

KARADENİZ TECHNICAL UNIVERSITY * THE INSTITUTE OF SOCIAL SCIENCES

DEPARTMENT OF WESTERN LANGUAGES AND LITERATURE

APPLIED LINGUISTICS MASTER'S PROGRAM

**A LONGITUDINAL LEARNER CORPUS INVESTIGATION OF THE USE AND
DEVELOPMENT OF FORMULAIC
SEQUENCES IN ARGUMENTATIVE ESSAYS OF EFL LEARNERS**

MASTER'S THESIS

Hanife ÖZTÜRK

JUNE-2021

TRABZON

KARADENİZ TECHNICAL UNIVERSITY * THE INSTITUTE OF SOCIAL SCIENCES

DEPARTMENT OF WESTERN LANGUAGES AND LITERATURE

APPLIED LINGUISTICS MASTER'S PROGRAM

**A LONGITUDINAL LEARNER CORPUS INVESTIGATION OF THE USE AND
DEVELOPMENT OF FORMULAIC
SEQUENCES IN ARGUMENTATIVE ESSAYS OF EFL LEARNERS**

MASTER'S THESIS

Hanife ÖZTÜRK

Supervisor: Asst. Prof. Dr. Ali Şükrü ÖZBAY

JUNE - 2021

TRABZON

DECLARATION OF ORIGINALITY

I, **Hanife ÖZTÜRK**, hereby confirm and certify that;

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

Signature.....

Date 22.06.2021

ACKNOWLEDGEMENT

Corpus-based studies have shown that language use in English comprises of a large number of repeated patterns. Formulaic sequences, which are widely used in both written and spoken language, are considered important in many respects for EFL learners.

The prevalence of the formulaic sequences in English and their advantages for the learners in language use have led many researchers to emphasize the benefits of using these sequences and to investigate them from different aspects. The use of these sequences enables to fluency and clarity while writing. In addition to these, it was assumed that the more they used these sequences, the more proficient the learners are likely to become in the language they are learning. Learner corpus studies have become a recent focus and revealed that the use of formulaic sequences may result in overuse or underuse problems.

In this Study, longitudinal learner corpora were compiled from 85 English as a Foreign Language (EFL) learners' argumentative essays during two semesters, and the use and development of three- to four-word sequences were investigated. For this purpose, a group analysis and an individual analysis were performed within the scope of quantitative analysis, and a retrospective protocol was made within the scope of qualitative analysis. The results of the quantitative analysis indicated that the frequency and range of formulaic sequences tended to increase. The majority of these sequences were comprised of verb phrase fragments on the basis of structural categorisation. On the other hand, referential expressions were the most common ones on the basis of functional categorisation. Individual analysis revealed that the learners used unique sequences. The results of the qualitative analysis showed that teacher feedback assisted to increase the learners' exposure and awareness of formulaic sequences, and the formulaic sequences were used frequently by the learners over time.

I would like to thank Asst. Prof. Dr. Ali Şükrü ÖZBAY and Asst. Prof. Dr. Nazan YILDIZ, whose knowledge and experiences contributed to all parts of the study.

June, 2021

Hanife ÖZTÜRK

TABLE OF CONTENTS

ACKNOWLEDGEMENT	IV
TABLE OF CONTENTS.....	V
ÖZET.....	VIII
ABSTRACT	IX
LIST OF TABLES	X
LIST OF FIGURES	XII
LIST OF ABBREVIATIONS.....	XIV

INTRODUCTION.....	1-4
-------------------	-----

CHAPTER ONE

1. FRAMEWORK OF THE STUDY.....	5-11
1.1. Background of the Study.....	5
1.2. Statement of the Problem	6
1.3. Purpose of the Study	8
1.4. Significance of the Study	8
1.5. Research Questions	9
1.6. Hypothesis.....	9
1.7. Assumptions.....	10
1.8. Definition of Terms.....	10
1.9. Organization of the Thesis	10

CHAPTER TWO

2. LITERATURE REVIEW.....	12-31
2.1. Phraseology.....	12
2.2. Lexicology	14
2.3. Formulaic Language	15
2.3.1. Structural and Functional Characteristics of Formulaic Language.....	17
2.4. Significance of Formulaic Sequences and Frequency of FSs	20
2.5. Acquisition of Formulaic Sequences by L2 Learners	22
2.6. Corpus Linguistics and Corpus Tools	23

2.7. Corpus Research on Formulaic Language	26
2.8. Native and Non-native Speakers in Applied Linguistics	30

CHAPTER THREE

3. METHODOLOGY	32-44
3.1. Introduction.....	32
3.2. Participants.....	32
3.3. Participant Training.....	33
3.4. Corpora.....	36
3.4.1. Longitudinal Learner Corpora	36
3.4.2. Native Written Corpus: LOCNESS	37
3.5. Analytical Procedures	37
3.5.1. Identification of Formulaic Sequences	40
3.5.3. EFLs' Formulaic Sequences Compared to Native Speaker	41
3.5.4. Individual EFLs' Inventories of Formulaic Sequences.....	42
3.5.5. EFL Inventories of Formulaic Sequences: Retrospective Protocol	43
3.6. Raters' Profile	44

