variety of texts. However, a brief scanning of available literature would show that almost every frequency-oriented study is consistently supported by an analysis of the functional classification of the available sequences in the texts. The analyses are carried out in the light of the frameworks established by prominent researchers, Biber and his colleagues being one of the out fronts.

Biber et al. (2004) set a framework comprising three main categories to explain how formulaic sequences function in a clause or sentence. When they mark the author's attitude, position, or interpretation, they are categorised as *stance bundles*. Much of the research in literature so far (Juknevičienė, 2009; Jaworska et al., 2015; Pan et al., 2016; Ruan, 2016; Nam, 2017; Lu & Deng, 2019) has shown that stance bundles are the mostly-used sequence type in the productions of mainly lower-level or non-native language learners. The second category of the sequences is specified based on their function of establishing a meaning connection between the preceding and following discourses. They are named *discourse organisers* when they open a topic up or expand it. *Referential bundles* that signify or highlight an important aspect of an entity come up as the last function formulaic sequences carry out. In the same paper, the researchers present that stance bundles are the most-preferred bundle type in spoken discourse while academic writing comprises mostly referential bundles, specifically among four-word bundles that occur 40 times per million words.

2.2.3. Advantages of Formulaic Sequences

Formulaic sequences, “minimal” language units that are comprised of either single words or a group of words, have long been of great importance to the researchers and language learners for the benefits they serve in the course of language learning and language use. Surprisingly enough, the benefits of formulaic language are not limited to language users only, but the target audience also takes considerable advantage of these language units.

To begin with, it is possible to say that formulaic sequences help language learners adapt to novel linguistic concepts, structures, signs, and contexts in case of limited or inadequate knowledge of language. Wray (2002) exemplifies that with examples of juxtapositions in poetry which are normally challenging to comprehend since they are unfamiliar word combinations to the learner, hearer, or reader. With the examples she has given in her book, she argues that it is human beings' ability and flexibility to interpret entirely new linguistic combinations rather than pure grammatical competence or rigid formulae in memory. At this point, it would be misleading to propose that language users rely only on ready-made phrases or phrases they have come across before, according to the writer. Drawing on Wray's (2002: 9) definition, "retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar", formulaic sequences provide language users with processing advantage, which, in turn, promotes economy of time and effort during both spoken interaction and written production. Although many studies carried out using the eye-tracking method produce unignorable evidence for this type of holistic approach to FS storage and retrieval, researchers are still cautious about making generalisations without fully understanding background psycholinguistic processing. One other benefit formulaic sequences offer is their function of forging or maintaining identity and unity with a speech community. Language as an invaluable element of societies and cultures is a great tool in serving the common good. It transfers shared knowledge, thoughts, and feelings among community members; therefore, it can be considered as a quite powerful tool in strengthening affinities. As Wray (2002) puts it, when individuals feel excluded or isolated in a community, integrating shared formulaic sequences into their speech helps them stand out as the members of that community. That is to say, it is an effective way of claiming group identity and association. In addition, formulaic sequences, especially the ones categorised as stance markers, reflect the very own stand and style of the language user, not only in spoken but also in written language. To conclude, the advantages they offer range from their facilitating role in handling new, sometimes confusing linguistic situations to their time- and effort-saving processing feature; from their utility in setting one's tone, style, and voice in the course of production to their potential for achieving unity among different sort of communities.

2.2.4. Advantages of Formulaic Sequences in Academic Writing

Relevant to the subject and objectives of the current study, it would be practical to touch upon the benefits of using formulaic sequences in academic writing more deeply. Producing academic texts is quite arduous since it requires a lot of reading, in-depth analysis of existing literature, reasonable synthesis, and good knowledge of the language of all levels. A competent writer, for instance, should have a large lexicon consisting of terms, phrases, or expressions related to the subject area. Along with the lexicon, a competent writer should know how to put the words together correctly each time, and that might be pretty burdensome especially for lower-level writers. However, when the holistic approach to formulaic sequences, that is, the suggestion that FSs are stored in memory as single word units and retrieved from memory all at once, is accepted to be accurate, then it would not be daring to say that FSs significantly shorten the time and smooth the pave for generating sentences since language users do not need to check concord each time or to try to remember every single word separately. Besides their compact and easy storage and retrieval, their frequent and often specific (to various topics, contexts, channels, and sources.) use significantly facilitates the process of academic writing. As Coxhead and Byrd (2007) put it, “as a result of their frequent use, such sets [formulaic sequences] become defining markers of fluent writing and are important for the development of writing that fits the expectations of readers in academia” (p. 135). To be more precise, including sequences related to one’s subject area in writing makes the writer’s content register-, genre-, and discipline-specific, which, in turn, enables them to be accepted as a competent/proficient writer in the field. Nonetheless a text containing formulaic sequences is not always seen as satisfactory. Writers should pay meticulous attention not to overuse or underuse them or use them inaccurately because any case is argued to be a sign of incompetence or lack of knowledge in sheer amount of research articles. In his paper, Hyland (2008) mentions some of the frequently-used FSs in specific domains and states that their use in one’s very own product creates not only familiarity to audience but also sense of proficiency in the eyes of the target audience, thus underusing or total lack of those word units might signal “lack of fluency of a novice or newcomer to that society [target audience]” (p. 5). Similarly, Paquot (2008) points out that overusing some specific

sequences while avoiding the ones that are typical of native writers' indicates *foreignness* or idiosyncrasy. In brief, one's effective control over formulaic sequences specific to their field alleviates the challenging mental and mechanical process of academic writing and contributes to improving their linguistic competence. Moreover, it adjusts their style, tone, diction, and voice to the demands of the target audience and promotes native-like language use, which, in turn, gains them praise and approval in their field.

2.3. Literature Review

This section reviews previous studies regarding FSs are reviewed to portray the current literature on this topic. Although there are studies conducted with spoken language, the ones dealing with written corpora in academic contexts were preferred within the scope of the current research. Based on this brief literature review, it is possible to make the following conclusions. First, most of the studies have focused on comparing EFL learners from different proficiency levels in terms of their FSs use. Second, several studies compare EFL corpora with native corpora to reveal the effect of proficiency on FS use. Third, there are some scholars working with the writings of EFL learners from various L1 backgrounds to find out any L1 effect on the use of FSs. Fourth, almost all the studies examine FS use based on their frequencies and functions whereas there are also some studies that include a structural analysis as well. Finally, the number of such studies on FS use in the Turkish EFL context is very limited.

As stated above, the majority of the studies in FS literature have compared EFL learners from different proficiency levels regarding the frequency, functions and structures of FSs in their academic writings. For instance, Goulart (2020) examined a sub-corpus of written English at the University of Coimbra with an aim to explore if there were any differences in the type, frequency and structure of formulaic sequences across five proficiency levels (A1, A2, B1, B2 and C1 according to the CEFR. The corpus consisted of 624 written productions of 458 students who had 39 different L1s and came from 50 different countries. The students were taking Portuguese classes for foreigners at the University of Coimbra. The researcher took only 3-word bundles into

account for her study for some reasons which were discussed in her paper in detail on page 83. After a delicate analysis, she collected 356 formulaic sequences in total, whose biggest proportion belonged to the B2-level student writings. B1-level writing followed the B2 group, and the A1 and A2 level writing came next respectively. However, the researcher asserted that although the number of the types of FSs was the highest in the B2 group, A1 and A2 groups used more sequences in their writings as tokens, which signified the repetition of the same sequences in the lower language levels. When it comes to the structural features of the sequences across levels, it was observed that noun-based, prepositional and verb-based sequences were the most used sequence type at all levels. One of the three differences among the levels was that the reference point of the noun- and preposition-based sequences changed from concrete to abstract as the language level of the students increased; that is to say, the reference to people, objects, and places evolved into such abstract notions as time. The second difference was that B2-level productions contained more prepositional sequences in terms of both type and token than the previous levels. And the last one was that C1-level students tended to use fewer verb-, clause- and adverb-based sequences in their writing than the other levels.

The paper presented by Hejazi (2020) in an international conference focused on the effect of language proficiency on the use of formulaic sequences in academic texts. He studied on two written corpora, one consisting of 130 argumentative essays by 42 intermediate- and advanced-level ESL students in the UK and the other (the reference corpus) including 65 short essays written by proficient students majoring in the departments of linguistics and English language at different universities in the UK. Lists of 3- and 4-word sequences in the essays were compiled, and all the content-based sequences were cleared up to ensure that the sequences would only be related to academic skills. First of all, the author found out that the B2-level corpus contained a greater number of FSs compared to the C1-level corpus. Secondly, both groups of students were found to use 3-word sequences more frequently than 4-word bundles which were less prevalent in the B2-level corpus. Thirdly, almost none of the frequently-used bundles was shared between the two corpora. The researcher also identified some differences between the EFL students' and proficient language users' structural use of FSs. The former group applied phrase bundles more than clausal

bundles while the latter tended to use noun-based sequences more often in their texts. Besides, C1-level ESL learners were discovered to use verb-based sequences more frequently than B2-level ESL learners. Pointing out that increasing level of language proficiency did predict a wider variety of structures and functions of the FSs though the variety of FSs in the texts did not necessarily foster academic performance, Hejazi concluded that C1-level language learners' essays bore a closer resemblance to the reference corpus than the B2-level language learners' essays.

Similarly, Li and Volkov (2017) investigated the discourse functions of formulaic sequences in the written responses of students of varying proficiency levels to an e-mail writing part of the Canadian English Language Proficiency Index Program-General (CELP-IP-General) test, which was designed to determine the language levels of the learners. The aim was to inspect different discourse functions of formulaic sequences across proficiency levels. The first finding that the researchers came up with was that lower-level language users employed fewer formulaic sequences both in number and type. Across proficiency levels, there were 14 shared bundles which happened to occur in varying frequencies. In terms of functional use of the FSs, stance markers, discourse organising sequences and other sequences serving other functions made up of 90% of the total sequences at each language level, and referential sequences constituted the least common sequence. The researchers concluded that the occurrence and variety of the FSs in the corpus increased in tandem with the increasing language level, although functional use of FSs did not show significant differences across levels.

Ruan (2016) attempted to explore if Chinese students at an English-medium university tended to use more formulaic sequences in their written production and the variety of structures and functions of those sequences changed over time throughout their university education. The analysis of student corpus of academic writing showed that as students' level of English increased, the occurrence of FSs in their writing decreased while the range extended clearly. The study also revealed that the language level predicts the functional distribution of the sequences in that FSs fulfilled more functions in writing as the students progressed towards their senior years, and surprisingly, it

was found that discourse organising formulaic sequences were preferred more often by the students than stance or referential expressions regardless of language level.

In the same year, Appel and Wood (2016) published an article that focused on the effect of language levels on the use of formulaic sequences. They analysed the argumentative essays of EAP (English for Academic Purposes) students which were retrieved from the Canadian Academic English Language Test. The sequences to be examined were chosen to be minimum four and maximum seven words in length, and they were analysed quantitatively and qualitatively. One of the findings of the research was that lower-level essays contained a lot more 4-word sequences compared to higher level essays. Even when the 4-word sequences were combined with the coinciding more-word sequences, the situation did not change. However, the fact that lower-level texts were shorter in length than the higher-level ones might have resulted in a higher ratio of formulaic sequences to the existing words in the whole text. Regardless of the reason, the researchers summarise that lower-level groups were more dependent on the prefabricated chunks and used them repeatedly. When it comes to the functional use of the sequences, lower-level student essays consisted more of stance and discourse markers whereas higher-level student essays comprised referential bundles.

With the aim of disclosing the relationship, if any, between language proficiency and the use of formulaic sequences in writing, Staples et al. (2013) examined the formulaic sequences in participants' written productions in TOEFL iBT. As in Ruan's (2016) study, dependence on formulaic sequences was observed to reduce as language proficiency increased. Similarly, the researchers found that the lower-scoring participants used more prompt-based FSs in their texts than the intermediate-level and highest-scoring participants. Compared to the highest-level participants, intermediate-level participants were found to use more formulaic sequences in their writing. Moreover, unlike the lower-level participants, they did not stick to prompt-based sequences, indicating in this context that gradual autonomy in language enables the production of creative language with relatively fewer *prefabricated chunks*. Lastly, the research showed that the test takers, regardless of their proficiency level, had a tendency to use discourse-organising and stance expressions, rather than referential bundles, at similar rates.

Another study focusing on the use of formulaic sequences across proficiency levels comes from Huang (2015). He examined 5990 essays written by 3rd and 4th-year Chinese university students who were majoring in an EFL department to see a pattern, if any, in their use of formulaic sequences. It was pinpointed by the researcher that senior students' level of English is accepted as higher than junior students' according to the national language standards. After a year of a longitudinal study, the senior students' writings were observed to include more and a wider variety of sequences compared to the juniors'. The research showed a positive correlation between the improving level of language proficiency of students and their preference for use of sequences. However, the researcher warns to interpret the results cautiously since the literature is full of studies demonstrating contrasting results. The second significant finding of the study was that regardless of the language level, both groups tended to use five out of ten formulaic sequences quite often, which might signal L1 transfer. Lastly, the research revealed no significant difference between the two groups in their correct use of sequences, and this means proficiency does not necessarily predict the accurate use of formulaic sequences.

In short, the following conclusions can be drawn from the studies comparing FS use among EFL learners from different proficiency groups. First, it is not possible to state that there is a positive or negative correlation between the use of FSs and proficiency levels, since the studies in literature so far have demonstrated both an increase and decrease in the use of FSs as the proficiency level of learners increases. There are also studies which show an increase and then decrease in the number of FSs from A to C level. Second, the findings have indicated that referential expressions are the least common functional FS type at all proficiency levels, and when preferred, they were observed to occur more frequently at higher proficiency levels.

The second group of studies has examined FS use among EFL learners by comparing with native speaker corpora. For instance, Juknevičienė (2009) carried out a study making use of three corpora of three different language levels to gain an insight into the differences in the use of formulaic sequences between the Lithuanian students majoring in the department of English philology and native-speakers of English whose essays were retrieved from the LOCNESS corpus. The non-native speaker corpora

were consisted of first-year students' essays (AFK1 corpus) and third- and fourth-year students' essays (LICLE corpus), and the examined essays were related to both departmental subjects and general issues such as euthanasia, and educating students on sex at school. The first finding of the research was that AFK1 corpus included formulaic sequences, mostly verb phrases (VPs) and dependent clauses as stance and discourse organising bundles rather than referential bundles, similar to those used by native speakers of English in spoken discourse. Secondly, the essays in the native corpus, not surprisingly, were observed to retain much more features which were typical to a written discourse such as frequent use of noun and prepositional phrases referential bundles as referential bundles. Lastly, the LICLE was found to have more similarities with the LOCNESS than the AFK1 had, the ratio of topic-related bundles being the most obvious. In general, the LICLE shared more common features with the native corpus; however, both of the non-native corpora exhibited repeated sequences in use.

Wei and Lei (2011) compared advanced Chinese EFL learners' academic texts with professional writers' published articles to investigate the pattern of formulaic sequence use, and the study demonstrated similar results to those by Staples et al.'s. The academic texts produced by the advanced EFL learners were found to consist of manifold sequences, while professional writers did not draw much on those kinds of ready-to-use phrases. One of the other differences between the two groups was their preference of formulaic sequences to perform functions. The EFL learners were found to usually make use of sequences to form passives and to abstain from anticipatory *it* and participant-oriented phrases in their texts. Other than that, both groups used formulaic sequences to fulfil similar functions.

In their study, Jaworska et al. (2015) focused on the language of native speakers of German and English learners of German to investigate the frequency and functions of formulaic sequences in both groups' written productions. If the native group is accepted as more proficient speakers, although the researchers did not make such an assertion, it can be said that the results showed close similarities with many other proficiency-based studies. The English learners of German were found to use formulaic sequences more frequently than the native speakers of German did.

However, this finding does not mean that the former group used a wider range of sequences, since their texts included a limited number of sequences used repeatedly. On the contrary, native speakers' essays were richer in variety. The other finding put forth by the study was that there was a remarkable difference between the functional features of the sequences used by the language learners and the native speakers. While the native speakers of German were observed to use personal stance bundles frequently contrary to the research findings on the language of native speakers, the learners of German prefer to apply hedging to ensure impersonality.

The study by Lyu and Gee (2020) aimed to investigate the difference between the Chinese undergraduate students' graduation theses and American students' master theses in terms of the use of formulaic sequences. The research corpora revealed that the written productions of the Chinese students contained a larger number of types and tokens, and the shared sequences of the two corpora made only up 2% of the total FSs in the theses. When it comes to the structural features of the sequences in these two corpora, the noun phrases (NPs), prepositional phrases (PPs) and verb phrases (VPs) were observed to be equally dispersed as both types and tokens in the Chinese corpus while in the American corpus, the PPs outnumbered NPs and VPs as types, and NPs and PPs, being equally dispersed, surpassed VPs as tokens. Lastly, no statistically significant difference between the two groups of students was found in the functional use of sequences in terms of distributional percentage; however, research- and text-oriented functions were found to be larger in number in the Chinese students' theses than in the other group's theses.

Similarly, Lu and Deng (2019) examined the dissertation abstracts of Chinese and English-speaking postgraduate students in terms of structural and functional distribution and frequency of formulaic sequences. 13,596 abstracts of doctoral dissertations written in the English language by Chinese students and 4,755 by English students were investigated by the researchers, and the two corpora attained from Tsinghua University and MIT were carefully selected so as them to be at academically equal levels. 4-word bundles were compared in the two sets of data, and the findings were as follows. The researchers stated that the non-native corpora included a higher number of types and tokens of sequences, and it proportionally comprised more clause-

based and VP-based sequences than the NP- or PP-based ones. And VP-based sequences were used more frequently in the non-native data in every subcategory than in the native data. The functions of the sequences were grouped into three main categories in the research: research-oriented, text-oriented, and stance bundles, and it was revealed that the abstracts by the Chinese students included more FSs belonging to each category. The text-oriented sequences were the most common FS type in both groups' writings, while stance bundles were the least frequent. However, when compared, the Chinese students used proportionally more stance bundles than their native counterparts, and the English-speaking students used more text-oriented ones.

For the functional analysis of formulaic sequences in written productions, Nam (2017) worked with two corpora which included argumentative essays written by Korean junior and senior students and upper-level undergraduate American students, who were all majoring in English language-related departments. For the categorisation of the sequences, the researcher made use of the taxonomies by Biber et al. (2004) and Chen and Baker (2010), and he basically categorised the functions as “stance bundles, discourse organisers and referential bundles” (p. 708) and for the sequences that did not fit into any of these categories, he put them under the title of “others”. He analysed 4-word sequences that had occurred in at least five times in the non-native corpus and at least three times in the native corpus at least three different texts, and they were taken into account only if they had occurred in three or more texts. The examination of the functional use of the FSs in both corpora revealed that the two groups tended to use the FSs for significantly different functions. While the native corpus included referential bundles at quite a high ratio, they made up only a small part of the non-native corpus. Instead, the non-native corpus comprised mostly stance bundles and discourse organising bundles. Having referred to some existing studies on the relation between the use of FSs and proficiency levels, the researcher pinpointed that Korean students needed to be trained on the referential bundle to be regarded as proficient academic writers.

The study carried out by Qin (2014) presented results supporting Huang's first finding. She compared 136 articles produced by 20 graduate students of applied linguistics and 15 published research papers by 11 American professors who are native speakers of

English with corpora of 128 articles which were found in reputable journals on applied linguistics. With the assumption that the professors were more competent in the use of English, she tried to understand how the groups used formulaic sequences in terms of frequency and function. The results presented that the language users whose language levels were higher tended to use a wider variety of formulaic sequences and to use the sequences much more frequently. Moreover, the language users with a higher level of knowledge of language were observed to make use of the sequences to fulfil a larger variety of functions.

Ahmadi et al. (2013) designed their study to find out the pattern of the use of formulaic sequences in the research article abstracts of Iranian authors and English-speaking native authors in applied linguistics. The two bodies of corpora subjected to examination and comparison consisted of 200 abstracts produced by the Iranian and native English-speaking authors each, and the abstracts were all qualified enough to be published in recognised journals. As part of the study, only 4-word sequences which appeared at least in five articles were examined by the researchers, and they were found to occur more than twice as many sequences as in the non-native data. Also, out of the most frequent 10 sequences, none of them was shared between the two corpora. The structural features of the sequences varied between the corpora as well. The non-native speakers of English tended to use more phrasal sequences, verb phrases with dependent clauses succeeding and verb phrase fragments. And the functional distribution of the sequences showed that the sequences were mostly used for the statement of the purpose of the study, and then for giving a summary of the findings and drawing conclusions by both group of the authors. Moreover, both groups used fewer sequences while giving information about the methodology of their studies. In the conclusion part, the writers highlighted that as the speakers' level of English had increased, they had come across more formulaic sequences and functions specific to the genre.

Similar results to Qin's study were obtained from the study by Ädel and Erman (2012). They worked on a language corpus consisting of 325 essays of native and non-native English-speaking college students of different language levels majoring in the department of linguistics. The hypothesis of the research had been that non-native

speakers of English would produce fewer formulaic sequences both in number and variety, and the results verified this assumption. The corpus study demonstrated that the native speakers of English had utilised more sequences, and their writings were richer than the other group in terms of sequence variety. The finding was promising in the field of language education in that it demonstrates that non-native speakers' high level of knowledge of language did not necessarily enable native-like language use, at least in terms of the use of formulaic sequences.

Karabacak and Qin (2013) examined Chinese, Turkish and American college students' argumentative essays in comparison with a target corpus which was based on *New York Times* magazine to see the level of similarity between the two corpora in terms of use of formulaic sequence. Although the participants' language level was not indicated in the article, their academic level was stated to be quite similar, which might make it possible to interpret the results drawing on the L1 difference between the participants. The research showed that the American students applied the sequences the most frequently and elaborately (5-word bundles) among the three groups, and the Chinese students used them the least. However, the researchers were cautious about commenting on the American students' statistics because of the content-related usage. When the topic was about the US, for instance, 24 types of sequences that were used in texts produced by the American students did not occur in the other groups' texts. Or the Turkish and Chinese students might have avoided using some sequences due to a lack of thorough knowledge.

Durrant and Schmitt's study (2009) aimed to investigate the pattern of use of formulaic sequence in native and non-native texts through an examination of extended texts by native post-graduate students of an applied linguistics MA program at a British university along with essays which had an academic writing format in a magazine written by native speakers of English and relatively shorter texts by non-native English learners in an English for Academic Purposes. The non-native texts belonged to two groups of students: post-graduates from a pre-sessional English for Academic Purposes (EAP) program at a British university and freshmen in an in-sessional EAP program at an English-medium college in Turkey. To see if the findings were also applicable to even shorter texts, the researchers included into their research the essays

by pre-sessional students of a British university and the sub-corpus of the International Corpus of Learner English (ICLE) in the Bulgarian language. The study firstly revealed that the sequences used by the natives were of low-frequent type. Secondly, there was almost no difference in the use of collocations between the native and non-native group, though the non-native texts could be said to be characterised by a repetition of the “strong” (p. 174) collocations. Lastly, the collocations with “high mutual information scores” (p. 174) were found to be underused in the non-native texts. As it comes to the third group of texts, the shortest ones, they showed similar tendency with less apparent features.

Although the focus was not basically on the proficiency levels, Uçar (2017) carried out research to see the pattern of use of formulaic sequences between Turkish non-native scholars and English-speaking scholars. She compared the two groups’ 15 academic writings in terms of frequency, structure and function of the sequences they had used. The results supported many other study results in that Turkish non-native but proficient writers applied more sequences in their works with less variance. That is to say, the works of native scholars were richer in sequence type, with noun phrases’ being the most common, while the Turkish scholars preferred to use prepositional phrases the most. When it comes to functional use, referential bundles were observed to dominate the Turkish-speaking writers’ texts.

In short, the studies examining FS use by comparing native and non-native corpora have revealed the following conclusions. First, the academic writings of non-native speakers of English were found to include more FSs as tokens than the native speakers’, with a few exceptions. Second, the difference in the range of the FSs used by non-natives and natives varied across studies. In some studies, non-native texts were found to include a wider range of FS, while in others, native texts were richer in variety. Third, non-native texts were observed to comprise more verb phrases (VPs), whereas the most remarkable feature of native texts was that they were richer in prepositional phrases (PPs) and relatively richer in noun phrases than the non-native ones. Lastly, native speakers were found to use referential bundles more often than the non-native speakers did, and they were observed to use more text- and genre-oriented

FSs in general. The academic writings of non-native speakers, on the other hand, consisted mostly of stance or discourse marking FSs.

Third, there are several studies which were concerned with the variation in FS use among EFL learners regarding their L1 backgrounds. For example, Leńko-Szymańska's (2014) research is one of the many enlightening studies in the area that seeks to understand the probable impact of L1 differences among participants and language proficiency on the use of formulaic sequences. She collected the essays of students of varying grades, from 4 to 12, from different countries/states (Austria, China, Hong Kong, Japan, Poland, Israel, Taiwan, and Spain) where English is taught as a foreign language. According to their length, the essays were grouped into proficiency levels, the shortest being low, mid, and the longest being high, though it was highlighted that the levels were not accepted as the absolute language level of the students. The sequences to be examined were limited to 3-word bundles which were collected from native corpora, and native-like occurrences and tokens of sequences were counted. The findings were quite remarkable in that no influence of L1 on the use of formulaic sequence was found. Rather, the level of language proficiency seemed to have an impact. The general pattern that was found among all language levels was that the variety of the sequences in the texts tended to increase in parallel with the level. However, that increase took place in different stages with varying densities. It was steeper in the Spanish, Polish, and Austrian students' texts and was observed to take place between the middle and high stages in texts from Poland, Spain, Taiwan, and Israel. For Austrian students, the growth occurred between the low and middle stages, which was very similar for the Japanese students, though the increase in the variety of the sequences dropped in the highest stage. The increase in the frequency of the sequences also showed differences across texts and levels. While it was observed between the mid and high stages for the Spanish and Polish students, the Austrian students started to use the sequences more frequently in an earlier stage. Surprisingly, stability in the frequency was observed in the Taiwanese and Israeli students' texts, and the number of the sequences decreased in Japanese students' productions throughout the stages.

Öztürk and Köse (2016) conducted a study by examining three sub-corpora, a compilation of Turkish and English-speaking postgraduate students' MA and Ph.D. theses and English-speaking scholars published works to investigate the pattern in the use of formulaic sequences by those three groups. The researchers focused on the types, frequency, structures, and functions of the formulaic sequences in those three genres and revealed some interesting results. Relevant to current research, the theses by the Turkish students included far more sequences than the two other groups' academic prosés. They not only used more sequences in number, but they also employed more varied ones in their texts. Besides, the formulaic sequences in their texts were not much different from the other two groups', and even some of them were not identified in the other texts, which might signal L1 transfer. When it comes to the frequency of the FSs, the fact that Turkish students used the majority of the sequences repetitively makes it important to interpret the frequency statistics cautiously. Lastly, the researchers reported that there was no statistically significant difference in the structural and functional distribution of the sequences.

Another research focusing on the effect of different L1s on the use of formulaic sequence was conducted by Bychkovska and Lee (2017). They examined 101 essays produced by English-speaking college students and 105 by Chinese-speaking college students, both of which were argumentative essays and highly rated by the assessors. The results again showed that Chinese-speaking students used many more sequences in their writing in number and type. As identified by many studies in literature, the structures and functions of the formulaic sequences used by the two groups also showed different patterns. English-speaking native writers were found to use PPs and PPs, while L1-Chinese writers mostly used VPs although their texts included NP- and PP-based sequences more than half of the time. In terms of the functional distribution of the sequences, although the groups applied different types of sequences frequently to fulfil different functions, the researchers found no statistically significant distribution.

Considering the possible effects of different L1s on the structural and functional use of formulaic sequences, Pan et al. (2016) examined a corpus of English articles written by L1-Chinese and L1-English professionals and published in journals in the field of

telecommunication. One of the findings of the research was that Chinese writers applied similar sequences as the English ones in their writing but less frequently and with limited distribution. Secondly, Chinese professionals were found to have used various 4-word sequences in their writing, which the authors interpreted as an indication of greater dependence on formulaic sequences compared to the L1-English professionals. Regarding structural dispersion of the sequences, L1-English professionals' texts mostly included phrasal sequences; noun and prepositional phrases while the other groups used verb phrases (VPs) more. Lastly, L1-Chinese-speaking professionals were found to use, and sometimes misuse, the sequences to fulfil different functions, and they tended to use more stance bundles and much fewer research-oriented sequences than the L1-English-speaking professionals.

In brief, the studies concerned with the L1 effect on FS use among EFL learners have drawn the conclusion that it would not be misleading to say that there is no statistically significant difference among language learners in terms of use, frequency, functions and/or structures of the identified FSs regarding varying L1 background. Variations did occur among the groups; however, the researchers attributed these divergences to the proficiency levels and the tendency to overuse or underuse FSs.

Finally, two more studies are reviewed in this section: one focusing on the effect of instruction on FS use in academic writing and another with a longitudinal research design. First, the study on the effect of explicit instruction of formulaic sequences by AlHassan and Wood (2015) generated compelling results. Twelve participants with varying proficiency levels were provided with training which was designed to integrate 65 new sequences into the participants' repertoire. One of the main aims of the research was to see if the instruction would improve the participants' competence in writing and consequently help them get better evaluations by the judges. The scores given by two judges in the pre-, post-, and delayed-tests applied at the end of the training demonstrated that the participants could benefit from the instruction. Surprisingly though, the scores given by the third judge were found not to be influenced by the improving ability of the participants to use formulaic sequences rich in number and variety. Second, Li and Schmitt (2009) carried out a longitudinal study by observing the development of a Chinese MA student at an English-medium

university in Britain during her MA. They examined the lexical phrases she used in her nine assignments required to complete the program. Amy, the pseudonym of the participant, added 166 new lexical phrases to her repertoire. However, that did not result in an increase in the occurrence of lexical phrases in her assignments; rather, she preferred to make use of newly-learned phrases over the existing ones. What made a difference during the study was the accuracy of the phrases in the texts over time. Her texts clearly displayed a high rate of accurate use of phrases compared to the ones she had produced before. Although that might result from her improving level of English, the researchers asserted that the reason might also be that she overused some lexical phrases that she felt confident with.

3. METHODOLOGY

This section aims to give comprehensive information about the research methodology adopted to observe the pattern of formulaic sequence use across three proficiency levels; levels A, B and C. To this end, the first part explains the research design and why it was thought to be suitable for this study. The second part gives information about the data collection procedure. Finally, all the phases of data analysis were described in detail, and the academic resources were introduced within the scope of the objectives of the current study.

3.1. Research Design

The current study aims to investigate the use of tertiary level EFL learners' use of 3- and 4-word formulaic sequences in their academic writings and to reveal any patterns in their FS use across proficiency levels. The types and frequencies of the sequences in their writings and the functional and structural features of the identified sequences were the focal points of the study. To this end, a qualitative research design was utilized in this study due to various reasons. First, qualitative research is, as Patton (1985) defines it, "an effort to understand situations in their uniqueness as part of a particular context and the interactions there. [...] The analysis strives for depth of understanding" (p. 5). Likewise, an in-depth analysis of EFL learners' writings was performed in the context of the current study. Second, the above-mentioned aims of this study do not require any experimentation or manipulation. Therefore, the documents written by EFL learners in their natural setting (writing exams) were analysed. Third, this study does not include any statistical analysis as it is not aimed to correlate variables or to generalize from the research findings to all language learners although some numerical data like frequencies were obtained to interpret the non-numerical data. Finally, this descriptive study intends to examine solely the learner behaviour of FS use in academic writing and how it varies across proficiency levels. Therefore, this study was carried out through an investigation of existing

documents produced by the EFL learners studying at prep schools, and document analysis is one of the typical qualitative research techniques, together with observation of participant behaviour and interview with participants. (Creswell, 2009).

3.2. Data Collection Procedure

Since the present study was carried out over three sets of research corpora belonging to the language levels A, B and C, no participant had directly involved in the research. The research corpora comprised the academic writings of EFL learners in two distinct state universities in Istanbul, Turkey. A great majority of the EFL learners were studying in the preparatory programs of their own universities. However, the writing documents of the junior students were also included in the C-level research corpus because of the small number of C-level prep-class students, since most of the preparatory programs in universities in Turkey graduate their students with B1+ or B2 language level. The inclusion of the writing documents of the senior students might be accepted as one of the limitations of the study, but the fact that the documents belonged to the very first writing exam in the department would compensate for this limitation, in that the language level of the senior students would not increase considerably.

To clarify the concept of “academic writing” in the context of this research, it could be noted that the documents formed the research corpora were the EFL learners’ first writing exam papers of the academic year. Two questions might arise at this point as to if the documents reflected the true level of the EFL learners and what if there were a change in the use of formulaic sequences as the same students progressed through the end of the academic year. In the first place, it could be said that the academic writings were believed to reflect the real language level of the learners because the grouping of the learners was made after proficiency and placement tests, which were given at the beginning of the preparatory program in all universities in Turkey. Considering the second question, it could be accepted that it is one of the limitations of the study. The study had actually intended first to examine several academic writings of a relatively limited number of EFL learners throughout modules. That is, the writing exam documents had been requested from the institutions which apply a modular system in their preparatory program. The institutions either did not prefer to

share the documents with the researcher because of the inconveniences that the unexpected Covid-19 pandemic resulted in, or they offered to share online writing exams. However, the online exams would not be perfectly reliable data because the learners might have benefitted from online tools, dictionaries, or expert guidance during the exam. Also, cheating became more common in online testing during the pandemic which created concerns about the reliability of tests. As a result, the documents were obtained from the institutions that apply a semester system. To transcend this limitation and to maximise internal validity, as many documents of different lengths as possible were investigated at each language level.

In order to get the exam documents from the preparatory schools, some official procedures were followed. First, the ethics committee approval was obtained from Istanbul Medeniyet University Faculty of Educational Sciences Ethics Committee. Second, the official permissions were gained from the Rectorates and Administrations of the Schools of Foreign Languages by submitting petitions for permission with the research ethics committee's approval. Then, the institutions shared the required exam documents as scanned PDF documents by concealing all the identity information like names or student numbers.