CHAPTER FOUR

4. RESULTS AND DISCUSSION.....	45-146
4.1. Introduction.....	45
4.2. Group Analysis	45
4.2.1. Analysis of 3- and 4-Word Formulaic Sequences.....	45
4.2.2. Structural and Functional Analysis.....	51
4.2.2.1. Structural Analysis	51
4.2.2.2. Functional Analysis.....	66
4.2.3. Comparison of Native and Non-native Formulaic Sequences	73
4.3. Individual Analysis: EFL Inventories of Formulaic Sequences.....	78
4.3.1. Gizem's Inventory of Formulaic Sequences.....	78
4.3.1.1. Structures of Formulaic Sequences in Gizem's each Sub-corpus	80
4.3.1.2. Functions of Formulaic Sequences in Gizem's each Sub-corpus	82
4.3.1.3. Unique Formulaic Sequences in Gizem's each Sub-corpus.....	83
4.3.2. Elif's Inventory of Formulaic Sequences.....	84
4.3.2.1. Structures of Formulaic Sequences in Elif's each Sub-corpus.....	86
4.3.2.2. Functions of Formulaic Sequences in Elif's each Sub-corpus	88

4.3.2.3. Unique Formulaic Sequences in Elif’s each Sub-corpus	90
4.3.3. Arda’s Inventory of Formulaic Sequences.....	91
4.3.3.1. Structures of Formulaic Sequences in Arda’s each Sub-corpus.....	93
4.3.3.2. Functions of Formulaic Sequences in Arda’s each Sub-corpus	95
4.3.3.3. Unique Formulaic Sequences in Arda’s each Sub-corpus	97
4.3.4. Lale’s Inventory of Formulaic Sequences	98
4.3.4.1. Structures of Formulaic Sequences in Lale’s each Sub-corpus	100
4.3.4.2. Functions of Formulaic Sequences in Lale’s each Sub-corpus.....	102
4.3.4.3. Unique Formulaic Sequences in Lale’s each Sub-corpus	104
4.3.5. Betül’s Inventory of Formulaic Sequences.....	105
4.3.5.1. Structures of Formulaic Sequences in Betül’s each Sub-corpus	107
4.3.5.2. Functions of Formulaic Sequences in Betül’s each Sub-corpus	109
4.3.5.3. Unique Formulaic Sequences in Betül’s each Sub-corpus.....	111
4.3.6. Ali’s Inventory of Formulaic Sequences	112
4.3.6.1. Structures of Formulaic Sequences in Ali’s each Sub-corpus	114
4.3.6.2. Functions of Formulaic Sequences in Ali’s each Sub-corpus.....	116
4.3.6.3. Unique Formulaic Sequences in Ali’s each Sub-corpus	118
4.3.7. Aslı’s Inventory of Formulaic Sequences.....	119
4.3.7.1. Structures of Formulaic Sequences in Aslı’s each Sub-corpus.....	121
4.3.7.2. Functions of Formulaic Sequences in Aslı’s each Sub-corpus	123
4.3.7.3. Unique Formulaic Sequences in Aslı’s each Sub-corpus.....	125
4.3.8. Zeynep’s Inventory of Formulaic Sequences.....	126
4.3.8.1. Structures of Formulaic Sequences in Zeynep’s each Sub-corpus.....	128
4.3.8.2. Functions of Formulaic Sequences in Zeynep’s each Sub-corpus	130
4.3.8.3. Unique Formulaic Sequences in Zeynep’s each Sub-corpus	132
4.4. Data Analysis of the Retrospective Protocols	133
4.5. Discussion	139
CONCLUSION AND RECOMMENDATIONS	147
REFERENCES.....	151
APPENDIXES	169
CURRICULUM VITAE.....	212

ÖZET

Bu tez, İngilizce'yi yabancı dil (EFL) olarak öğrenen 85 öğrencinin iki dönemlik yazılı verilerinden oluşan boylamsal öğrenen derleminde üçlü ve dördlü kelime dizinlerinin kullanımını araştırmayı amaçlamaktadır. Analiz iki yönlüdür: Birincisi, farklı zaman aralıklarında söz dizinlerinin kullanımı açısından toplu eğilimleri gözlemlemek için sağlanan grup analizidir; ikincisi, iki dönem boyunca 8 öğrencinin söz dizinlerini kullanımını gözlemlemeye yardımcı olan bireysel analizdir. Frekans yaklaşımına dayalı olarak, en sık kullanılan üçlü ve dördlü kelime dizinleri boylamsal öğrenen derleminin iki grubunun her alt derleminde bulunmuş ve yapısal ve işlevsel olarak sınıflandırılmıştır. Sonrasında, bu dizinlerin kullanımı açısından, anadili İngilizce olan (LOCNESS) ve anadili İngilizce olmayan derlemler arasında karşılaştırma yapılmıştır. Bireysel analiz, iki dönemlik gözlem boyunca elde edilen beş alt derleminde belirli öğrenciler tarafından sıklıkla kullanılan özgün dizinlerin belirlenmesini içerir. Boylamsal öğrenen derlem bulgularının nicel analizini, öğrencilerin geriye dönük protokollere verdiği yanıtların nitel analizi takip etmiştir. Söz dizinlerinin frekans, korelasyon, yapısal ve işlevsel analizleri göstermiştir ki öğrencilerin yazmış oldukları metinlere iki dönem boyunca geri bildirim verildiğinde, söz dizinlerinin sayısı ve çeşitliliği bakımından artışlar olmuştur. Yapısal olarak, söz dizinlerinin çoğunluğunun fiil öbeği parçaları olduğu, ardından isim öbeği ve edat öbeği parçaları olduğu bulunmuştur. İşlevsel olarak, bu çalışmada analiz edilen söz dizinlerinin çoğu, önceki literatürde olduğu gibi referans ifadeleri ve ardından tutum ifadelerini içermiştir. Ayrıca Pearson korelasyon testi, iki öğrenen derlem grubunda: 1. grubun beş alt derleminin sıklıkla kullanılan söz dizinleri açısından anadil İngilizce olan derlemi ile orta derecede korelasyon olduğu 2. gruptan farklı olarak görülmüştür. Bireysel analizin sonuçları, belirli öğrencilere özgü bazı söz dizinlerinin olduğunu göstermiştir. Başka bir deyişle, söz dizinlerin yarısından fazlası kelimesi kelimesine veya kısmen boylamsal öğrenen derlemleri ortak kullanılırken, anadili İngilizce olan derlem (LOCNESS) ile daha az söz dizinleri ortak kullanılmıştır. Son olarak, geriye dönük protokolün sonuçları, katılımcıların söz dizinlerine maruz kalmalarının ve farkındalıklarının öğretmen geri bildirim yoluyla arttığını ve bu nedenle zaman içinde daha fazla söz dizini kullandıklarını göstermiştir.

Anahtar Kelimeler: Söz Dizimleri (FSs), Öğrenen Derlemi, İngilizce'yi Yabancı Dil (EFL) Olarak Öğrenenler, Özgün Söz Dizimleri

ABSTRACT

The current thesis investigates the usage patterns of three- to four-word sequences in longitudinal learner corpora composed of two semesters written data from 85 English as a Foreign Language learners (EFL). The analysis is twofold: the first one is a group analysis that made available to observe collective trends in terms of usage patterns of formulaic sequences across different time intervals, and the second one is an individual analysis that assisted in observing the use of formulaic sequences of the same 8 individual learners across two semesters. Based on the frequency approach, the most frequent three and four-word recurrent formulaic sequences were extracted from each sub-corpus in two groups of longitudinal learner corpora and classified structurally and functionally. Then, the use of these sequences was compared across native (LOCNESS) and non-native corpora. The individual analysis involved identifying unique sequences used frequently by particular learners across two semesters observation in the five sub-corpora. Quantitative analysis of the longitudinal learner corpora findings was followed by a qualitative analysis of the learners' responses to retrospective protocols. The frequency analyses, correlation statistical test, and the structural and functional analyses of the formulaic sequences showed that the number and the range of FSs seemed to show an increasing pattern in number and type, as the learners were given more instruction and teacher feedback regarding their essays for each week during two semesters. Structurally, the majority of formulaic sequences were found to be verb phrase fragments, followed by noun phrase and prepositional phrase fragments. Functionally, most formulaic sequences analysed in this study include referential expressions, followed by stance expressions like in previous literature. The Pearson correlation test also showed that the frequent FSs in the two learner corpora seemed to moderately correlate, in Group 1, with the native learner corpora in the five sub-corpora unlike Group 2. The results of the individual analysis showed some formulaic sequences that were unique to the particular learners. In other words, more than half of the formulaic sequences shared word-for-word or partially with longitudinal learner corpora whereas they shared less formulaic sequences with native written corpus (LOCNESS). Last but not least, the results of the retrospective protocol showed that participants' exposure and awareness of FSs increased through teacher feedback, and they used more FS across time intervals.

Keywords: Formulaic Sequences (FSs), Learner Corpus, EFL Learners, Unique Sequences

LIST OF TABLES

Table No	Table Name	Page Nr.
1	Structural Classification of Language Formulas	19
2	Functional Classification of Language Formulas	19
3	Previous Research on Formulaic Sequences	27
4	Demographic Information of Participants	33
5	Further Training on the Weekly Basis	33
6	The Profiles of Longitudinal Learner Corpora in Two Groups	36
7	The Profiles of Native Written Corpus and Longitudinal Learner Corpora	37
8	Participants' Demographic Information of Individual Analysis.....	43
9	Timeline of the Protocol Sessions with Each Sample	44
10	Demographic Information of Raters	44
11	Total Number of Token and Type Frequency of 3- to 4-Word Sequences in Group 1 and Group 2	45
12	Top 10 Frequent Formulaic Sequences over Time in Group 1.....	46
13	Top 10 Frequent Formulaic Sequences over Time in Group 2.....	48
14	Shared Frequent 3- and 4-Word Formulaic Sequences over Time in Group 1.....	49
15	Shared Frequent 3- and 4-Word Formulaic Sequences over Time in Group 2.....	50
16	Structure of the Top 100 Frequent Formulaic Sequences in Sub-corpus 1 and Sub-corpus 2 of Group 1 and Group 2.....	53
17	Structure of the Top 100 Frequent Formulaic Sequences in Sub-corpus 3 and Sub-corpus 4 of Group 1 and Group 2.....	58
18	Structure of the Top 100 Frequent Formulaic Sequences in Sub-corpus 5 of Group 1 and Group 2	62
19	Function of the Top 100 Frequent Formulaic Sequences in Sub-corpus 1 and Sub-corpus 2 of Group 1 and Group 2.....	67
20	Function of the Top 100 Frequent Formulaic Sequences in Sub-corpus 3 and Sub-corpus 4 of Group 1 and Group 2.....	69
21	Function of the Top 100 Frequent Formulaic Sequences in Sub-corpus 5 of Group 1 and Group 2	70
22	Correlations between Sub-corpus 1 of Group 1 and LOCNESS	74
23	Correlations between Sub-corpus 2 of Group 1 and LOCNESS	74
24	Correlations between Sub-corpus 3 of Group 1 and LOCNESS	74
25	Correlations between Sub-corpus 4 of Group 1 and LOCNESS	74