3.3. Research Setting

The two universities from which the research data were collected apply a semester system in their preparatory programs. In this system, learners sit a placement test at the very beginning of the academic year, and they are placed in classes according to their language levels described in the CEFR. A-level classes include beginner and elementary learners whereas B-level classes are formed with pre-intermediate and intermediate learners. Finally, C-level classes are for upper-intermediate students. One academic year consists of two semesters, named as fall and spring, and students keep being instructed in the same groups (A, B or C) throughout the year even though their language levels improve and their classes (but not levels) are changed at certain periods. Learners take such exams as writing, speaking, mixed-type, mid-term and final regularly until the end of the academic year. Percentage points from those exams

and proficiency exams at the end of the year are then calculated to decide if a student passes or fails the program. Since each exam partly determines EFL learners' pass/fail status, exams are one of the significant assessment tools for both learners and practitioners, and writing exams are maybe one of the most significant ones. The preparatory programs of the two universities place increased emphasis on developing writing skills, which is an irreplaceable element of academic literacy, because one of them is an English-medium university, and the other one has many English-medium departments. Within the context of writing exams, students are expected to write essays of varying types; descriptive, analytical, persuasive, and critical. At beginner levels, short paragraphs or short essays are acceptable forms of writing. In writing exams, students produce their texts within the scheduled time under teacher observation, and they are not allowed to make use of any kind of sources, neither published nor online. The exam documents are then evaluated according to a rubric by two distinct raters to ensure interrater reliability.

3.4. Data Analysis

In the framework of the present study, the analyses of the 3- and 4-word formulaic sequences were performed in three contexts. Firstly, the types and frequencies of the FSs in the research data were examined at all proficiency levels separately. Secondly, the identified FSs were functionally analysed. And lastly, each FS was structurally categorised. These three phases were preceded by the identification of the 3- and 4-word FSs in the writings of the EFL learners.

3.4.1. Identification of 3- and 4-Word Formulaic Sequences Across Proficiency Levels

In the first phase of the research, three distinct research corpora appertaining to the language levels A, B and C were generated from the writing exam documents provided by the universities. Then, each corpus was analysed separately to identify all the possible 3- and 4- word formulaic sequences by means of a free software programme #LancsBox, “a new-generation software package for the analysis of language data and corpora developed at Lancaster University” (<http://corpora.lancs.ac.uk/lancsbox/>).

Along with many valuable functions, it helps researchers form their own corpus and work on them by generating n-grams. After the identification of the so-called 3- and 4-word formulaic sequences in the research corpora, the FSs whose occurrence rate is below five times were excluded from the research context. In other words, the study set at least five time of occurrence as a cut-off point for a word group to be recognized as a formulaic sequence. The leftover formulaic sequences were then checked against a native corpus *The Louvain Corpus of Native English Essays (LOCNESS)*, which consists of the writings of A-level British pupils and British and American university students, to see if they also existed in a native corpus. If not, they were also excluded from the research context. The remaining formulaic sequences were again checked against another native corpus *The British National Corpus (the BNC)* for verification. The parameter this time was that the formulaic sequences identified in the research corpora must occur 20 times or more per million words in the native corpus. The cut-off point of 20-40 times of occurrence per million was one of the parameters that Chen and Baker (2010) set in their paper. After all these processes, the remaining 3- and 4-word formulaic sequences were listed separately from the most frequent to the least and tabularized in *Microsoft Excel* with their frequency of occurrences both in the research corpora and in the BNC to be used in the course of the analysis of types and frequencies. Along with the examination of the FSs in their own categories, i.e., 3-word FSs in A-level corpus, or 4-word FSs in B-level corpus, the shared 3- and 4-word sequences among the three research corpora were also identified and their frequencies in each corpus were compared to each other. Since the number of the tokens in each corpus varied, the comparison could only be made after the number of the tokens in each corpus was proportioned to 20,000, which was the highest number of tokens among three distinct corpora.

3.4.2. Identification of Functions of 3- and 4-word Formulaic Sequences Across Proficiency Levels

As a second step, all the 3- and 4-word formulaic sequences were functionally analysed to see whether the EFL learners of varying proficiency levels displayed a similar tendency towards the functional use of the FSs. Biber et al.'s taxonomy on functions (2004) provided a basis for the analysis in this phase. They specify three

main functional categories for formulaic sequences in their paper, being stance bundles, discourse organising bundles and referential bundles.

Stance bundles were defined as frames for the expressions of attitudes or allegation of certainty shaping following statements. They could be of two types; epistemic stance and attitudinal/modality stance bundles. The former is mostly personal and used to show the speaker's/writer's certainty or uncertainty, while the latter, still personal, is used to convey the person's approach towards the cases in the following statements.

Second, discourse organising bundles fulfil structural and organisational functions in the discourse, as Appel and Wood (2016) word it. They could be used for their numerous functions, such as introducing, clarifying, or summarising a topic, elaborating on the issue, and linking parts of sentences or paragraphs. They are relatedly divided into two sub-categories being topic introduction/focus bundles and topic elaboration/clarification bundles.

The last major category of functions is referential bundles, used to identify a subject matter, mention or explain aspects of an entity, and point out concepts of time, place, and manner. Biber et al. (2004) specify four sub-categories for referential bundles. The first one is identification/focus bundles, which, as the name suggests, identify or focus on the subject matter. The second is imprecision bundles, which are used "either to indicate that additional references of the same type could be provided" (Biber et al., 2004: 394). The third is the bundles that specify attributes. They could be used to refer to abstract features, mention the size, sort or form of an entity, build up the relationship between the concepts, or events. And the last category is time/place/text-deixis bundles, which refer to the time of an action/event, location of an entity/concept or some specific parts of a text in the discourse.

As the Biber et al.'s (2004) categorisation of functions being the basis, all the 3- and 4-word formulaic sequences identified in the research corpora were categorised according to the EFL learners' proficiency levels. The most and the least frequently used FS types of functions were compared and interpreted within and between each language level to investigate a possible pattern across the groups. And some student excerpts that were randomly chosen from the research corpora were included in the

results of the study to exemplify each major category of functions at all proficiency levels.

3.4.3. Identification of Structural Features of the 3- and 4-Word Formulaic Sequences Across Proficiency Levels

The third and the last step of the study was to carry out a structural analysis of the 3- and 4- word formulaic sequences in the research corpora in order to observe if the structural types of FSs were relevant to proficiency levels. The findings of each level were first evaluated within their own contexts before they were compared and contrasted with each other. In the course of structural analysis, Biber et al. (2004) again provided an invaluable guide in that the identified FSs were categorised according to their structure mainly under three major categories: formulaic sequences incorporating 1) *verb phrase* (VP) fragments, 2) *dependent clause* fragments and 3) *noun phrase* (NP) and *prepositional phrase* (PP) fragments.

According to Biber et al. (2004), the FSs that incorporate VP fragments consist of seven sub-categories. They are *(connector +) 1st/2nd person pronoun + VP fragments*, *(connector +) 3rd person pronoun + VP fragments*, *discourse marker + VP fragments*, *verb phrase with non-passive verb*, *verb phrase with passive verb*, *yes-no question fragments* and *WH-question fragments*. The dependent clause fragments have five distinct sub-categories as *1st/2nd person pronoun + dependent clause fragments*, *WH-clause fragments*, *if-clause fragments*, *(verb/adjective+) to-clause fragments* and *that-clause fragments*. And the last major structural category of NP and PP fragments includes the sub-categories of *(connector +) noun phrase with of-phrase fragments*, *noun phrase with other post-modifier fragments*, *other noun phrase expressions*, *prepositional phrase expressions* and *comparative expressions*. Although the researchers usefully summarise the main and sub-categories of structural types of FSs in their paper with several good and representative examples, it was seen that some of the FSs in the research corpora did not fit into any of the categories specified in Biber et al. (2004). At that point, another taxonomy was used as a basis for the categorisation of the remaining FSs, which is designated in the book by Biber and his colleagues (1999); the *Longman Grammar of Spoken and Written English* (LSWE), and those FSs

were grouped under the category of “others”. The book is actually very comprehensive in that each category includes numerous samples of formulaic sequences of different length as corpus findings and discussions related to the findings. (See pages 1002-1027). Though the book introduces the abundant number of categories, the aforementioned remaining FSs were found to fall into seven major and sub-categories of structures. They were quantifier expressions, adverbial clause fragments, other adverbial clause fragments, other expressions, pronoun/noun phrase + (copula) be, prepositional phrases with embedded of-phrase fragments, and other prepositional phrase fragments. Each structural type was summarised in Table 1 below with sample sentences, most of which were taken from the research corpora.

Table 1. Structural Features of Formulaic Sequences with Samples

Structural Category	Structural Feature	Samples
Lexical Bundles That Incorporate Verb Phrase Fragments	(connector +) 1 st /2 nd person pronoun + VP fragments	When you live in a foreign country, you have to speak English to communicate with people.
	(connector +) 3 rd person pronoun + VP fragments	If you want to learn a foreign language from online classes, you need to have good microphone and camera.
	Discourse marker + VP fragments	You know it was one of the drawbacks of distance learning.
	Verb phrase (with non-passive verb)	First of all, people are able to know everything that happens in other places of the world and reaching [this information] makes the world become a small village.
	Verb phrase with passive verb	First of all, most of the people who are using social media want to be liked by other people.
	Yes-no question fragments	Do you want to attend classes regularly or to watch the recordings when you are available?
	WH-question fragments	How many of you are competent enough to write an argumentative essay to point the issue?
Lexical Bundles That Incorporate Dependent Clause Fragments	1 st /2 nd person pronoun + dependent clause fragments	I believe that online shopping websites/apps have made people become addicted to shopping.
	WH-clause fragments	When I was a child, my father bought a computer.

Lexical Bundles That Incorporate Noun Phrase and Prepositional Phrase Fragments	<i>If</i> -clause fragments	[...] or if you do not know how to use it, there can be problems with your learning.
	(Verb/adjective+) <i>to</i> -clause fragments	Children might need assistance to be able to limit the time they spend in front of TV.
	<i>That</i> -clause fragments	At the same time, I stopped saying that I was the best at video games.
	(connector +) Noun phrase with <i>of</i> -phrase fragments	Education is one of the most essential ingredients for the development of not only the individual but also the nation as whole.
	Noun phrase with other post-modifier fragment	I support the idea that children at early ages should not use technological devices for some reasons.
	Other noun phrase expressions	Secondly, the only thing that society expect from women is running the house and taking care of the children.
	Prepositional phrase expressions	During the live classrooms, teacher has a webcam connected [to] his/her computer or laptop, and the students can hear the voice of their teacher and watch him/her at the same time and can take notes, ask questions or talk to [the] teacher.
	Comparative expressions	However, we must remember that the more people reach proper education means the more advanced and developed individuals as well as nations.
	Quantifier expressions	While I was trying to beat up my feelings, I faced a lot of obstacles.
	Adverbial clause fragments	First of all , women cannot find a job properly.
Others	Other adverbial clause fragments	To sum up, online education is not a good way to learn a foreign language because it is not [efficient,] and it has many problem[s].
	Other expressions	As a result , children should not use technological devices at a very early age for their education and health.
	Pronoun/Noun phrase+(copula) be	[...] there are many restaurants in Alexandria, and it serves local food so that tourists could taste delicious Egyptian food.
	Prepositional phrase with embedded <i>of</i> -phrase fragments	Thus, traditional education can be more creative and attractive in terms of analysis.
	Other prepositional phrase fragments	He, on the other hand , was trying to find a job in İstanbul at the time of economic crisis.

As can be seen in Table 1, there are four major categories for the structural aspects of formulaic sequences in the study. The categorisation of the 3- and 4-word FSs in the research corpora was devised by means of the specifications in Biber et al. (1994) and Biber et al. (2014), and again the findings were interpreted both individually, that is, within the language group, and in relation to each other by comparing and contrasting.

3.5. Reliability and Validity of the Study

3.5.1. Reliability

The lists of 3- and 4-word formulaic sequences in the research corpora at each proficiency level were finalized after the identified FSs had been checked against two distinct native corpora. The first native corpus was the Louvain Corpus of Native English Essays (LOCNESS). As its name suggests, the LOCNESS comprises the writings of native English speakers. The second corpus was the British National Corpus (BNC), which is a remarkably wide collection of written and spoken British English from a broad range of linguistic sources. The comparison of the FSs with these two invaluable native corpora was one of the means to ensure reliability in the research. The other means was the identification and categorisation processes of the FSs in the research data. All the FSs were first categorised by the researcher herself, and they were then double-checked by the supervisor. The status of the FSs were negotiated, and the inter-rater reliability was calculated as .91 by means of the formula by Miles and Huberman (1994). When needed, native speakers of English and a couple of experienced teachers were consulted for intuitive and/or qualitative judgements. These processes were the mainsprings of reliability in the context of the current study.

3.5.2. Validity

As mentioned before, the study was carried out through the investigation of existing documents which were the writing exam documents of the EFL learners in two distinct state universities. The exams were taken face to face without the use of dictionary, or

the assistance of teachers, peers, or online tools. In other words, the collected documents were the learners' immediate linguistic output in an academic context. This was thought to be a creditable parameter to establish research validity. Similarly, the native corpora against which the FSs in the research data were checked consisted of real-time productions of native English speakers. They were academically and internationally-accepted sources of linguistic production, which again helped establish research validity.

4. RESULTS

The first objective of the current study is to identify 3- and 4-word formulaic sequences (FSs) used in the writing exam documents of A-, B- and C-level EFL learners in two distinct state universities in İstanbul. Upon the identification of the FSs in the research corpora, the researcher aims to examine the types and frequencies of those FSs across the three language levels. Correspondingly, the first part of the “Results” presents the findings about the types and frequencies with tables and explanations for the data displayed in the tables. The third objective of the study is to find out about the functions of all the identified FSs, so the “Results” part proceeds with the results of the functional analysis of the 3- and 4-word FSs in the research corpora. Finally, the researcher focuses on the structures of the identified FSs at three distinct language levels and addresses the structural categories of the FSs in the closing part. To conclude, the “Results” part includes the findings of the analyses of types, frequencies, functions and structural forms of the 3- and 4-word formulaic sequences found in the writing exam documents of the EFL learners.

4.1. Findings Regarding the Types and Frequencies of the Formulaic Sequences

After the research data were obtained from the writing exam documents of university students of three language levels; being A, B and C, and analysed to identify the most frequently-used 3- and 4-word word formulaic sequences (FSs) in the research data through #LancsBox, corpus check was implemented to ensure the identified FSs in the research data match with the native corpus LOCNESS first and then the British National Corpus (BNC). The FSs that met the criteria were then tabularised. The tables provide some information about the number of tokens, types and lemmas in the research corpora, types and frequencies of the identified FSs, their rate of occurrence in the native corpora and shared FSs and their frequencies in all the three corpora.

4.1.1. 3-Word Formulaic Sequences in A-Level Writing Exam Documents in terms of Type and Frequency

The writing exam documents of five different A-level groups were collected from two different universities. They were analysed using #LancsBox by being compared to LOCNESS and then manually to British National Corpus (BNC) to identify the most frequently-occurring formulaic sequences. A list of FSs in the research data was compiled, and the FSs were lined up as to be from the most to the least frequent. Table 2 below presents this list of sequences with their frequency values in both the A-level research corpus and the native corpus BNC.

Table 2. Comparison of 3-Word Formulaic Sequences in Data of Level A with British National Corpus

Level A		British National Corpus (BNC)	
19,076 tokens		100,119,205 tokens	
2,281 types			
2,155 lemmas			
Formulaic Sequence (FS)	Frequency of Use in Research Data	Frequency of Use in BNC	Ratio of Frequency in BNC
a lot of	49	8,885	88.74
he is a	25	335	32.1
do not like	24	1,381	138.60
there are many	19	2,138	21.35
you can go	18	360	36.13
I want to	16	3,459	34.549
first of all	15	764	76.68
I do not	15	3,024	30.204
he does not	14	773	77.58
is the most	14	335	335
because it is	14	515	33.59
you have to	13	3,262	32.581
you do not	12	5,106	236.28
you want to	9	3,351	33.470
it is very	9	482	23.13
one of the	8	32,284	322.456
I think that	8	2,784	279.42
in the world	8	5,58	55.734
was born in	7	560	26.88
do not know	7	9,034	906.69
I have a	7	525	52.69

is in the	6	2,774	27.707
but it is	6	5,765	57.581
go to the	6	863	86.61
It is a	6	8,916	89.054
and you can	5	653	65.54
we did not	5	916	91.93
you go to	5	412	41.35
we do not	5	2,225	223.31
I had a	5	642	64.43
I think this	5	313	31.41
have to go	5	724	72.66
is not a	5	3,676	36.716
I believe that	5	201	20.17
do not have	5	1,552	155.77
you can see	5	781	78.38
as a result	5	7,584	75.750
in the city	5	448	42.80

The data collected from the writing exam documents of A-level prep class students comprise 19,076 tokens of 2,281 different types. The whole data were analysed to find out 3-word formulaic sequences in student writings, and the cut-off point was set as at least five times of occurrence. As a result, 38 types of 3-word formulaic sequences were identified in the writings of A-level language learners. The sequence *a lot of* was found to be by far the most frequently-used bundle with an occurrence of 49 times. *He is a* and *do not like* came after with an occurrence of 25 and 24 respectively. The list of 3-word FSs in the data shows that the A-level students preferred 3-word sequences that were recurring at high rates in the native corpus. Surprisingly though, the top ten FSs in student writings did not have the greatest frequency in the native corpus, with three exceptions; *you do not*, *do not like* and *is the most*. They were observed to appear more than 100 times per million. On the other hand, the analysed student writings were found to include FSs that had genuinely high rates of use in the native corpus. The sequences *you do not*, *one of the*, *I think that*, *do not know*, *we do not* and *do not have* are some illuminating examples to it.

4.1.2. 3-Word Formulaic Sequences in B-Level Writing Exam Documents in terms of Type and Frequency

The research data belonging to the language level B were obtained again from the writing exam documents of four different language classes at two different universities. The formulaic sequences in the data were identified with the same methods as the A-level data. They were listed from the most frequently-used to the least, and Table 3 below introduces the list of the identified FSs in the writings of B-level students and includes some numeric information as to their frequency of use both in the research data and the native corpus.

Table 3. Comparison of 3-Word Formulaic Sequences in Data of Level B with British National Corpus

Level B		British National Corpus (BNC)	
14,757 tokens		100,119,205 tokens	
1,567 types			
1,420 lemmas			
Formulaic Sequence (FS)	Frequency of Use in Research Data	Frequency of Use in BNC	Ratio of Frequency in BNC
they do not	34	2,263	22.603
you do not	27	5,106	512.46
do not have	21	1,552	155.77
a lot of	18	8,885	88.744
first of all	14	764	76.68
not want to	12	1,565	157.07
we do not	11	2,225	223.31
not have to	10	2,953	26.342
the most important	9	3,159	31.552
I believe that	8	201	20.17
as a result	8	7,584	75.750
you want to	7	3,351	33.470
in the future	7	2,212	22.094
the idea that	6	563	36.72
one of the	5	32,284	322.456
a number of	5	14,217	142.001
it is not	5	10,035	100.231
I think that	5	2,784	279.42
do not know	5	9,034	906.69
if you are	5	3,917	39.123
go to the	5	2,247	22.443

I want to	5	3,459	34.549
I do not	5	3,024	30.204
you need to	5	630	63.23

As could be seen in Table 3, out of 14,757 tokens and 1,567 types in the B-level research data, 24 different types of 3-word formulaic sequences with an occurrence of 5 times or more were identified in the exam documents. *They do not* was found to be the most preferred sequence with an occurrence of 34, followed by *you do not*, which also had a high rate of use, 27. An interesting point should be stressed here: While the most-frequent sequence (*they do not*) in the research data was one of the least recurring FS in the native corpus, the second most common FS (*you do not*) in the research data was surprisingly the second most-recurring FS in the native corpus, among the shared FSs between the two corpora. When the remaining sequences on the list were compared to the native corpus in terms of the ratio of use, no clear parallelism was observed between the two. While some sequences were high in frequency in the research data, they were not among the most repeatedly-used ones in the BNC. Nevertheless, it should be highlighted that most of the identified FSs in the B-level research data was extensively used by native speakers of English.

4.1.3. 3-Word Formulaic Sequences in C-Level Writing Exam Documents in terms of Type and Frequency

The research data for language level C come from two classes of EFL students at a prep school and at an ELT program. The analysis of the written data from these two groups revealed that although the number of classes, and so the number of writing documents was lower than the other two language levels, the amount of the tokens surpassed those of the other two groups, since higher proficiency level compared to the other groups enabled the C-level learners to write long essays. Table 4 presents necessary information about the frequencies of the identified 3-word FSs in the research data and the BNC.

Table 4. Comparison of 3-Word Formulaic Sequences in Data of Level C with British National Corpus

Level C		British National Corpus (BNC)	
20,045 tokens		100,119,205 tokens	
2,401 types			
2,070 lemmas			
Formulaic Sequence (FS)	Frequency of Use in Research Data	Frequency of Use in BNC	Ratio of Frequency in BNC
I did not	21	6,533	65.252
a lot of	17	8,885	88.744
you do not	16	5,106	512.46
I could not	16	2,942	29.385
most of the	13	7,999	79.895
when I was	12	2,259	22.563
you have to	11	3,262	32.581
one of the	10	32,284	322.456
to go to	10	3,195	31.912
I had to	10	2,421	24.181
as a result	10	7,584	75.750
they do not	10	2,415	24.121
the most important	9	3,159	31.552
and it is	9	5,851	58.440
be able to	9	11,164	111.507
there is no	8	11,205	111.917
there are some	8	1,322	132.042
do not have	8	2,299	22.963
in terms of	7	8,596	85.858
did not know	7	2,747	27.437
does not matter	7	606	60.82
do not know	7	6,917	69.088
not have to	6	1,044	104.78
because of the	6	6,574	65.662
I was a	6	313	31.41
if you do	6	2,353	23.502
the only thing	6	470	47.17
it is not	6	10,035	100.231
not want to	6	4,138	41.331
there is not	6	434	43.56
and I had	6	337	33.82
want to be	6	416	41.75
I was in	6	377	37.84
he or she	6	555	36.20
we do not	5	2,554	25.510
in front of	5	5,773	54.665
that I was	5	287	28.80

and I was	5	801	80.39
have to go	5	724	72.66
as well as	5	17,622	176.010
in order to	5	11,549	115.352
they have to	5	418	41.95
they have a	5	281	28.20
is one of	5	4,8	47.943
are able to	5	373	24.33
one of them	5	2,702	26.988
was going to	5	3,677	36.726
we have been	5	297	28.38
I do not	5	18,502	184.800
did not want	5	2,2	21.974

Table 4 shows that the corpus of language level C consisted of 2,401 types and 20,045 tokens. After the 3-word FSs recurring at least five times in the research data were extracted, 50 different types of sequences were observed to find use among C-level language learners. The table demonstrates that the frequency of each type of FS was very close to each other. In other words, it is difficult to say that any formulaic sequence in the list outnumbered the remaining ones. Only a slight discrepancy between the first two FSs; *I did not* and *a lot of*, was observed in such a way that *I did not* occurs 21 times while *a lot of* occurred 17 times. When the corresponding ratio of the frequency of the FSs in the native corpus was examined, it could be noticed that there was no parallel increase or decrease across the most and the least frequent sequences. To give an example, the third FS in the list of the research data had the highest frequency in the native corpus whereas the fourth FS had the lowest frequency. On the whole, it would not be misleading to state that the most common FSs in the research data, except for *you do not*, *one of the*, *as well as* and *I do not*, do not seem to exhibit noticeably high density in the native corpus.

4.1.4. Comparison of 3-Word Formulaic Sequences Across Language Levels

After the identification of the formulaic sequences at each language level separately, an analysis of common sequences across three levels was carried out. Since the number of the tokens in each corpus belonging to the distinct language levels varied, the frequencies of the FSs were calculated on 20,000 tokens, which was the approximate number that the three corpora could find common ground, to make a valid comparison

between them. The table below shows the shared formulaic sequences among language levels A, B and C with their frequencies per 20,000 tokens.

Table 5. Shared Formulaic Sequences Across Levels

Formulaic Sequence	Frequency in Level A	Frequency in Level B	Frequency in Level C
a lot of	51	24	17
I do not	16	7	5
you do not	13	37	16
one of the	8	7	10
do not know	7	7	7
we do not	5	15	5
do not have	5	28	8
as a result	5	11	10

Table 5 demonstrates that the academic writings of the students of differing language levels included eight formulaic sequences in common with varied frequencies. It is not possible to say that a type of formulaic sequence was favoured most by all the students. However, the sequence *a lot of* could be stated to have quite a high occurrence rate in all three corpora. In fact, it was the most frequently-used sequence among the eight sequences in the corpora of level A and level C, and it was also one of the most recurring FSs in the corpus of level B. The frequencies of some sequences such as *one of the*, *do not know* and *as a result* were observed to be close to each other, insomuch that *do not know* was actually equally used by all the three levels. Nevertheless, it should be noted that it is again not much viable to make obvious inferences as to the most favoured sequences by the three levels.

4.1.5. 4-Word Formulaic Sequences in A-Level Writing Exam Documents in terms of Type and Frequency

The A-level research data were obtained from five different prep-classes of two different universities in İstanbul. The A-level research data were compared to two distinct native corpora. After the comparisons, the final list of 4-word FSs was tabulated in Table 6 with frequency rates of FSs in the research corpus and in the BNC.

Table 6. Comparison of 4-Word Formulaic Sequences in Data of Level A with British National Corpus

Level A		British National Corpus (BNC)	
19,076 tokens		100,119,205 tokens	
2,281 types			
2,155 lemmas			
Formulaic Sequence (FS)	Frequency of Use in Research Data	Frequency of Use in BNC	Ratio of Frequency in BNC
you do not know	5	518	51.99

Out of 19,076 tokens in the research data, only one formulaic sequence was found to appear in the native corpus, which was *you do not know*. It occurred five times in the research data and 52 times per million in BNC. However, it should be noted that only the FSs that occurred at least five times in the research data and 20 times per million in the native corpus were taken into consideration within the scope of this research.

4.1.6. 4-Word Formulaic Sequences in B-Level Writing Exam Documents in terms of Type and Frequency

As for each set of data at three language levels, the 4-word FSs were checked against two distinct corpora, and the table of the results were finalised according to the second corpus check, which was a comparison with BNC. Upon the analysis of the data of four B-level prep classes, the identified 4-word formulaic sequences in the research data were listed in Table 7 below, in which one can see the frequencies of 4-word sequences identified in both corpora.

Table 7. Comparison of 4-Word Formulaic Sequences in Data of Level B with British National Corpus

Level B		British National Corpus (BNC)	
14,757 tokens		100,119,205 tokens	
1,567 types			
1,420 lemmas			
Formulaic Sequence (FS)	Frequency of Use in Research Data	Frequency of Use in BNC	Ratio of Frequency in BNC
you do not have	11	595	59.72
do not want to	10	3,324	29.651
do not have to	10	705	70.76
at the same time	5	6,462	64.543

The total number of the tokens in B-level students' writing exam documents is 14,747 with 1,567 different types. Four 4-word FSs occurring five times in the research data and 20 times per million in the native corpus were identified in the data. The most frequently-used FS is *you do not have* with a frequency rate of 11 times, which was closely followed by *do not want to* and *do not have to* with a frequency rate of 10 times. The fourth and the last 4-word FS found in the data was found to recur 5 times in the research data, and it was observed to occur half as many times again as the other three FSs. The comparison of the two sets of corpora revealed no linear relationship between the frequencies of the shared FSs in both corpora. In other words, increases and decreases in the frequency of identified FSs in the research data did not go in the same direction in the native corpus. To give an example, the FS having the highest occurrence rate in the research corpus did not match with an FS that has the highest rate in the native corpus. However, it could be seen that B-level language learners tended to include native-like 4-word formulaic sequences in their academic writing hinging on the fact that three of the sequences appeared in the native corpus over 50 times per million whereas the second most frequent FS in the research data occurred approximately 30 times in the native corpus.

4.1.7. 4-Word Formulaic Sequences in C-Level Writing Exam Documents in terms of Type and Frequency

Succeeding the identification of all the 4-word combinations in the corpus of the C-level language learners, the FSs appearing at least five times in the research data and encountered in the native corpora at least 20 times per million were listed to be scrutinised in detail. Table 8 below displays the types and frequencies of the identified FSs comparatively with the British National Corpus.

Table 8. Comparison of 4-Word Formulaic Sequences in Data of Level C with British National Corpus

Level C		British National Corpus (BNC)	
20,045 tokens		100,119,205 tokens	
2,401 types			
2,070 lemmas			
Formulaic Sequence (FS)	Frequency of Use in Research Data	Frequency of Use in BNC	Ratio of Frequency in BNC
on the other hand	14	5,054	50.480
at the same time	6	6,462	64.543
do not have to	6	705	70.76
one of the most	6	3,969	39.643
if you do not	6	724	72.66
at the end of	5	8,26	82.502

The C-level research data comprising of 20,045 tokens and 2,401 types were found to include six types of 4-word formulaic sequences. *On the other hand* was, as seen, the most preferred 4-word formulaic sequence in the research data with an occurrence of 14 times. It appeared more than twice as much as the other frequently-used FSs in the data. The sequences *at the same time*, *do not have to*, *one of the most* and *if you do not* were found to occur 6 times, and the last 4-word FS *at the end of* had a frequency of 5 times. Here again, there seems to be no direct or inverse proportion between the research and native corpus regarding the frequency of use of the identified FSs. *At the end of*, for instance, has the highest rate of occurrence in the native corpus while it has the lowest in the research data among all the six sequences. Even so, the FSs in question could be reported to have quite a high frequency in the native corpus; the highest being 82.502 times and the lowest 39.643 per 100 million tokens.

4.1.8. Shared 4-Word Formulaic Sequences across Language Levels A, B and C

Four and six 4-word bundles were identified for B and C level groups whereas there was only one from group A. When the identified 4-word formulaic sequences at each level were compared, it was seen that no 4-word FS was commonly used among the research corpora, contrary to the 3-word FSs.

4.2. Functional Categorisation of the Formulaic Sequences in the Research Data

To find out the relationship between the language levels of the students and the use of formulaic sequences, an analysis of the types of formulaic sequences in student writing exam documents and their frequencies across documents was carried out within the scope of current research first. With the assumption that unaccompanied analysis of the types and frequencies only would not be sufficient enough to show the relation, if any, between the use of FSs and proficiency level, a functional analysis of each 3- and 4-word formulaic sequence in the research corpora was performed in the light of the functional categorisation of formulaic sequences in *Biber et al. (2004)*. The study introduces three main categories for functions: stance expressions, discourse organisers and referential expressions, and each category has several subcategories. *Biber et al. (2004)* examine stance expressions either as epistemic stance or attitudinal/modality stance bundles, discourse organisers as topic introduction/focus or topic elaboration/clarification bundles, and referential expressions as identification/focus, imprecision or attribute-specifying bundles. Although current research focused only on the main and their subcategories during analysis, it should be noted that some of the subcategories in *Biber et al. (2004)* have their own subgroups. The following part of the study presents the findings on the functional usage of both 3- and 4-word bundles separately across three distinct proficiency levels.

4.2.1. Functional Categorisation of 3-Word Formulaic Sequences for Level A

As mentioned earlier, the corpus of the student writings of Level A was found to include 38 different types of 3-word formulaic sequences with varying frequencies, but with an occurrence rate of a minimum five times in the research data. A second analysis was carried out to discover the functions they fulfilled in an academic context, which was an EFL writing exam at two distinct universities. The functions of the FSs were identified in light of the classification made by *Biber et al. (2004)*. Accordingly, the functions of the 3-word FSs in the A level data could be examined in the table below under three main categories: *stance expressions, discourse markers, referential expressions* (*Biber et al., 2004*).

Table 9. Functional Categorisation of 3-Word Formulaic Sequences in the Data of Level A

Stance Expressions	Discourse Organisers	Referential Expressions
<i>Epistemic Stance Bundles</i>	<i>Topic Introduction/Focus Bundles</i>	<i>Identification/Focus Bundles</i>
I think this	first of all	is the most
I think that		one of the
I believe that	<i>Topic</i>	it is very
do not know	<i>Elaboration/Clarification Bundles</i>	I do not
		you do not
<i>Attitudinal/Modality Bundles</i>	as a result	he does not
I want to	because it is	we do not
and you can	but it is	we did not
do not like		do not have
have to go		I have a
you have to		I had a
you can see		he is a
you want to		it is a
you can go		is not a
		<i>Bundles Specifying Attributes</i>
		a lot of
		there are many
		<i>Time/Place/Text-deixis Bundles</i>
		in the city
		in the world
		you go to
		go to the
		is in the
		was born in

As can be seen in Table 9, the functional categorisation of the 3-word formulaic sequences used by A-level students presented that most of the identified FSs (22 of them) served referential purposes in the academic writings of the EFL learners. Comprising of 12 FSs, stance bundles were the second-mostly preferred functional type by the A-level EFL learners. And only 4 of the all 3-word FSs were observed to fulfil the discourse organising function in the A-level research corpus.

First, 22 out of 38 sequences were found to be referential expressions, which accounted for more than 50% of all the sequences. The identified referential expressions in the corpus could be examined under three sub-categories, as being identification/focus bundles, bundles specifying attributes, and time/place/text-deixis bundles. The formulaic sequences *is the most*, *it is very* and *one of the* were observed to serve as identification/focus bundles. The FSs *he is a*, *it is a*, *is not a*, *I have a*, *I had a*, *do not have*, *I do not*, *he does not*, *you do not*, *we do not* and *we did not* were also categorised under this sub-category, although the specifications for identification/focus bundles did not completely apply to these sequences. However, it should be noted that they by no means had the functional properties of the other three sub-categories. *A lot of* and *there are many* were the two sequences that specified attributes, which was quantity in both cases. And the rest of the referential expressions; *in the world*, *was born in*, *is in the*, *go to the*, *in the city* and *you go to*, were classified as place-deixis sequences. The following sentence was taken from a student writing, in which the sequence *one of the* was used as an identification/focus bundle: “There are many ways to improve your English. Some people improve it at school. Some people improve it by living in Anglo-Saxon countries. I think **one of the** best ways to improve your English is that.”

Following referential expressions, stance expressions were the second most applied formulaic sequences to convey “epistemic and attitude/modality” (Biber et al., 2004). 12 of the total number of identified 3-word FSs in the data were found to account for stance expressions. While four of them; *I think that*, *do not know*, *I think this* and *I believe that*, functioned as epistemic stance bundles, the other eight; *you can go*, *I want to*, *you have to*, *you want to*, *and you can*, *have to go*, *you can see* and *do not like* served attitudinal/modality functions. “**I believe that** [-] technology [has] changed our world and the [world now] is a better place to live” is one of the samples in the research corpus for the sequences functioning as a stance expression.

In the third and the last category, discourse organisers made up the smallest proportion in the data of language level A constituting 10.52% of all the 3-word sequences. To put it differently, only four of the 38 formulaic sequences functioned as a discourse organiser in the students’ writing exam documents. Of the four discourse organising bundles, three of them; *because it is*, *but it is* and *as a result*, were used to elaborate

on the topic or provide clarification whereas *first of all*, the fourth frequently-used 3-word formulaic sequence, could be listed under the category of topic introduction/focus bundles. Here in this excerpt “In my opinion, technology has made the world a better place to live for many reasons. **First of all**, people can live longer [thanks to] medical technology”, it could be seen that an A-level language learner could successfully use a formulaic sequence to introduce a topic.