26	Correlations between Sub-corpus 5 of Group 1 and LOCNESS	75
27	Correlations between Sub-corpus 1 of Group 2 and LOCNESS	76
28	Correlations between Sub-corpus 2 of Group 2 and LOCNESS	76
29	Correlations between Sub-corpus 3 of Group 2 and LOCNESS	76
30	Correlations between Sub-corpus 4 of Group 2 and LOCNESS	76
31	Correlations between Sub-corpus 5 of Group 2 and LOCNESS	77
32	Top 20 Frequent Formulaic Sequences Produced by Gizem across 5 Sub-corpora.....	79
33	Structure of Formulaic Sequences in Gizem’s each Sub-corpus	80
34	Function of Formulaic Sequences in Gizem’s each Sub-corpus	82
35	Top 20 Frequent Formulaic Sequences Produced by Elif across 5 Sub-corpora.....	85
36	Structure of Formulaic Sequences in Elif’s each Sub-corpus.....	86
37	Function of Formulaic Sequences in Elif’s each Sub-corpus	89
38	Top 20 Frequent Formulaic Sequences Produced by Arda across 5 Sub-corpora.....	92
39	Structure of Formulaic Sequences in Arda’s each Sub-corpus.....	93
40	Function of Formulaic Sequences in Arda’s each Sub-corpus	95
41	Top 20 Frequent Formulaic Sequences Produced by Lale across 5 Sub-corpora	99
42	Structure of Formulaic Sequences in Lale’s each Sub-corpus.....	100
43	Function of Formulaic Sequences in Lale’s each Sub-corpus	102
44	Top 20 Frequent Formulaic Sequences Produced by Betül across 5 Sub-corpora.....	106
45	Structure of Formulaic Sequences in Betül’s each Sub-corpus	107
46	Function of Formulaic Sequences in Betül’s each Sub-corpus	109
47	Top 20 Frequent Formulaic Sequences Produced by Ali across 5 Sub-corpora.....	113
48	Structure of Formulaic Sequences in Ali’s each Sub-corpus.....	114
49	Function of Formulaic Sequences in Ali’s each Sub-corpus	117
50	Top 20 Frequent Formulaic Sequences Produced by Aslı across 5 Sub-corpora.....	120
51	Structure of Formulaic Sequences in Aslı’s each Sub-corpus	121
52	Function of Formulaic Sequences in Aslı’s each Sub-corpus	124
53	Top 20 Frequent Formulaic Sequences Produced by Zeynep across 5 Sub-corpora.....	127
54	Structure of Formulaic Sequences in Zeynep’s each Sub-corpus.....	128
55	Function of Formulaic Sequences in Zeynep’s each Sub-corpus	130
56	Protocol Questions and their Coded Categories	133
57	The Encoded Analysis of the Protocol Questions	134

LIST OF FIGURES

Figure No	Figure Name	Page Nr.
1	The Process of Compiling Longitudinal Learner Corpora	37
2	Research Design	39
3	The Distribution of the Type Frequency of the Structural Categories of the Top 100 Frequent Formulaic Sequences across 5 Sub-corpora of Group 1.....	64
4	The Distribution of the Type Frequency of the Structural Categories of the Top 100 Frequent Formulaic Sequences across 5 Sub-corpora of Group 2.....	65
5	The Distribution of the Type Frequency of the Functional Categories of the Top 100 Frequent Formulaic Sequences across 5 Sub-corpora of Group 1	71
6	The Distribution of the Type Frequency of the Functional Categories of the Top 100 Frequent Formulaic Sequences across 5 Sub-corpora of Group 2	72
7	Scatter Plots of the Relationship between Frequency Scores in LOCNESS and Each Sub-corpus of Group 1 in Longitudinal Learner Corpora	75
8	Scatter Plots of the Relationship between Frequency Scores in LOCNESS and Each Sub-corpus of Group 2 in Longitudinal Learner Corpora	77
9	Distribution of Structural Types of Formulaic Sequences Produced by Gizem across 5 Sub-corpora.....	82
10	Distribution of Functional Types of Formulaic Sequences Produced by Gizem across 5 Sub-corpora.....	83
11	Distribution of Structural Types of Formulaic Sequences Produced by Elif across 5 Sub-corpora.....	88
12	Distribution of Functional Types of Formulaic Sequences Produced by Elif across 5 Sub-corpora.....	90
13	Distribution of Structural Types of Formulaic Sequences Produced by Arda across 5 Sub-corpora.....	95
14	Distribution of Functional Types of Formulaic Sequences Produced by Arda across 5 Sub-corpora.....	97
15	Distribution of Structural Types of Formulaic Sequences Produced by Lale across 5 Sub-corpora.....	102
16	Distribution of Functional Types of Formulaic Sequences Produced by Lale across 5 Sub-corpora.....	104
17	Distribution of Structural Types of Formulaic Sequences Produced by Betül across 5 Sub-corpora.....	109
18	Distribution of Functional Types of Formulaic Sequences Produced by Betül across 5 Sub-corpora.....	111

19	Distribution of Structural Types of Formulaic Sequences Produced by Ali across 5 Sub-corpora.....	116
20	Distribution of Functional Types of Formulaic Sequences Produced by Ali across 5 Sub-corpora.....	118
21	Distribution of Structural Types of Formulaic Sequences Produced by Aslı across 5 Sub-corpora.....	123
22	Distribution of Functional Types of Formulaic Sequences Produced by Aslı across 5 Sub-corpora.....	125
23	Distribution of Structural Types of Formulaic Sequences Produced by Zeynep across 5 Sub-corpora.....	130
24	Distribution of Functional Types of Formulaic Sequences Produced by Zeynep across 5 Sub-corpora.....	132



LIST OF ABBREVIATIONS

EFL	: English as Foreign Language
ESL	: English as Second Language
L1	: First / Native language
L2	: Second / Foreign Language
FS/FSs	: Formulaic Sequence(s)
LOCNESS	: Louvain Corpus of Native English Essays
VP	: Verb Phrase
NPPP	: Noun Phrase and Prepositional Phrase
DC	: Dependent Clause
OE	: Other Expressions
SE	: Stance Expressions
DO	: Discourse Organizers
RE	: Referential Expressions
N / n	: Number / Frequency
Norm	: Normalized
Freq/f	: Frequency

INTRODUCTION

The study of phraseology is one of the fields that appears at the core of applied linguistic studies and has been studied frequently in recent years by Simpson-Vlach and Ellis (2010), Pastor and Mitkov (2019), Oakey (2020). Cowie (1994) defines phraseology as “the study of the structure, meaning and use of word combinations” (1994: 3168), and it is grounded on Sinclair’s (1991: 110) idiom principle that is based on the premise that “a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments”. It is also a broader term that covers many types of word combinations and refers to different phraseological variations such as formulas, formulaic sequences, multi-word units, collocations, idioms and lexical bundles (Granger and Paquot, 2008; Philip, 2008; Piirainen, 2008; Huang, 2014; Benigno and Kraif, 2016; Pastor and Mitkov, 2019). Phraseology finds numerous applications in such fields as lexicography, English language teaching and learning, stylistics etc. Specifically, corpus-based phraseological research has brought out the importance of frequent FSs in writing (Salazar, 2011). On the other hand, it can be seen in the literature that the accessibility and manageability of large written and spoken language data in the form of electronic corpora and the spread of new research methods in the field of corpus linguistics assisted to grasp the pervasive nature of formulaic sequences in language use.

Reviewing the recent literature shows that formulaic sequences are viewed as ‘big word’, and they are stored in mind as single items (Ellis and Sinclair, 1996: 245), and they comprise of multi-word expressions and behave like single word lexis (Wong, 2012). Different subtypes such as recurrent word sequences and lexical bundles have been examined in various earlier studies (e.g., Altenberg, 1998; Cortes, 2004; Hyland, 2008). The term formulaic sequence was first introduced into the literature by Wray (2000), and this framework has been applied in many subsequent studies as an umbrella term or “the most comprehensive term” (Schmitt and Carter, 2004: 4) to refer to some of the other types and as a particular term to refer to one of the other subtypes such as lexical phrases, prefabricated patterns, collocations, fixed expressions, lexical bundles, multi-word units and chunks (e.g., Wray, 2000; Wray and Perkins, 2000; Wray, 2002; Wray, 2017; Schmitt, 2004; Schmitt and Carter, 2004; Wood, 2015; Conklin and Schmitt, 2008; O’Donnell et al. 2013). The operational definition of formulaic sequence here is: “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (Wray, 2002: 9), and for the sake of convenience, this study uses the term formulaic sequence such other studies as an umbrella among over fifty subtypes.

A thorough analysis of the existing literature has shown that there has been a growing interest in formulaicity and formulaic sequences since they are widespread in both written and spoken language discourse (Conklin and Schmitt, 2008; Ohlrogge, 2009; Schmitt, 2013; Hatami, 2015), and they play a significant role in the development and usage of L2 to ensure successful communication. Thus, these sequences have been one of the main concerns of the studies in the field of applied linguistics (Cowie, 2001; Wray, 2002; Schmitt, 2004; Ellis, 2008, Granger and Meunier, 2008; Römer, 2010; O'Donnell et al. 2013). L2 learners are perceived as more proficient learners when they use formulaic sequences in their linguistic productions. That is, the use of these sequences is an indicator of proficient language use (Cortes, 2004; Geluso and Yamaguchi, 2014; Hatami, 2015). Likewise, Pawley and Syder (1983) considered that they are essential to fluent linguistic production, particularly in spoken discourse. Ranjbar et al. (2012) found that there is a significant relationship between the use of FSs and writing fluency. In the same way, both efficient and appropriate language usages require the use of formulaic sequences since “interlocutors expect them, and they are the preferred choices” (Schmitt and Carter, 2004: 10). It is claimed that the lack or scarcity of these sequences is regarded as inadequate and too colloquial for the learner’s writing (Jones and Haywood, 2004; Coxhead and Byrd, 2007; Gilquin and Paquot, 2008; Hyland, 2012). The lack and misuse of formulaic language are one of the main reasons why the written productions of non-native learners may sound unnatural (Schmitt, 2013; Peters and Pauwels, 2015). Thanks to the use of more and appropriate FSs, language learners can achieve naturalness in L2, particularly in writing (Allen, 2009; Razak, 2015). Likewise, Hyland (2008) suggested that they “are familiar to writers and readers who regularly participate in a particular discourse, their very ‘naturalness’ signalling competent participation in a given community” (2008: 5). In these statements, naturalness is used as being synonymous with authenticity (Warren, 2006).