4.2.2. Functional Categorisation of 3-Word Formulaic Sequences for Level B

After a multi-step analysis of the research corpus was carried out, 24 different types of 3-word formulaic sequences, which had a cut-off point of at least five times of occurrence, were identified in the level B data. As part of the second objective of the current research, the identified formulaic sequences were examined to find out the functions they fulfilled. The table below groups the 3-word FSs found in the writing exam documents of B-level students in accordance with the categorisation of Biber et al. (2004).

Table 10. Functional Categorisation of 3-Word Formulaic Sequences in the Data of Level B

Stance Expressions	Discourse Organisers	Referential Expressions
<i>Epistemic Stance Bundles</i>	<i>Topic Introduction/Focus Bundles</i>	<i>Identification/Focus Bundles</i>
I believe that	first of all	I do not
I think that	if you are	you do not
do not know		we do not
you want to	<i>Topic</i>	they do not
	<i>Elaboration/Clarification Bundles</i>	it is not
<i>Attitudinal/Modality Bundles</i>		do not have
not want to	as a result	the most important
not have to		the idea that
I want to		one of the
you need to		<i>Bundles Specifying Attributes</i>
		a lot of
		a number of
		<i>Time/Place/Text-deixis Bundles</i>
		in the future
		go to the

As seen in Table 10, the writing exam documents of B-level language learners mostly comprised referential expressions, as in the documents of A-level language learners. They made up 54.16% of the total number of sequences in the research corpus of Level B, which corresponded to 11 out of 24 as a raw number. Stance expressions came after referential expressions with 33.33% (8 in raw number), which were followed by three discourse organisers whose ratio is 12.5%.

First, the referential expressions in the data of Level B were found to function in three different sub-categories as identification/focus bundles, bundles specifying attributes and time/place/text-deixis bundles. Nine of the 3-word formulaic sequences; *the most important*, *the idea that*, *one of the*, *they do not*, *you do not*, *I do not*, *it is not*, *do not have* and *we do not*, were used to serve as identification/focus bundles in student writings. “**The most important** point is that using technological devices at those ages which are 3-9 has not [got] any benefits [for] children” is a fine example that illustrates a language learner’s ability to handle functional use of a formulaic sequence to generate text cohesion. The sequences *a lot of* and *a number of* were seen to be used as quantity indicators, so they could be listed under the category of bundles specifying attributes. The last two referential expressions in the data belonged to the sub-category of time/place/text-deixis bundles. They were the sequences *in the future* and *go to the*, which referred to time and place respectively.

Eight of the 24 3-word sequences in the documents were found to function as stance expressions, and that made 33.33% of the whole group of 3-word formulaic sequences. *I believe that*, *I think that* and *do not know* were used as epistemic stance markers while *not want to*, *not have to*, *you want to*, *I want to* and *you need to* came up as attitudinal/modality bundles. The student excerpt “Secondly, you do **not have to** send only text messages[.] You can also share something such as photos, videos, locations, documents, [and] files” exemplifies the use of a referential expression to denote modality.

Lastly, it could be said that the remaining three 3-word FSs; *first of all*, *as a result* and *if you are*, were used as discourse organisers by the language learners of Level B. Here, *as a result* could be categorised under topic elaboration bundles. On the other hand,

first of all and *if you are* were used as topic introduction/focus bundles. The FS in bold in the following extract “For example, **if you are** selling something online, you can use WhatsApp to take orders and get the money. You can talk about [the] price and send [the] pictures [of the] thing that you sell” includes an FS classified as a topic introduction/focus bundle.

4.2.3. Functional Categorisation of 3-Word Formulaic Sequences for Level C

When the C-level research corpus was examined, it was revealed, as stated before, that 50 distinct types of formulaic sequences were used by the students in their writing exam documents. Those FSs were then analysed to find out about the functions they served in their contexts. The findings were categorised as displayed in Table 11 below.

Table 11. Functional Categorisation of 3-Word Formulaic Sequences in the Data of Level C

Stance Expressions	Discourse Organisers	Referential Expressions
<i>Epistemic Stance</i>	<i>Topic Introduction/Focus</i>	<i>Identification/Focus</i>
do not know	when I was	I was a
did not know		they have a
	<i>Topic Elaboration/Clarification</i>	you do not
<i>Attitudinal/Modality</i>		we do not
not have to	as a result	they do not
not want to	as well as	I did not
want to be	in order to	we have been
be able to	if you do	do not have
are able to		does not matter
have to go		that I was
was going to		and I was
did not want		and I had
I could not		and it is
you have to		it is not
they have to		there is not
I had to		most of the
		one of the
		one of them
		the most important
		the only thing
		is one of
		<i>Imprecision</i>
		he or she

Specifying Attributes

there are some
there is no
a lot of
in terms of
because of the

Time/Place/Text-deixis

I was in
to go to
in front of

As shown in Table 11, the C-level corpus comprised 31 referential expressions with the highest percentage of 62%, 14 stance expressions making up 28% and five discourse organising sequences with the lowest percentage of 10%.

The functional categorisation of the 3-word formulaic sequences in the research data disclosed that more than half of the FSs functioned as referential expressions. Out of 50 FSs, 31 of them fell into this category. What made the C-level corpus distinctive among all the three corpora in terms of referential expressions used was that contrary to the corpora of Level A and Level B, it included one imprecision bundle: “In traditional education, [a] person has to be [at] school and being [at] school requires some rules. For example, **he or she** has to wear [a] uniform and this is not comfortable.” No 3-word imprecision bundle was traced in the previous two corpora. With an aim to denote place, the language learners used three place-deixis bundles which are *I was in*, *in front of* and *to go to*. The sequences *a lot of*, *there are some*, *there is no*, *there are some*, *in terms of* and *because of the* could be categorised as bundles specifying attributes. The rest of the sequences were used as identification/focus bundles. There were 22 identification/focus bundles in the corpus of Level C, and *most of the*, *and it is*, *does not matter*, *the only thing*, *there is not*, *that I was*, and *we have been* were some of the examples to this sub-category of referential expressions.

The FSs signifying the stance of the author constituted the second broadest functional category in the C-level corpus with a percentage of 28. In other words, the students used 14 sequences to serve as a stance expression. The two of stance bundles, *did not know* and *do not know*, were identified to function as epistemic stance markers. Remaining twelve FSs were attitudinal/modality bundles which separately implied ability, desire, obligation, intention, prediction, or directive. *Are able to*, *have to go*, *not want to*, and *was going to* were some of the attitudinal/modality bundles, each being an example of specified six functions. The following cited sentence from a student writing exemplifies a formulaic sequence's being used as an ability marker: "I stayed with them during 3-[day-]long exam. At least, I could pass the exam[,] and now I **am able to** write this essay."

Discourse organisers, on the other hand, did not seem to be benefited much by the C-level language learners. Only five, 10%, of the 3-word sequences were used as discourse organisers. As specified in Biber et al. (2004), discourse organisers comprised two sub-categories as topic introduction/focus bundles and topic elaboration/clarification bundles. As in the student excerpt; "**When I was** 9 years old, my whole world turned upside down. We were living in Mersin back then and I had a great childhood", *when I was* was used to introduce a topic, and it was the only topic introduction/focus bundle in the research data that met the aforementioned criteria (five times of occurrence and having a frequent use in the native corpora). The other four formulaic sequences in the corpus; *as a result*, *as well as*, *in order to* and *if you do*, facilitated topic elaboration, and they could be said to have been accurately used by the C-level language learners.

4.2.4. Functional Categorisation of 4-Word Formulaic Sequences for Level A

The analysis of the A-level research corpus comprising 19,076 tokens revealed that A-level language learners were competent enough to use a number of 3-word formulaic sequences to fulfil various functions in an academic context. However, when it came to 4-word sequences, the number of the FS types dropped drastically, and the functional groupings, consequently, narrowed down, as displayed in Table 12 below.

Table 12. Functional Categorisation of 4-Word Formulaic Sequences in the Data of Level A

Stance Expressions	Discourse Organisers	Referential Expressions
<i>Epistemic Stance Bundles</i> you do not know	-	-

As could be seen in Table 12 above, the research corpus included only one 4-word formulaic sequence, *you do not know*, that met the requirement of at least five times of occurrence. *You do not know* in this context fell into the category of stance expressions according to the classification by Biber et al. (2004). More specifically, it could be described as an epistemic stance expression. The following example was taken from an A-level student writing, which depicts the functional use of the only 4-word formulaic sequence found in the A-level corpus. “When you learn new words, you will begin understanding your listening and reading. So you try to learn words that **you do not know**.” It bears repeating that no 4-word discourse organiser or referential expression was identified in the A-level research corpus.

4.2.5. Functional Categorisation of 4-Word Formulaic Sequences for Level B

The research corpus generated from the writing exam documents of B-level language learners included four type of 4-word bundles. When the functions they fulfilled were examined according to the categorisation proposed by Biber et al. (2004), the findings were as follows.

Table 13. Functional Categorisation of 4-Word Formulaic Sequences in the Data of Level B

Stance Expressions	Discourse Organisers	Referential Expressions
<i>Attitudinal/Modality Stance Bundles</i> do not want to do not have to you do not have		<i>Time/Place/Text-deixis Bundles</i> at the same time

Table 13 displays that the 4-word bundles in the B-level corpus were found to serve two functions in student writings: three of them function as stance expressions and the last one was used as a referential expression. The stance expressions in the data, *do not want to*, *do not have to* and *you do not have*, could be listed under the subcategory of attitudinal/modality stance bundles. In the following example, one of the students whose writing exam document had provided data for the study formed a sentence by using a modality bundle to refer to lack of obligation. “For example, you can text or call someone very fast and **you do not have to** pay for it. [The] only thing you need is [an] internet connection.”

When it came to referential expressions, it could be seen that student writings included only one 4-word formulaic sequence satisfying the parameters of the present study, which was *at the same time*, and it is exemplified in the following student excerpt. “Last of all, we can send documents such as PDF, Word and Excel or voice messages to many people in one touch. So, we can talk or reach many people **at the same time**.” Here it is observed that the referential expression in question functions as a time reference.

4.2.6. Functional Categorisation of 4-Word Formulaic Sequences for Level C

The C-level research corpus was the broadest corpus among all the research corpora, and it was also the richest in the types of both 3-word and 4-word bundles used in the writing documents of the language learners. In terms of 4-word bundles, it was found to include six formulaic sequences, each fitting into one of the three categories proposed by Biber et al. (2004).

Table 14. Functional Categorisation of 4-Word Formulaic Sequences in the Data of Level C

Stance Expressions	Discourse Organisers	Referential Expressions
<i>Attitudinal/Modality Stance Bundles</i>	<i>Topic Introduction/Focus Bundles</i>	<i>Identification/Focus Bundles</i>
do not have to	if you do not	one of the most

*Topic Elaboration/
Clarification Bundles*
on the other hand

*Time/Place/Text-deixis
Bundles*
at the same time

*Multi-functional
Referential Bundles*
at the end of

Table 14 above demonstrates that the EFL learners used 4-word FSs for referential functions the most. Three different referential expressions were used to fulfil this function. The corpus consisted of two discourse-organising FSs, which made it the second most-preferred functional type. And lastly, it was seen that the C-level EFL learners used only one stance expression, so they were the least applied formulaic sequence type in terms of function.

First, a careful examination of the C-level corpus revealed that the language learners used only one stance expression, *do not have to*, as in the example “Besides, contrary to [the] traditional way, you **do not have to** carry the materials that you use for classes”, functioning as a modality stance marker indicating obligation. To this respect, it could be stated that stance expressions made up only a small, 16.66, percentage of all the identified 4-word formulaic sequences. The referential expressions, on the other hand, constituted the highest proportion in the corpus with a percentage of 50%. *At the same time*, *one of the most* and *at the end of* were the three referential expressions identified in the corpus, and each served distinct functions. The sequence *at the same time* performed time reference. The second sequence, *one of the most*, functioned as an identification/focus bundle, and the third sequence *at the end of* was a multi-functional referential bundle, which could indicate time-, place- or text-deixis in varying contexts. The following excerpt from one of the research documents exemplifies the sequence being used as a time-deixis bundle: “The second mistake I made was I always studied the lessons I was good at. **At the end of** the day, I deceived myself thinking I studied a lot.”

Having made up the 33.33% of the total number of 4-word formulaic sequence, the functional category of discourse organisers comprised two formulaic sequences which were *on the other hand* and *if you do not*. The former was found to be used for topic

elaboration, while the latter was a bundle for topic introduction/focus. “He, **on the other hand**, was trying to find a job in Istanbul at the time of economic crisis” is an outstanding illustration of topic elaboration through the use of a formulaic sequence. Moreover, it could be stated that the author of this sample seemed to have enough expertise in the use of formulaic sequences based on the two 4-word FS used in the same sentence.

4.3. Structural Categorisation of the Formulaic Sequences in the Research Data

Structural analysis of 3- and 4-word formulaic sequences used in writing might not predict the relationship between the use of FSs and the proficiency level of an author on its own, but it might facilitate drawing conclusions when interpreted in the light of the findings about the types, frequencies and functions of formulaic sequences in written or spoken language. With this concern, all the 3- and 4-word formulaic sequences identified in the research corpora were structurally analysed under the guidance of the categorisation scheme in Biber et al. (2004). However, in the course of analysis, it was seen that some of the FSs in research corpora did not fit into any of the categories specified in the paper. Öztürk (2021) mentions the same shortcoming in her master’s thesis, and the author states she made use of another taxonomy formed by Douglas Biber and his colleagues in the book the *Longman Grammar of Spoken and Written English* (LSWE). Here in this research, the same approach will be adopted, and the FSs in the documents will be analysed under four main categories: *lexical bundles that incorporate verb phrase fragments, dependent clause fragments or noun phrase and prepositional phrase fragments* (Biber et al. 2004: 381) and *others*. The category *others* will include quantifier expressions, adverbial clause fragments, other adverbial clause fragments, pronoun/noun phrase+(copula) *be*, prepositional phrases with embedded *of*-phrase fragments, other prepositional phrase fragments and other expressions.

4.3.1. Structural Categorisation of 3-Word Formulaic Sequences Identified in the A-Level Corpus

As stated above, the structural analysis of the 3-word formulaic sequences in the A-level corpus was carried out according to the categorisation of Biber et al. (2004). The first examination revealed that there were no samples for some subcategories specified in the paper, and there were six FSs that should be examined under the category of “other” in the light of the categorisation in Biber et al. (1999). Table 15 below summarises the findings as in the simplest way as possible.

Table 15. Structural Categorisation of the 3-Word Formulaic Sequences in the Corpus of Level A

Structural Category	Structural Feature	Level A
Lexical Bundles That Incorporate Verb Phrase Fragments	(connector +) 1 st /2 nd person pronoun + VP fragments	you can go, I want to, I do not, you have to, you do not, you want to, I have a, and you can, we did not, you go to, we do not, I had a, you can see
	(connector +) 3 rd person pronoun + VP fragments	he is a, he does not, it is very, it is a
	Discourse marker + VP fragments	-
	Verb phrase (with non-passive verb)	do not like, is the most, do not know, is in the, go to the, have to go, is not a, do not have
	Verb phrase with passive verb	was born in
	<i>Yes-no</i> question fragments	-
	WH-question fragments	-
	1 st /2 nd person pronoun + dependent clause fragments	I think that, I think this, I believe that
Lexical Bundles That Incorporate Dependent Clause Fragments	WH-clause fragments	-
	<i>If</i> -clause fragments	-
	(Verb/adjective+) <i>to</i> -clause fragments	-
	<i>That</i> -clause fragments	-
	(connector +) Noun phrase with <i>of</i> -phrase fragments	one of the
Lexical Bundles That Incorporate Noun Phrase and Prepositional Phrase Fragments	Noun phrase with other post-modifier fragment	-
	Other noun phrase expressions	-
	Prepositional phrase expressions	in the world, in the city
	Comparative expressions	-

Others	Quantifier expressions	a lot of
	Adverbial clause fragments	first of all
	Other adverbial clause fragments	because it is, but it is
	Other expressions	as a result
	Pronoun/Noun phrase+(copula) <i>be</i>	there are many
	Prepositional phrase with embedded <i>of</i> -phrase fragments	-

As can be seen in Table 15, the identified 3-word FSs in the A-level corpus could be grouped into four categories in terms of their structural features. The corpus mainly included the FSs that incorporated verb phrase (VP) fragments, dependent clause fragments, noun phrase (NP) and prepositional phrase (PP) fragments and the other FSs. The A corpus was the densest in VP fragments in number. The “others” category was found to be the second densest category, and dependent clause fragments and NP and PP fragments were the least common structural type of the corpus.

The 3-word FSs in the corpus were found to fit into four distinct subcategories in the category of verb phrase fragments: 1st/2nd person pronoun + VP fragments, 3rd person pronoun + VP fragments, verb phrase with a non-passive verb and verb phrase with a passive verb.

The first category, 1st/2nd person pronoun + VP fragments, included the highest number of FSs. *You can go, I want to, I do not, you have to, you do not, you want to, I have a, and you can, we did not, you go to, we do not, I had a* and *you can see* were the examples found in the research corpus. 3rd person pronoun + VP fragments, on the other hand, included four FSs which were *he is a, he does not, it is very* and *it is a*. Except for *was born in*, which was a verb phrase with a passive verb, all the other sequences incorporating verb phrase fragments; *do not like, is the most, do not know, is in the, go to the, have to go, is not a* and *do not have*, were followed by a non-passive verb.

The second category introducing sequences that incorporate dependent clause fragments consisted only of 1st/2nd person pronoun + dependent clause fragments, to which *I think that, I think this* and *I believe that* could be given as examples. No FSs exemplifying WH-clause fragments, if-clause fragments, to-clause fragments or that-

clause fragments were identified in the research corpus. In the third category that grouped the formulaic sequences incorporating noun phrase and prepositional phrase fragments, only one noun phrase with of-phrase fragments; *one of the*, and two prepositional phrase expressions; *in the world* and *in the city* were identified. There were no FSs in the research corpus to typify noun phrases with other post-modifier fragment, other noun phrase expressions or comparative expressions. Upon the categorisation of the 3-word FSs so far, it could be said that the FSs *a lot of*, *first of all*, *because it is*, *but it is*, *as a result* and *there are many* did not befit any of the categories mentioned in Biber et al. (2004). However, they could have been examined using the categorisation in Biber et al.'s (1999) *Longman Grammar of Spoken and Written English* (pages 1001-1024). As a result of the second examination of these FSs, five more categories were established. According to this categorisation, *a lot of* was found to be a quantifier expression while *first of all* was categorised as an adverbial clause fragment. *Because it is* and *but it is* were the other two adverbial clause fragments, though they were listed as “other” adverbial clause fragments. *There are many* was structured as pronoun/noun phrase + copula *be*, and the last FS *as a result* was listed under the category of other expressions. In general, the examinations of the 3-word formulaic sequences in the A-level corpus shown that the beginner-level language users seemed to have a tendency to use formulaic sequences that incorporated verb phrase (VP) fragments.

4.3.2. Structural Categorisation of 3-Word Formulaic Sequences Identified in the B-Level Corpus

The number of the structural categories of the 3-word formulaic sequences in the B-level corpus was observed to have increase slightly when compared with the A-level corpus. It could also be said that the categories included relatively more samples. Again, there were some formulaic sequences in this corpus that could not be fitted in any of the categories specified in Biber et al. (2004), and at that point, *LSWE* provided an invaluable guide through the examination of the FSs. Table 16 below summarises the results of the examination.

Table 16. Structural Categorisation of the 3-Word Formulaic Sequences in the Corpus of Level B

Structural Category	Structural Feature	Level B
Lexical Bundles That Incorporate Verb Phrase Fragments	(connector +) 1 st /2 nd person pronoun + VP fragments	I do not, you do not, we do not, you want to, I want to, you need to
	(connector +) 3 rd person pronoun + VP fragments	they do not, it is not
	Discourse marker + VP fragments	-
	Verb phrase (with non-passive verb)	do not have, not want to, *not have to, do not know, go to the
	Verb phrase with passive verb	*not have to
	<i>Yes-no</i> question fragments	-
	WH-question fragments	-
Lexical Bundles That Incorporate Dependent Clause Fragments	1 st /2 nd person pronoun + dependent clause fragments	I believe that, I think that
	WH-clause fragments	-
	<i>If</i> -clause fragments	if you are
	(Verb/adjective+) <i>to</i> -clause fragments	-
	<i>That</i> -clause fragments	-
Lexical Bundles That Incorporate Noun Phrase and Prepositional Phrase Fragments	(connector +) Noun phrase with <i>of</i> -phrase fragments	one of the, a number of
	Noun phrase with other post-modifier fragment	the idea that
	Other noun phrase expressions	-
	Prepositional phrase expressions	in the future
	Comparative expressions	the most important
	Quantifier expressions	a lot of
Others	Adverbial clause fragments	first of all
	Other adverbial clause fragments	-
	Other expressions	as a result
	Pronoun/Noun phrase + (copula) <i>be</i>	-
	Prepositional phrase with embedded <i>of</i> -phrase fragment	-

* the VP fragments that were followed by both a passive and a non-passive verb

Table 16 above shows that there were at least three 3-word FSs exemplifying the four major structural categories in the research corpus. Again, the VP fragments were the most common FS type, followed by NP and PP fragments. The “others” category and the dependent clause fragments included the same number of 3-word FSs, being the least common structural type.

Thirteen of the FSs in the research corpus were discovered to incorporate VP fragments. *I do not, you do not, we do not, you want to, I want to* and *you need to* were the FSs structured as 1st/2nd person pronoun + VP fragment. *They do not* and *it is not* had the very same structure with a change in personal pronoun, which converted into 3rd person pronoun. The corpus also involved six verb phrases followed by either a non-passive or passive verb. *Do not have, not want to, not have to, do not know* and *go to the* preceded a non-passive verb; however, the formulaic sequence *not have to* was noticed to be also followed by a passive verb in the writing exam documents. When it came to the FSs incorporating dependent clause fragments, it was observed that B-level students used *I believe that* and *I think that* in the form of a 1st/2nd person pronoun + dependent clause fragment, and *if you are* as an if-clause fragment. And there were no 3-word bundles in the B-level corpus exemplifying the other three subcategories. Five 3-word FSs were seen to embody noun phrase or prepositional phrase fragments. *One of the* and *a number of* were examples for the FSs in the form of a noun phrase with of-phrase fragment. *The idea that* was in the form of a noun phrase with a post-modifier fragment. There was one prepositional phrase expression in the data, which was *in the future*, and *the most important* was again the only example of its category, being a comparative expression. The last category formed according to Biber et al. (1999) consisted of three formulaic sequences of distinct structures. Accordingly, *a lot of* was classified as a quantifier expression, *first of all* as an adverbial clause fragment and *as a result* as an FS in the category of “other expressions”. To summarise, the FSs in the B-level corpus were found to incorporate VPs most, as in the A-level corpus. Differently, the number of the FSs embodying noun phrase or prepositional phrase fragments was inclined to increase at this language level.

4.3.3. Structural Categorisation of 3-Word Formulaic Sequences Identified in the C-Level Corpus

Having the highest number of tokens and highest number of different types of formulaic sequences, the C-level corpus was also the one which had the highest number of structural categories among the three corpora. The corpus analysis revealed that the FSs could be examined in 15 different structural categories. What was again

notable in the C-level corpus was that both the structural categories and the samples for those categories could be said to increase in number. The results of the examination were displayed in Table 17 below.

Table 17. Structural Categorisation of the 3-Word Formulaic Sequences in the Corpus of Level C

Structural Category	Structural Feature	Level C
Lexical Bundles That Incorporate Verb Phrase Fragments	(connector +) 1 st /2 nd person pronoun + VP fragments	I do not, I did not, you do not, we do not, I could not, you have to, I had to, I was a, and I had, I was in, and I was, we have been
	(connector +) 3 rd person pronoun + VP fragments	and it is, it is not, they do not, they have to, they have a
	Discourse marker + VP fragments	-
	Verb phrase (with non-passive verb)	be able to, do not have, did not know, does not matter, do not know, not have to, not want to, *want to be, have to go, is one of, are able to, *was going to, did not want, to go to
	Verb phrase with passive verb	*want to be, *was going to
	<i>Yes-no</i> question fragments	-
	WH-question fragments	-
Lexical Bundles That Incorporate Dependent Clause Fragments	1 st /2 nd person pronoun + dependent clause fragments	-
	WH-clause fragments	when I was
	<i>If</i> -clause fragments	if you do
	(Verb/adjective+) <i>to</i> -clause fragments	-
	That-clause fragments	that I was
Lexical Bundles That Incorporate Noun Phrase and Prepositional Phrase Fragments	(connector +) Noun phrase with <i>of</i> -phrase fragments	one of the, one of them
	Noun phrase with other post-modifier fragment	-
	Other noun phrase expressions	the only thing
	Prepositional phrase expressions	in front of
	Comparative expressions	most of the, the most important, as well as
Others	Quantifier expressions	a lot of
	Adverbial clause fragments	-
	Other adverbial clause fragments	-
	Other expressions	as a result, he or she, in order to, because of the

Pronoun/Noun phrase + (copula) <i>be</i>	there is no, there is not, there are some
Prepositional phrase with embedded <i>of</i> - phrase fragment	in terms of

* the VP fragments that were followed both by a passive and a non-passive verb

Table 17 above demonstrates that the FSs in C-level corpus varied considerably in terms of their structural features, but still VP fragments made up the majority. VP fragments were followed by others and NP and PP fragments respectively, and FSs incorporating dependent clause fragments were again low in number as in A- and B-level corpora.

First, it could be said that most of the FSs in the C-level corpus were accompanied by verb phrase fragments. *I do not, I did not, you do not, we do not, I could not, you have to, I had to, I was a, and I had, I was in, and I was* and *we have been* were the FSs with a 1st/2nd person pronoun and a VP fragment. *And it is, it is not, they do not, they have to* and *they have a* included a 3rd person pronoun and a VP fragment. The following FSs; *be able to, do not have, did not know, does not matter, do not know, not have to, not want to, want to be, have to go, is one of, are able to, was going to, did not want* and *to go to* were seen to be in the form of a verb phrase followed by a non-passive verb. The two FSs; *want to be* and *was going to*, on the other hand, preceded a passive verb as well. The category of FS-incorporated dependent clause fragments encompassed three formulaic sequences. The first one, *when I was*, was a WH-clause fragment. The second one was an if-clause fragment. And the last formulaic sequence of the category, *that I was*, was a *that*-clause fragment. Seven FSs in the research corpus were found to incorporate either a noun phrase or a prepositional phrase fragment. *One of the* and *one of them* were examples of noun phrases with *of*-phrase fragments. *The only thing* was categorised under the title of “other noun phrase expressions”. The C-level corpus included only one prepositional phrase expression which was *in front of*. Lastly, the FSs *most of the, the most important* and *as well as* were classified as comparative expressions. *A lot of, as a result, he or she, in order to, because of the, in terms of the, there is no, there is not* and *there are some* were the formulaic sequence that could not be specified under any of the categories mentioned in Biber et al. (2004), but they did fit into some other categories according to the

LSWE. A lot of, for instance, was a quantifier expression. The successive four FSs were categorised as “other expressions”. *In terms of* exemplified a prepositional phrase with embedded of-phrase fragment, and the last three FSs in the C-level corpus were in the form of a pronoun/noun phrase with copula *be*. Once again, it was remarkable that like A- and B-level corpus, C-level corpus also included FSs that were combined with VPs the most. Still, it should be acknowledged that the structural variety of the FSs used by the C-level language learners was markedly wide.

4.3.4. Structural Categorisation of 4-Word Formulaic Sequences Identified in the A-, B- and C-Level Corpora

4-word formulaic sequences in the research data were found to be much fewer than the 3-word FSs in the same corpora, which resulted in fewer structural categorisations of the FSs. This being the case, it would be more practical to summarise the findings in one combined table. Table 18 below presents the identified 4-word formulaic sequences in the corpora of Level A, B and C pointing out the structural categories and the structural features of the FSs in the data.

Table 18. Structural Categorisation of the 4-Word Formulaic Sequences in the Corpora of Level A, B and C

Structural Category	Structural Feature	Level A	Level B	Level C
Lexical Bundles That Incorporate Verb Phrase Fragments	(connector +) 1 st /2 nd person pronoun + VP fragments	you do not know	you do not have	-
	(connector +) 3 rd person pronoun + VP fragments	-	-	-
	Discourse marker + VP fragments	-	-	-
	Verb phrase (with non-passive verb)	-	do not want to, do not have to	do not have to
	Verb phrase with passive verb	-	-	-
	<i>Yes-no</i> question fragments	-	-	-
	WH-question fragments	-	-	-
	1 st /2 nd person pronoun + dependent clause fragments	-	-	-

	WH-clause fragments	-	-	-
	<i>If</i> -clause fragments	-	-	if you do not
	(Verb/adjective+) <i>to</i> -clause fragments	-	-	-
	<i>That</i> -clause fragments	-	-	-
Lexical Bundles That Incorporate Noun Phrase and Prepositional Phrase Fragments	(connector +) Noun phrase with <i>of</i> -phrase fragments	-	-	one of the most
	Noun phrase with other post-modifier fragment	-	-	-
	Other noun phrase expressions	-	-	-
	Prepositional phrase expressions	-	at the same time	at the same time, at the end of
	Comparative expressions	-	-	-
Others	Other prepositional phrase fragments	-	-	on the other hand

4.3.5. Structural Categorisation of 4-Word Formulaic Sequences Identified in the A-Level Corpus

When the A-level corpus was scrutinised to identify 4-word formulaic sequences, only one type of FS that meets the requirements established for the current study was discovered. As a result, only one structural category exists for the A-level corpus, as shown in Table 18. *You do not know* is the only FS in the research corpus, and it is structured as a 1st/2nd person pronoun and a VP fragment, meaning that it is an FS that incorporate a verb phrase. There are no examples of a 4-word FS incorporating a dependent clause, noun phrase or prepositional phrase in the corpus of language level A.

4.3.6. Structural Categorisation of 4-Word Formulaic Sequences Identified in the B-Level Corpus

Table 18 reveals that the corpus of B language level included four different types of 4-word formulaic sequences. When examined, they were found to belong to three different structural categories. To begin with, it could be said that the writing

documents of the students did not provide any samples of dependent clause-incorporated formulaic sequences. On the other hand, the sequences used with verb phrase fragments; *you do not have*, *do not want to* and *do not have to*, outnumbered the FSs used with noun phrase and/or prepositional phrase fragments, *at the same time*. *You do not have* was structured as a 1st/2nd person pronoun+ VP fragment. The language learners were found to use *do not want to* and *do not have to* with a non-passive verb in their writing documents. Being the only 4-word FS of its category, *at the same time* exemplified prepositional phrase expressions. There was also no FS in the corpus to be classified under the category of “others”.

4.3.7. Structural Categorisation of 4-Word Formulaic Sequences Identified in the C-Level Corpus

Six types of 4-word formulaic sequences was identified in the C-level corpus, and among those, five of them were observed to be structured differently from each other, which could be seen in Table 18 above. Specific to C-level corpus, each structural category was discovered to include at least one 4-word FS although there were no examples of FSs for every subcategory in those categories. To start with, *do not have to* was an FS which constituted a verb phrase with a non-passive verb. *If you do not* was an if-clause fragment, being the only FS in the category of formulaic sequences that incorporated dependent clause fragments. *One of the most*, *at the same time* and *at the end of* were the FSs used with a noun phrase or prepositional phrase fragment in student writings. And the last 4-word sequence in the research corpus could have been categorised with the aid of Biber et al. (1999) since it fell into the category of “others”. According to the LSWE, *on the other hand* as categorised as other prepositional phrase fragment. On the whole, the C-level corpus included not only the widest variety of 4-word FS types but also the highest number of structural categories compared to A- and B-level corpora.

5. CONCLUSIONS, DISCUSSION AND RECOMMENDATIONS

In this study, it is aimed to describe tertiary level EFL learners' use of 3- and 4-word formulaic sequences in their academic writings and to reveal any patterns in their FS use across proficiency levels. In this section, first a brief summary of the main findings is given. Then, the results are discussed in the light of the related literature to draw some conclusions. Finally, some recommendations are made for practitioners, policymakers and further research.

5.1. Conclusions and Discussion

The question of whether there is a relationship between the language level of language learners and the use of formulaic sequence in academic writing has been under scrutiny of many researchers in the field of language education. The type, frequency and/or range of the formulaic sequences used in written academic texts have been areas of interest along with the structural and functional categorisations of the FSs. The current research was also carried out to examine the use of formulaic sequences in academic writing across three language levels. With this aim, three distinct research corpora, belonging to proficiency levels A, B and C, were created through the collection of EFL learners' writing exam documents. The 3- and 4-word formulaic sequences with an occurrence rate of at least five were identified, checked in two distinct native corpora, and treated as formulaic sequences if their occurrence rate in the native corpora was at least 20 times per million words. After an arduous analysis, all the identified 3- and 4-word FSs were scrutinised in terms of their types, frequencies, functions and structures at each proficiency level.

Regarding the types of 3-word formulaic sequences in the research corpora, the C-level corpus was found to have the largest range of FSs with 50 different types of FSs.

A-level corpus followed the C-level corpus with 38 types, and the B-level corpus was found to have the narrowest range of FSs. To be more precise, the number of the types of 3-word FSs was observed to have decreased when the proficiency level of the EFL learners transitioned from beginner to intermediate. When the proficiency of the EFL learners reached at the advanced level, the number of the FS types greatly expanded in comparison to the transition between the lower levels. Based on this finding, it is not possible to state that an increasing language level resulted in a gradual increase in the number of FS types in academic writing pieces of EFL learners. Similar irregular pattern in level transitions was observed in the study of Leńko-Szymańska (2014). Her study presented that the density of the increase in the variety of the formulaic sequences occurred between the elementary and advanced levels in the texts of Poland, Spain, Israel and Taiwan. That increase took place during the transition from lower to medium language levels for Austrian and Japanese students. However, she highlights that the increase in the range of FS took place in parallel with language proficiency in general, unlike the case in the present study. The study by Goulart (2020) also showed a non-linear increase and/or decrease across differing language levels, although no information about the transition stages was provided. The writings of B1 and B2-level learners were found to include the widest range of formulaic sequences, while the range decreased in the writings of A1-, A2- and C1-level learners respectively. Based on these three studies, including the current study, it could be concluded that increased language level does not always predict a wider range of FS use, or vice versa. There might be decreases and increases in the number of the types of the formulaic sequences in the data across language levels.