On the other hand, there are a number of advantages of learning and employing formulaic sequences. For example, they both assist to automatize the linguistic performance (Schauer and Adolphs, 2006) and are considered as a primary way of successful communication (Ohlrogge, 2009), and these sequences also reduce the overall processing load (Wray, 2013). To sum up, FSs are regarded as central in idiomatic and fluent use of language (Pawley and Syder, 1983; Nattinger and DeCarrico, 1992; Ellis, 1996; Erman and Warren, 2000), on the other hand, they assist speaker production and hearer comprehension by saving effort in processing (Wray, 2000).

The study of formulaic language for EFL learners plays an essential role in learner corpus research (Paquot and Granger, 2012). Furthermore, “corpus has become less of a buzzword and more of a necessary, acknowledged reference source for students, linguists, language professionals, teachers, translators, technical writers, lexicographers, etc.” (Bernardini, 2004: 32). In the field, the contrastive analysis of native and learner corpora has come into prominence in language learning and teaching contexts for many reasons. It is generally known that learner corpora are described as “electronic collections of authentic FL/SL” (Sinclair, 1991: 2), and “systematic computerized

collections of texts produced by language learners” (Nesselhauf, 2004: 125). In this regard, native speaker corpora bring out “what native speakers of the language in question typically write or say (either in general or in a certain situation / in a certain text type)” (Nesselhauf, 2004: 125). The propagation of the contrastive method leads up further studies on formulaic sequences in native and non-native corpora. It is seen that “the most frequently used method of investigating formulaic language in a learner corpus has been to compare the results of a learner corpus analysis with those obtained from the analysis of a comparable native corpus and identify errors and patterns of learner over- and underuse of formulaic sequences” (Paquot and Granger, 2012: 132). In addition to this, the studies of FSs compare the functional and structural characteristics of these sequences in native and learner corpora (Paquot and Granger, 2012).

The use of formulaic sequences by EFL learners are not without problems either. According to Ohlrogge (2009), authors remark that formulaic sequences “often break the ‘rules’ of language, whether through lexical abnormalities (e.g., *kith and kin*), grammatical abnormalities (e.g., *by and large*), or by an idiomatic or otherwise metaphorical meaning (e.g., *on the other hand*)” (2009: 375). It is therefore not surprising that these sequences lead to a special challenge for L2 learners, who must learn both “what form(s) and meaning(s) can and cannot be associated with a given sequence” (2009: 375) and “how to incorporate learned sequences into larger pieces” (2009: 375-376) in written and spoken language use. The acquisition and usage of formulaic sequences by second language learners seem as a problematic issue (Lenko-Szymanska, 2014; Wray 2002; Bishop, 2004; Schmitt, 2013). The results of Bishop’s (2004) study are consistent with the claim that formulaic sequences are not noticed L2 learners, thus they are not learned. In this direction, Lenko-Szymanska’s (2014) study suggests that the acquisition of word strings may be affected by the proficiency of learners. What is more, it is remarked that “producing natural, idiomatic English is not just a matter of constructing well-formed sentences, but of using well-learned lexical expressions in appropriate places” (Biber et al., 1999: 990). Iwatsuki and Aizawa (2018) asserted that when L2 speakers write scholarly papers, they often use the same sequences or word strings repeatedly, and also they are not confident in the sense of correctness of their usage of these FSs. Likewise, non-native learners do not always use FSs appropriately and accurately in their language usage (Howarth, 1998; Nesselhauf, 2003). According to the claim by Khusnita and Rukmini (2016), Indonesian EFL learners have tendencies to translate the native language into the target language which is English word by word instead of using formulaic sequences in their discourse. Thus, this results in unnatural expressions because of learners’ lack of formulaic competence, so the use or recall of FSs becomes difficult. The findings of the study by Geluso and Yamaguchi (2014) demonstrate that EFL learners may encounter difficulties using novel formulaic sequences in a natural way in language use. The non-native mastery of formulaic language can be analysed under three dimensions, and these are accuracy or appropriacy of use, amount of use and quality or speed of the underlying formulaic intuitions (Schmitt, 2013). It can be seen that the overuses and underuses are among the problems of formulaic sequences in non-native learners’ language use. Whereas L2 learners tend to overuse FSs that they know well the

reason why they are considered as safe bets, they also tend to underuse FSs that they do not know well how to use these sequences (Granger, 1998). When formulaic sequences are analysed in the sense of accuracy or appropriacy of usage by non-native learners, it may be noticed that how these FSs can be non-native (Nesselhauf, 2003, 2005). In the study of Nesselhauf (2003), the analysis of the use of verb-noun collocations (e.g., *take a break*) in free written production showed that even advanced learners of English have difficulty in the production of collocations. Both combining words in an inaccurate way and using combinations inappropriate way are the reason for mistakes in written productions of learners. In the sense of the quality of intuitions of formulaic language, the study of Siyanova and Schmitt (2008) demonstrated that non-native advanced learners have poor intuitions of recurrent sequence frequency in contrast to native speakers' accurate intuitions.