Although indecisive, it is possible to make some inferences about the underlying reasons. Considering the findings of this study about the types of 3-word formulaic sequences, the fact that A-level language learners used quite diverse FSs in their writings might have resulted from their limited control over the target language. Since they were not competent enough to generate their own words without assistance, they might have depended much on formulaic sequences during language production. As Wray (2000) points out, formulaic sequences are prefabricated chunks that do not need to be constructed and analysed upon each use, so the use of formulaic sequences in written or spoken language is time- and energy-efficient. Moreover, they are known

to minimise linguistic errors, which could be assumed to be another reason why the beginner-level EFL learners in the study applied variety of FSs in their texts. On the other hand, the B-level EFL learners were found to use the narrowest range of formulaic sequences in their written production. Compared to the beginners, they were considered more knowledgeable about the target language but not competent enough to produce uncontrolled language. Their perception of being “stuck” in between the two proficiency levels might have caused them to become hesitant to depend too much on ready-to-use linguistic units and to produce their texts creatively as advanced-level language learners/users do. The advanced-level language learners in this study were nevertheless found to be the group that used the highest number of FSs in terms of variety, although they were acknowledged to be the most competent language users to produce the most creative language. The intrinsic motivation behind their preference for formulaic sequences might be native-like language use. It is possible that they assumed they could not generate flawless and native-like language unless they used wide variety of FSs since they learnt English as a foreign language. Consequently, they might have wanted to use various prefabricated chunks that they thought the natives frequently use, and they would be able to use them accurately. Or the aim could be to benefit from the time and effort efficiency formulaic sequences provide under time constraints because they were known to have produced their texts that were subjected to examination during their writing exams. The other possible explanation could be that the C-level EFL learners might have had more formulaic sequences in their repertoire that were ready to use, for the C-level learners had been exposed to more linguistic input starting from the lower proficiency levels.

To conclude, the number of the types of 3-word formulaic sequences was higher in the A-level EFL learners’ writings than in the B-level learners’ while the number doubled from B to C language level, making the C-level corpus the richest data set in terms of FS variety. Regarding the probable causes of this variance, it should be noted that the presented arguments are only presumptions that need an exhaustive analysis.

The frequency of 3- and 4-word formulaic sequences was another point of interest in this study, and the findings demonstrated that the frequencies of the identified 3-word FSs in the corpus of Level A were much higher than the corpora of Level B and C. In

other words, the A-level EFL learners had a tendency to use FSs in their repertoire much more repeatedly than the other two levels did. When the remaining two, B- and C-level corpora were compared in terms of FS frequency, it was seen that the frequency rate of the 3-word FSs in the B-level corpus was slightly more than the C-level corpus. Considering the results of frequency analysis, it could be concluded that the language users with higher levels of proficiency were more capable of avoiding repetition. A similar result was obtained in the studies of Appel and Wood (2016) and Juknevičienė (2009). Their studies illustrated that lower-level language learners used the same formulaic sequences repeatedly, which consequently increased the frequency ratio of the FS. Though they did not focus directly on language levels but on native and non-native language use; non-natives being presumably lower-level language users, the studies by Bychkovska and Lee (2017), Durrant and Schmitt (2009), Jaworska et al. (2015), Lu and Deng (2019), Lyu and Gee (2020), Öztürk and Köse (2016) and Uçar (2017) also displayed a higher frequency of FSs in the productions of non-native language users. The decrease in the frequency of FSs as language level increased might have resulted from the expanding lexicon, developing grammatical competence and improving linguistic skills of the higher-level language learners. As pointed out before, since lower-level language learners did not have complete control over the target language, they might have preferred to stay in the “safe zone” by repeatedly using the FSs in their repertoire. Or that might only be a constraint caused by linguistic incompetence rather than preference. On the other hand, the higher-level EFL learners were observed during the analysis to have expressed themselves in various ways instead of sticking too much to certain ways of expression. The last finding regarding the frequencies of the 3-word formulaic sequences in the research corpora was particularly remarkable in respect of the frequencies of the oft-used FSs in the native corpus, the BNC. It was surprising to see that although the B-level corpus included the narrowest range of 3-word FSs, more than one-third of them, i.e., *you do not*, *not want to*, *one of the*, and *do not know*, (See Table 2) happened to occur in the BNC at quite high ratios, and this ratio decreased in the corpus of Level C and became the lowest in the corpus of Level A. To put it another way, B-level EFL learners seemed to use the FSs that the native speakers of English also prefer to use quite often, which might indicate that they produced a more native-like language. In

Juknevičienė's study (2009), on the other hand, out of two proficiency levels, the productions of higher-level language learners showed more resemblance to the native corpora. Still, it is beneficial to be cautious while making inferences because no analysis was conducted for the accuracy of the use of the FSs in the current study.

As the last part of the type and frequency analysis, shared 3-word formulaic sequences were identified in the research corpora and eight FSs; *a lot of*, *I do not*, *you do not*, *one of the*, *do not know*, *we do not*, *do not have* and *as a result*, were found to be used by all the three proficiency levels with varying frequencies. It was observed that the frequencies did not follow any pattern between or across proficiency levels. In other words, the frequency of the shared 3-word FSs were not informative about the general tendencies of the EFL learners of different language levels. What was remarkable though was that *a lot of* was the most frequently-used FS among the shared ones in A- and C-level corpora and the third most recurring FS in the B-level corpus. Öztürk (2020) similarly concluded that *a lot of* was found to be one of the most frequently-used formulaic sequences among the shared FSs in each sub-corpus of her study. Elkılıç (2012) and Şen and Şimşek (2020) stated that inaccurate use of quantifiers was one of the most common reasons that led to language errors in Turkish EFL learner writings, which means that quantifiers are one of the most common linguistic units in Turkish learners' writing. And because *a lot of* is an unmarked quantifier in terms of its usability with both countable and uncountable nouns as in the Turkish language, it might have been preferred by great amount of Turkish EFL learners at all levels. And if this is the case, its high frequency in the research corpora at all levels is predictable.

The types and frequencies of the 3-word formulaic sequences in the research data showed nonlinear distribution across levels, making it difficult to draw obvious conclusions about the relationship between the FS use and proficiency level. The analyses of the 4-word formulaic sequences in the research data, however, showed apparently more explicit patterns. Firstly, a considerable decrease in both types and frequencies of formulaic sequences was observed when the focus of the analysis was shifted to 4-word sequences in the data. Hejazi (2020) also pointed out that both the intermediate and advanced-level groups in his study used fewer 4-word FSs than 3-word FSs. Secondly, the analyses showed that the types and tokens of 4-word FSs in

the research data could inform well of the language levels of the EFL learners in that as the language level of the learners increased, 4-word FSs were observed to have diversified and become more frequent in the data. The finding is inconsistent with Hejazi (2020) which reported that it was the lowest level that used the 4-word FSs the most frequently. The fact that 3-word formulaic sequences outnumbered 4-word sequences in the research corpora and that the written productions of higher-level language included more 4-word FSs compared to the lower levels was most likely to be result of the developing linguistic competence of EFL learners. 4-word FSs are seemingly more complicated and sophisticated in nature than 3-word FSs, and as a result, lower-level language users might not have had the means for their production. Biber et al.'s (1999) conclusion that 3-word FSs could be regarded as "a kind of extended collocational association, and they are thus extremely common" while the formulaic sequences that are greater in length (4- or more word FSs) are "more phrasal in nature and correspondingly less common" (p. 992) seems to support this assumption. Additionally, the processing efficiency and retrieval rate that formulaic sequences provide quite possibly decreased in longer FSs, and that consequently led to reduced occurrence in the lower-level texts.

Considering the findings of the current research about the types and frequencies, it could be summarised that the distribution of 3-word formulaic sequences in the research data and the frequency of the examined 3-word FSs did not demonstrate linearity across levels. The analysis of the 4-word FSs, on the other hand, demonstrated a linear increase in the use and also frequency of formulaic sequences from the lowest to the highest level. Related literature does embody studies concluding that the written texts of the language users of higher language levels were richer in variety in terms of use of FSs (i.e., Alhassan & Wood, 2015; Ädel & Erman, 2012; Huang, 2015; Karabacak & Qin, 2013; Li & Volkov, 2017; Ruan, 2016; Qin, 2014). On the contrary, there are also instances of a narrower range of FS use in the advancing levels, examples of which could be seen in Ahmadi et al. (2013), Appel and Wood (2016), Bychkovska and Lee (2017), Hejazi (2020), Jaworska et al. (2015), Lu and Deng (2019), Lyu and Gee (2020), Öztürk and Köse (2016), and Pan et al. (2016), Staples et al. (2013) and Wei and Lei (2011).

The second motivation for the present study was to examine the functional qualities of both 3- and 4-word formulaic sequences in EFL learners' academic texts across three proficiency levels. The functional analysis of the study was grounded on Biber et al.'s (2004) taxonomy of functional categorisation, and a possible relationship between the proficiency level and the dispositions of the EFL learners towards the functional roles of the FSs was examined.

To start with the functions of the 3-word formulaic sequences in the corpora, the academic writings of the A-level EFL learners were found to include referential expressions the most, stance expressions following the referential expressions and the discourse organising sequences being the least preferred sequences. While the ratio of the referential sequences to the total number of FSs was 60.5%, the ratios of stance expressions and discourse organising sequences followed it at 28,94% and 10,5% respectively. In the B-level corpus, referential expressions constituted the 54% of all the 3-word FSs in the data. Stance expressions followed them with a ratio of 33%, and the discourse organising sequences made up only the 12.5% of the total number. Again, in the corpus of Level C, referential expressions were the most frequent 3-word FSs with a ratio of 62%. Stance expressions had the second highest ratio with 28%, and discourse organising sequences comprised the 10% of all the 3-word FSs in the corpus.

The findings clearly showed that referential expressions were the mostly-used 3-word sequences in each corpus, contrary to the findings of Juknevičienė (2009), Li and Volkov (2017) and Staples et al. (2013). The referential bundles in those studies were found to be the least-preferred formulaic sequences at each language level, and in Appel and Wood (2016) and Nam (2017), they constituted more of the writings of the higher-level language users. One of the findings of the study by Uçar (2017) and the present study bears a surprising similarity in that both sets of data disclosed a strong tendency of Turkish speakers of English for the use of referential bundles in academic writing. Whether this tendency is representative of all or at least most of the Turkish-speaking English learners could be an intriguing research area for further studies.

After referential bundles, stance marking formulaic sequences were found to be the second most-common functional type within the same language level and among all the three language levels, with a ratio of over 25% in each. They are actually the most-commonly encountered formulaic sequences in speech, according to Biber et al. (2004). And seemingly, they are not limited to spoken language because there are numerous studies in literature having revealed the very frequent use of stance expressions especially among lower-level language users and non-native speakers of the target language in written language as well. (Appel and Wood, 2016; Juknevičienė, 2009; Nam, 2007; Staples et al., 2013). Lu and Deng (2019), however, drew an unpredictable conclusion. They presented that stance expressions were the least recurring formulaic sequence in the corpus of non-native and native speakers of English, but still, the non-native speakers were still found to use stance expressions more often than the native speakers did. Another study that put forth a distinguished finding belongs to Jaworska et al. (2015). The researchers highlighted that the native speaker group in their study applied personal stance bundles quite frequently, contrary to what was discovered in other studies in the literature about the nature of native language.

Lastly, it was observed that discourse organising sequences had the smallest proportion of all the FSs, as detailed in one of the preceding paragraphs, again within and among the three sets of corpora. Considering the vast and major functions they fulfil, as Appel and Wood (2016) touched upon, their low frequency of use could be anticipated. The frequency of discourse organising bundles was indeed expected to increase as the EFL learners became more competent and insightful in the target language, yet they were discovered to be more prevalent in the B-level corpus than the C-level corpus when the categorised FSs were proportioned to the total number of FSs in the corpus. Still, it is possible to say that there was no significant divergence in the use of discourse organising bundles across proficiency levels.

The analysis of the 4-word formulaic sequences in the corpora produced relatively more predictable results compared to the analysis of 3-word FSs. The most apparent change was observed in the use of referential expressions. Their frequency in the EFL learners' writings linearly increased as the learners' language level increased. A

similar pattern applied to discourse organising formulaic sequences. That is to say, the proficiency level of the language learners could predict the frequency of use of the discourse organising FSs. While the C-level corpus included two distinct discourse organising FSs, they were not encountered in both of the lower-level corpora. Only the stance expressions could be said to have demonstrated inconsistent existence in the data. The EFL learners of A and C levels used one stance expression whereas two stance expressions were identified in the writings of B-level EFL learners. This finding was not anticipated because most of the studies in the field, as summarised in previous paragraphs about 3-word formulaic sequences, revealed a high frequency of stance expressions either in the lowest-level language learners' or in the non-native language learners' writings. The finding at issue is insufficient to draw reasonable conclusions as to this inconsistency because of the inadequacy of the identified 4-word FSs in the three research corpora. Replicating the study with broader corpora might lead to some reasonable and valid conclusions.

When the results of analyses of both 3- and 4-word formulaic sequences in the research corpora are examined together, referential expressions and stance expressions appeared to be the most recurring formulaic sequence in the corpora while the occurrence of discourse organising 3- and 4- FSs was pretty low at each proficiency level. The frequency of stance and referential expressions might be explained by the fact that most of the EFL learners' writings that formed the research corpora were either argumentative or descriptive essays. Considering that the language learners were non-native speakers of English, it could be anticipated that personal tone would be at work while producing argumentative essays, at least in the writings of lower-level learners. Or since descriptive essays require giving references to "concepts, elements, or tangible/intangible items" (Appel & Wood, 2016: 63), the learners instinctively made use of referential bundles performing these functions. Lastly, stance and referential expressions are not as complex in nature as discourse organising sequences with regard to their functions in academic contexts, and that might have made them favourable for lower-level or non-native learners of a language in the course of production.

The third and last objective of the current study was to investigate the probable effect of the proficiency level on the structure of formulaic sequences used by EFL students in written language; therefore 3-word and 4-word FSs were structurally categorised according to the specifications defined in Biber et al. (1999, 2004). The examination of the findings revealed that the 3-word formulaic sequences incorporating verb phrase (VPs) fragments were by far the most occurring structural type at all language levels in the research data. Their ratio in each corpus became lower first from A to B level and then higher from B to C level, being the highest in the A-level corpus. In most of the studies in the literature (i.e., Bychkovska & Lee, 2017; Goulart, 2020; Juknevičienė, 2009; Lu & Deng, 2019; Lyu & Gee, 2020; Pan et al., 2016), VP fragments were also discovered to be as widespread; however, the occurrence of this type of FS structure was observed to decrease as the proficiency level of the participants increased. In the same way, the written productions of non-native language learners were observed to consist of fewer VPs in comparison with the native speaker texts. In this respect, the first research finding could be said not to be consistent with the studies in the literature. An instance of greater use of VP fragments at C1 language level than at the B2 level was encountered in Hejazi (2020). The second finding of the structural analysis of the 3-word FSs disclosed that FSs incorporating dependent clause fragments were the least common at all levels, and the B-level corpus was found to consist of more dependent clause fragments than the other two corpora with a ratio of 12.5%. The finding is consistent with the results of many studies, i.e., Bychkovska & Lee (2017), Lu & Deng (2019), Lyu & Gee (2020) and Uçar, (2017), except for Juknevičienė (2009). In her study, AFK1 corpus belonging to the lowest-language level was found to include dependent clauses quite frequently. When it came to the 3-word FSs that incorporated noun phrase (NP) and prepositional phrase (PP) fragments, it could be said that their frequency of use was much more related to the proficiency level of the EFL learners in that their ratio became greater at higher language levels as in Bychkovska and Lee (2017) and Lyu and Gee (2020), contrary to Goulart (2020) and Uçar (2017). Nonetheless, it should be noted that the B-level corpus of the present study consisted of more NPs and PPs (20.83%) than the C-level corpus (14%). The sub-types of the “others” category, on the other hand, varied among the three proficiency levels, but quantifier expressions, adverbial clause fragments,

and other expressions were the shared FS type among the research corpora. And pronoun/noun phrase + copula *be* was the type of 3-word FS occurred only in the A- and B-level corpus.

As to the possible explanations for the distribution of the structural categories, the classroom language, syllabus/curriculum design and/or course materials might be considered to have played a major role if the studies in the field could put forth varying patterns across language levels or between native and non-native languages, making it difficult to make conspicuous deductions. Only the frequency of 3-word VPs the proficiency level *seemed* to be inversely proportional according to the results of the studies in the field (Bychkovska & Lee, 2017; Goulart, 2020; Lu & Deng, 2019; Pan et al., 2016; Lyu & Gee, 2020). Considering this and the infrequency of 3-word dependent clause fragments, it might be induced that the EFL learners whose writing exam documents were examined could not reach a native or near native-like level for some reasons. However, the distribution of the 3-word NP and PP fragments showed the opposite. To conclude, it could be stated that the structural distribution of the 3-word FSs across proficiency levels alone was not a determining factor in predicting language learners' proficiency levels. Surprisingly though, the analysis of the structures of 4-word formulaic sequences was more informative about the EFL learners' proficiency levels. It was found that as the EFL learners became more competent in the target language, the use of 4-word VP fragments decreased, except for Level B, and the use of 4-word NP, PP, dependent clause fragments and other structural types increased. Developing language skills seemed to have diversified the types, functions and structures of the 4-word sequences used by the EFL learners.

5.2. Recommendations

This current research which was carried out to examine the use of formulaic sequences in academic writing across three language levels has brought about significant conclusions. Below are made some practical recommendations for practitioners, policymakers and further research separately regarding the conclusions and discussions of the study.

5.2.1. Recommendations for Practitioners and Policymakers

The following suggestions can be made regarding the results of this current study. First, considering their prevalent and extensive use in native and non-native texts and the advantages they provide the learners with in both written and spoken production EFL teachers' awareness of the role and advantages of FSs should be increased via in-service training. Second, EFL learners should also be instructed and trained on how and why to include formulaic language in their use, when needed. Third, it is of no use to say that educational personnel at all levels should always be up-to-date about the new developments, discoveries, and trends in education. Likewise, the policymakers and curriculum developers should consider the studies on FS use since classroom language could be adjusted, course materials could be reviewed and adapted, and language teaching curricula could be redesigned in the light of existing research in the related literature. Lastly, EFL teachers should be encouraged and trained to conduct research because the FSs which are frequently used in Turkish EFL learners' productions but quite rare in native corpora are recommended to be subjected to elaborate examination to discover if the language taught in language classes diverges markedly from the native language.

5.2.2. Recommendations for Further Research

Based on the overall research findings, it could firstly be stated that the study of 4-word formulaic sequences is more likely to produce representative results of the groups investigated; hence it seems more practical to conduct research on 4- or more word FSs. Secondly, carrying out longitudinal research with a relatively limited number of participants/documents could produce more comprehensive and reliable results in understanding the nature of proficiency level-related FS use. For researchers, the language teaching programs that apply a modular system would be a favourable research environment for this. When a semester system has to be chosen for the research for any reason, it is advisable to generate much wider research corpora to get representative results. Thirdly, experimental research with control groups could be

conducted to scrutinize the possible relationship between FS use and proficiency level or the effect of instruction (e.g., use of a particular methodology or instruction by native versus non-native teachers) on FS use. And lastly, a study with a mixed method design is likely to provide much more generative results.

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>
- Ahmadi, H. S., Ghonsooly, B., & Fatemi, A. H. (2013). An analysis of lexical bundles in research article abstracts by Iranian and Native English-speaking authors of applied linguistic articles. *The Asian ESP Journal* 9(1), 5-25.
- AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes*, 17, 51-62. <https://doi.org/10.1016/j.jeap.2015.02.001>
- Al-Oliemat, A. (2013). Attitudes and motivations for learning Italian as a foreign language in Jordan. *International Journal of Linguistics*, 5(6), 154-170. <https://doi.org/10.5296/ijl.v5i6.4982>
- Appel, R., & Wood, D. (2016). Recurrent word combinations in EAP test-taker writing: Differences between high- and low-proficiency levels. *Language Assessment Quarterly*, 13(1), 55-71. <https://doi.org/10.1080/15434303.2015.1126718>
- Aydın, S. (2008). An investigation on the language anxiety and fear of negative evaluation among Turkish EFL learners. *Asian EFL Journal*, 421-444.
- Badstübner, T., & Ecke, P. (2009). Student expectations, motivations, target language use, and perceived learning progress in a summer study abroad program in

- Germany. *Die Unterrichtspraxis/Teaching German*, 42(1), 41-49. <https://doi.org/10.1111/j.1756-1221.2009.00034.x>
- Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based approaches to issues in applied linguistics. *Applied Linguistics*, 15(2), 169-189. <https://doi.org/10.1093/applin/15.2.169>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Pearson Education Limited. <https://doi.org/10.1162/089120101300346831>
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), 275-311. <https://doi.org/10.1075/ijcl.14.3.08bib>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371-405. <https://doi.org/10.1093/applin/25.3.371>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263-286. <https://doi.org/10.1016/j.esp.2006.08.003>
- Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). #LancsBox v. 5.x. [software]. Available at: <http://corpora.lancs.ac.uk/lancsbox>.
- Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, 30, 38-52. <https://doi.org/10.1016/j.jeap.2017.10.008>
- Casanave, C. P. (2002). Writing games: multicultural case studies of academic literacy practices in higher education. Routledge. <https://doi.org/10.4324/9781410612366>

- Chen, Y-H., & Baker, Paul. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.
- Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing. Generating text in L1 and L2. *Written Communication*, 18(1), 80-98. <https://doi.org/10.1177/0741088301018001004>
- 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>
- Conrad, S., & Biber, D. (2005). The Frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*, 20, 56-71. <https://doi.org/10.1515/9783484604674.56>
- Coşkun-Demirpolat, B. (2015). *Türkiye'nin yabancı dil öğretimiyle imtihanı: Sorunlar ve çözüm önerileri*. Turkuaz Matbaacılık Yayıncılık A.Ş.
- Council of Europe Portal (n.d.). *Reference Level Descriptions (Language by language)*. Retrieved July 25, 2022 from <https://www.coe.int/en/web/common-european-framework-reference-languages/reference-level-descriptions>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing* 16(3), 129-147. <https://doi.org/10.1016/j.jslw.2007.07.002>
- Creswell, J. W. (2009). *Research Design. Qualitative, quantitative, and mixed methods approach*. SAGE Publications, Inc.
- Crossley, S., & Salsbury, T. L. (2011). The development of lexical bundle accuracy and production in English second language speakers. *International Review of*

- Applied Linguistics in Language Teaching*, 49(1), 1-26. <https://doi.org/10.1515/iral.2011.001>
- Dağtan, E., & Cabaroğlu, N. (2021) Status of English speaking skills in Turkish ELT departments: A nationwide survey. *Eurasian Journal of Applied Linguistics*, 7(1), 359-382. <https://doi.org/10.32601/ejal.911454>
- Davis, W. S. (2020). Encouraging continued university foreign language study: a self-determination theory perspective on programme growth. *The Language Learning Journal*. <https://doi.org/10.1080/09571736.2020.1740768>
- Duan, S., & Shi, Z. (2021). A longitudinal study of formulaic sequence use in second language writing: Complex dynamic systems perspective. *Language Teaching Research*, 1-34. <https://doi.org/10.1177/136216882110029>
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL – International Review of Applied Linguistics in Language Teaching*, 47(2), 157-177. <https://doi.org/10.1515/iral.2009.007>
- Elkılıç, G. (2012). Mother tongue traces of Turkish university students on composition papers written in English. *Procedia – Social and Behavioral Sciences*, 47, 656-664. <https://doi.org/10.1016/j.sbspro.2012.06.713>
- Fang, Z. (2021, June 15). *What is academic writing? (and other burning questions about it)*. Routledge. Retrieved July 10, 2022, from <https://www.routledge.com/blog/article/what-is-academic-writing-and-other-burning-questions-about-it>
- Gao, F. (2010). Learning Korean language in China: motivations and strategies of non-Koreans. *International Journal of Bilingual Education and Bilingualism*, 13(3), 273-284. <https://doi.org/10.1080/13670050903006929>
- Goulart, L. (2020). Formulaic sequences and writing development in Portuguese as a second language. *Spanish and Portuguese Review*, 6, 79-107.

- Hejazi, H. (2020, June). *Investigating the use of lexical bundles and keyness in B2 and C1 ESL learners' academic writing*. Internet and Modern Society (IMS-2020), St. Petersburg.
- Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System*, 53, 13-23. <https://doi.org/10.1016/j.system.2015.06.011>
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4-21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Iwai, Y. (2011). The effects of metacognitive reading strategies: Pedagogical implications for EFL/ESL teachers. *The Reading Matrix*, 11(2), 150-159.
- Jaworska, S., Krummes, C., & Ensslin, A. (2015). Formulaic sequences in native and non-native argumentative writing in German. *International Journal of Corpus Linguistics*, 20(4), 500-525. <https://doi.org/10.1075/ijcl.20.4.04jaw>
- Jiang, N., & Nekrasova, T. M. (2007) The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91, 433-445. <https://doi.org/10.1111/j.1540-4781.2007.00589.x>
- Jin, L., Liang, X., Jiang, C., Zhang, J., Yuan, Y., & Xie, Q. (2014). Studying the motivations of Chinese young EFL learners through metaphor analysis. *ELT Journal*, 68(3), 286-298. <https://doi.org/10.1093/elt/ccu011>
- Juknevičienė, R. (2009). Lexical bundles in learner language: Lithuanian learners vs. native speakers. *Kalbotyra*, 60(3), 61-72.
- Karabacak, E., & Qin, J. (2013). Comparison of Lexical Bundles used by Turkish, Chinese, and American University Students. *Procedia - Social and Behavioral Sciences*, 70, 622-628. <https://doi.org/10.1016/j.sbspro.2013.01.101>

- Kim, E-S., & Lee, E-J. (2020). A corpus-based study of lexical bundles and moves by English L1 and L2 writers in medical journal abstracts. *Korean Journal of English Language and Linguistics*, 20, 768-800. <https://doi.org/10.18823/asiatefl.2021.18.1.9.142>
- Kroll, B. (1994). *Second Language Writing. Research insights for the classroom*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524551>
- Lenko-Szymanska, A. (2014). The acquisition of formulaic language by EFL learners: A cross-sectional and cross-linguistic perspective. *International Journal of Corpus Linguistics*, 19(2), 225-251. <https://doi.org/10.1075/ijcl.19.2.04len>
- Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18(2), 85-102. <https://doi.org/10.1016/j.jslw.2009.02.001>
- Li, Z., & Volkov, A. (2017). “To whom it may concern”: A study on the use of lexical bundles in email writing tasks in an English proficiency test. *TESL Canada Journal*, 34(3), 54-75. <https://doi.org/10.18806/tesl.v34i3.1273>
- Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes*, 39, 21-36. <https://doi.org/10.1016/j.jeap.2019.03.008>
- Lyu, M., & Gee, R. W. (2020). Lexical bundles in thesis abstracts by L1 Chinese learners of English and U.S. students. *English Language Teaching*, 13(1), 141-155. <https://doi.org/10.5539/elt.v13n1p141>
- Matsuda, P. K., & Silva, T. (2002). Writing. In *An introduction to applied linguistics*, Eds. N. Schmitt & M. P. H. Rodgers, Routledge, p. 279-293.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage Publications, Inc.

- Nam, D. (2017). Functional distribution of lexical bundle in native and non-native students' argumentative writing. *The Journal of Asia TEFL*, 14(4), 703-716. <https://doi.org/10.18823/asiatefl.2017.14.4.8.703>
- Noprival, Rafli, Z., & Nuruddin. (2021). The motivations of learning foreign languages: A descriptive case study of polyglots. *The Qualitative Report* 2021, 26(12), 3823-3833. <https://doi.org/10.46743/2160-3715/2021.4619>
- Nunan, D. (1989). *Understanding language classroom*. Prentice Hall.
- Özmat, D., & Senemoğlu, N. (2021). Difficulties in learning English by EFL students in Turkey. *Ankara University of Faculty of Educational Sciences*, 54(1), 141-173. <https://doi.org/10.30964/auebfd.742803>
- Öztürk, H. (2021, June). *A longitudinal learner corpus investigation of the use and development of formulaic sequences in argumentative essays of EFL learners* (Master's Thesis). Karadeniz Technical University.
- Öztürk, Y., & Köse, G. D. (2016). Turkish and native English academic writers' use of lexical bundles. *Journal of Language and Linguistic Studies*, 12(1), 149-165.
- Paltridge, B. (2004) Academic writing. *Language Teaching*, 37, 87-105. <https://doi.org/10.1017/S0261444804002216>
- Pan, F., & Liu, C. (2019). Comparing L1-L2 differences in lexical bundles in student and expert writing. *Southern African Linguistics and Applied Language Studies*, 37(2), 142-157. <https://doi.org/10.2989/16073614.2019.1625276>
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71. <https://doi.org/10.1016/j.jeap.2015.11.003>
- Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In *Phraseology in foreign language learning and teaching*, Eds. Fanny M. &

- Sylviane G., Amsterdam: Benjamins Publishing, p. 101-119. <https://doi.org/10.1075/z.138.09paq>
- Patton, M. Q., (1985). *Quality in qualitative research: methodological principles and recent developments*. Invited address to Division J of the American Educational Research Association, Chicago, IL.
- Peaci, D. (2016). The problematic interaction between the mother tongues, the national language and foreign language instruction in Turkish education. *Procedia – Social and Behavioral Sciences*, 232, 479-485. <https://doi.org/10.1016/j.sbspro.2016.10.065>
- Peréz-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84-94. <https://doi.org/10.1016/j.jeap.2014.01.002>
- Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System*, 42, 220-231. <https://doi.org/10.1016/j.system.2013.12.003>
- Ruan, Z. (2016). Lexical Bundles in Chinese Undergraduate Academic Writing at an English Medium University. *RELC Journal*, 48(3), 327-340. <https://doi.org/10.1177/0033688216631218>
- Shin, Y. K., & Kim Y. (2017). Using lexical bundles to teach articles to L2 English learners of different proficiencies. *System*, 69, 79-91. <https://doi.org/10.1016/j.system.2017.08.002>
- Sinclair, J. McH. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sinclair, J.McH. & Renouf, A. (1988). A lexical syllabus for language learning. In *Vocabulary and language teaching*. Eds. R. Carter & M. McCarthy, Longman, p. 140-160.

- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12(3), 214-225. <https://doi.org/10.1016/j.jeap.2013.05.002>
- Suna, Y., & Durmuşçelebi, M. (2013). Türkiye'de yabancı dil öğrenme-öğretme problemine ilişkin yapılan çalışmaların derlenmesi [A compilation work about why Turkey suffers from learning and teaching English]. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, 3(5), 7-24.
- Şen, N. & Şimşek, A. (2020). An analysis of Turkish students' written errors: A case of an EFL context. *Journal of Language Research (JLR)*, 4(1), 58-68.
- Tahaine, Y. & Daana, H. (2013). Jordanian undergraduates' motivations and attitudes towards learning English in EFL context. *International Review of Social Sciences and Humanities*, 4(2), 159-180.
- Türker, M. S. & Çelik, H. (2022). Investigation of Syrian refugees' motivations for learning Turkish. *International Journal of Education & Literacy Studies*, 10(1), 65-74. <https://doi.org/10.7575/aiac.ijels.v.10n.1p.65>
- Ucar, 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. <http://doi.org/10.5539/elt.v10n12p28>
- Usó-Juan, E., & Matínez-Flor, A. (2006). *Current trends in the development and teaching of the four language skills*. De Gruyter Mouton. <https://doi.org/10.1515/9783110197778>
- Wei, Y. & Lei, L. (2011). Lexical bundles in the academic Writing of advanced Chinese EFL learners. *RELIC Journal*, 42(2), 155-166. <https://doi.org/10.1177/0033688211407295>

- Wood, D. (2015). *Fundamentals of formulaic language. An introduction*. Bloomsbury Publishing Plc.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 21(4), 463-489. <https://doi.org/10.1093/applin/21.4.463>
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511519772>
- Zarrabi, F. (2018). English in an EFL context: Teachers' and learners' motivations for English Language Learning. *English Language Teaching*, 11(9), 17-25. <https://doi.org/10.5539/elt.v11n9p17>
- Zheng, Y. (2015). The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English. *System*, 56, 40-53. <https://doi.org/10.1016/j.system.2015.11.007>

APPENDICES

Appendix 1. Ethics Committee Approval (Etik Kurul Onay Formu)

CURRICULUM VITAE

PERSONAL INFORMATION

Name – Last Name Öyküm Tuğçe ÇETİN
Nationality Turkish

EDUCATION

2019 - 2022 **İstanbul Medeniyet University, İstanbul, Turkey**
English Language Education, M.A.

2010 - 2016 **Boğaziçi University, İstanbul, Turkey**
Foreign Language Education Department

2014 Apr. - June **Pädagogische Hochschule Ludwigsburg, Germany**
Foreign Language Education
Courses taken: How to Teach Vocabulary & How to Teach Writing

2006 - 2010 **Remzi Yapıcı Teacher Training Anatolian High School,**
Kırklareli, Turkey

WORK EXPERIENCE

Oct 2018 - Nov. 2020 **Beykoz University, İstanbul**
Instructor & Member of the Testing Office

June 2016 - Oct 2018 **YDS Publishing, İstanbul**
Content Writer & Editor

Nov 2015 - June 2016 **Beşiktaş Sınav High School, İstanbul**
English Teacher

July 2015 - Nov 2015 **Boğaziçi University Lifelong Learning Centre & ÜniversitePlus, Online English School**
Mentor & English Teacher; Speaking Instructor

Sep 2014 - June 2015 **Şişli Terakki High School, İstanbul**
Internship as an Assistant English Teacher

Sep 2014 – July 2015	Prestige English Language School, İstanbul English Teacher
April – July 2014 Germany	Wilhelm-Keil-Schule, Remseck am Neckar, Assistant English Teacher
July 2012 – July 2013	Geleceğin Yıldızları Summer Camps Assistant English Teacher & Coach

PROJECTS

April – July 2014	Wilhelm-Keil-Schule, Remseck am Neckar, Germany Researcher
--------------------------	--

The Title and Subject of the Project:

“Migration and Intercultural Competence As Pedagogical and Methodological Responsibilities” Along with my German-Turkish project partner, I carried out research for three months on the effects of migration background on social relationships of students in Germany and the effects of those relationships on the academic success of the students. The findings of the research were presented by me and my partner in a conference at Boğaziçi University in 2015, and the research was published in the book titled *Bildung Dialog Kultur* in 2016.

RESEARCH

- Sanli, C. & Cetin, O. T (2016). Freundschaften von Schülerinnen und Schülern mit Migrationshintergrund. In Aicher-Jakob, M. & Marti, L., *Bildung Dialog Kultur* (pp. 39-48). Baltmannsweiler, Germany: Schneider Verlag Hohengehren.

SKILLS

Computer	Microsoft Office
Languages	Turkish - Mother Tongue English - Advanced German - Intermediate