CHAPTER ONE

1. FRAMEWORK OF THE STUDY

1.1. Background of the Study

In recent years, the advent of computer technology and software tools have made available more complicated and fully operational facilities for corpus linguistics. In virtue of this development, the compilation of large collections of naturally occurring texts was made more accurately. On the other hand, these developments help the researchers to explore, process and analyse language data in various ways in different research areas (Aijmer and Altenberg, 2004). With the advancement of computerized corpora, the use of corpora “in the field of lexical studies has expanded a great deal” (Alquraishi, 2014: 23). Aijmer and Altenberg (2004) add that the studies on multi-word units, word lists and collocations offer insight into the significance of the use of corpora in various fields. This notion is supported by Özbay and Kayaoğlu (2016) who stated that “the use of corpus for lexical investigation is not a recent phenomenon but its full significance and value has, in the last decade, been realized especially after the introduction of computerized corpus tools by a much larger group of linguists all around globe” (343). In the field of foreign language learning and teaching, it is emphasised that the use of FSs is crucial for language skills, especially on writing proficiency. With the use of computer-assisted techniques in language environments, the researchers focused on these recurrent FSs, and they conducted large-scale studies. For instance, Simpson-Vlach and Ellis (2010) studied on FSs and derived the list of formulaic sequences for academic writing and speech by conducting corpus-based research, similarly, Coxhead (2000) generated an academic word list. Corpus-based research has provided evidence for revealing a great number of different repeated patterns in language use. Schmitt and Carter (2004) remarked that it has enlightened “the field by identifying formulaic language and describing how it is used in discourse” (2004: 11). Furthermore, according to corpus evidence, native language is wealthy in respect to formulaic sequences (Schmitt and Carter, 2004).

Because of the fact that FSs become more of an issue in linguistic production (Hyland, 2012), on the whole these sequences have started to receive considerable attention in writing (Biber and Barbieri, 2007; Hyland, 2008; Chen and Baker, 2010; Adel and Erman, 2012; Cortes, 2013; Huang, 2015; Staples and Reppen, 2016; Bychkovska and Lee, 2017), and the most frequent three and four-word sequences are the recurrent sequences within per million words such as *in order to, one of the, there is a, there is no, part of the, the number of, the presence of, the use of, the fact that; on the other*

hand and in the case of (Biber et al., 1999). Written language holds a large stock of different FSs and according to Qi and Ding (2011), “the past several decades have witnessed a growing body of research into the use of FSs in language learning” (165). Consequently, it is a need to investigate the usage patterns of three- and four-word formulaic sequences in writing of L2 learners, and search for to what extent non-native learners produce these frequently occurring sequences (Cortes, 2004; Hyland, 2008; Ishikawa, 2013; Elturki, 2015; Güngör and Uysal, 2016; Taşkaya, 2019; Ulfa and Muthalib, 2020).

According to Granger (2002), “the area of linguistic enquiry known as learner corpus research, which has only existed since the late 1980s, has created an important link between the two previously disparate fields of corpus linguistics and foreign/second language research” (4). That is to say, the compilation of learner corpora comprising L2 learners’ essays is the current trend in corpus-based research since it offers detailed descriptions of non-native language. It can be said that International Corpus of Learner English (ICLE) which was compiled by Granger (2003) is the pioneer of studies on learner corpora in EFL setting. Turkish Corpus of Learner English (TICLE) as the Turkish component of the ICLE corpus is one of the learner corpora in Turkish context. On the other hand, the Louvain Corpus of Native English Essays (LOCNESS) is native learner corpus which is frequently used as a control corpus to compare the non-native learner’ texts with native counterparts (e.g., Şanal, 2007; Lozano and Mendikoetxea, 2007; Pang, 2009; Taşkaya, 2019). This type of study assists to find out the acquisition, use and development of formulaic sequences; furthermore, it is possible to compare the proper use of these sequences in two different settings produced by native speakers of English (native learner corpora) and non-native speakers of English (non-native learner corpora).

1.2. Statement of the Problem

The efficiency of the use of formulaic sequences in both written and spoken language is perceived as a trace of competence in linguistic performance, in turn leading to more proficient learner writing (Boers et al., 2006; Coxhead and Byrd, 2007; Hyland, 2008; Wood, 2010; Housen et al., 2012; Martinez and Schmitt, 2012; Bestgen and Granger, 2014; Kyle and Crossley, 2015; Garner et al., 2019). As noted by Hyland (2008: 4) “multi-word expressions are an important component of fluent linguistic production and a key factor in successful language learning”, and therefore to notice the development of these sequences on foreign language learners’ text provides an insight on how they are progressed in writing. Several studies put forward the contrastive analysis between L1 and L2 learners on these sequences in order to see how they are similar or different from each other, and eventually this comparison provides a common framework for researchers. So, formulaic sequences are the basics of fluency in writing or the key to achieving efficiency in language use. “Recurrent word combinations are more frequent overall in native than non-native production” (Adel and Erman,

2012: 83). In Adel and Erman's study (2012), the number and diversity of sequences utilized by native learners in their texts are much greater than that of L2 learners do.

As seen in the existing literature, FSs contribute to the increase in the production and fluency in language use, but the learners of a second language have faced difficulty in both learning and using these sequences as compared to single word lexical items. In other words, there are several common issues in the usage of formulaic sequences belonging to the L2 learners that the well-known or safe FSs have been employed to providing confidence for the learners (Granger, 1998). Besides, when they are compared to native speakers of the language, the overuses, underuses and misuses of FSs by L2 learners are observed (De Cook, 2000). The studies conducted in Turkey indicated that Turkish EFL learners have had several problems in terms of the use of the formulaic sequences while writing. The use of wrong lexical combinations or inappropriate words in writing is one of the problems of EFL learners, as stated by Özbay (2015). In addition to wrong word combinations, the overuse and underuse problems were found in several studies. For instance, the findings of the study of Öztürk and Köse (2016) showed that Turkish EFL learners overused the sequences in the written outputs compared to both native learners and scholars in terms of both the range and frequency. Similarly, the studies of Bal (2010) and Güngör and Uysal (2016) pointed out that Turkish EFL learners had the overuse problems compared to L1 users. Similarly, Taşkaya (2019) found that Turkish EFL learners generally overused and underused the sequences compared to native counterparts. On the other hand, the results of the study carried out by Güngör (2016) revealed that Turkish writers were likely to transfer the sequences from Turkish to English. It was also found by Müjdecı (2014) that the reliance on L1 collocational knowledge while using collocations in the L2 was observed in the usage of FSs by Turkish EFL learners.

According to Ellis et al. (2008), language instructors should be aware of the extensive usage of sequences and their prominence in language, and this knowledge "inform which formulas should be prioritized for instruction in learners at different stages of development and need" (379); henceforth, it is important to specify the development of language formulas in students' texts. In that way, it will be easier to notice the underuses and overuses of sequences in the sense of the variety and frequency compared to the preference of L1 learners' written productions. It is generally agreed by both common sense and expert knowledge that "learning a language other than the mother tongue (a second, foreign, or heritage language; henceforth, an L2) is a complex process that happens through and over time" (Ortega and Iberri-Shea, 2005: 26). Therefore, it is required to conduct longitudinal investigations in the L2 research area (Ortega and Iberri-Shea, 2005; Duan and Shi, 2021). According to Bestgen and Granger (2014), when the development of the same individual learners across a given time period is tracked, the result of this observation supplies longitudinal data regarding the development of the L2 writing, especially phrasicon. Hence, it is "essential to apply phraseological indices to truly longitudinal data" (Bestgen and Granger, 2014: 30). Although, there have been ongoing issues related to usage patterns of formulaic sequences by L2 learners in their written

outputs, the number of the studies may not be satisfactory, and they are scarce (Siyanova-Chanturia and Spina, 2020). So, this situation has the potential to motivate researchers to conduct longitudinal studies on formulaic sequences in a Turkish context. In the context of these issues, what motivated this study is the need to understand the development and use of FSs in longitudinal EFL learner corpora across different time intervals.

1.3. Purpose of the Study

The purpose of this study is to investigate the usage of 3- and 4-word sequences in 85 EFL learners' written outputs as well as to gain an understanding of the use and development of formulaic sequences. The focus of the study is also to examine the structures and functions of these formulaic sequences in the learners' essays. These written productions were gathered during two semesters, three- to four-word formulaic sequences were extracted from longitudinal learner corpora, and the use and development of formulaic sequences were analysed in two different groups. For both group analysis and individual analysis, obtained initial lists of these sequences were filtered, and the structural and functional characteristics of FSs were categorized by using Biber et al.'s (1999; 2004) structural and functional framework. The analysis was expected to reveal the correlation in terms of usages of the EFL learners and the possible correspondences were also investigated. For this investigation, longitudinal learner corpora and LOCNESS were used. Lastly, a retrospective protocol was made with the eight participants to better understand how they learn and use formulaic sequences and how these patterns change over time.

1.4. Significance of the Study

The corpus-based study of multi-word units, especially of formulaic sequences, can provide a clearer understanding of how these sequences are used in learners' texts. First and foremost, this study can be significant to both foreign language researchers and teachers of English language, particularly in their writing classes. As noted by Chenu and Jisa, formulaic sequences "provide a stepping-stone into language development" (2009: 27) and similarly, they are regarded as a cornerstone of L2 with regards to production and processing (Ellis, 1996; Oakey, 2002; Jones and Haywood, 2004; Schmitt and Underwood, 2004; Ellis et al., 2008; Wray and Fitzpatrick, 2008; Wray and Fitzpatrick, 2010).

In recent years, there has been a strong demand for the longitudinal corpus-based research on the FSs in an attempt to better understand the use and development of these sequences. In fact, raising awareness on how EFL learners employ these sequences in their written outputs across different time intervals is fundamental. Moreover, the importance of the current study as a corpus-based research lies in its deeper exploration of both collective trends and unique patterns, meaning recurrent sequences utilized by individual learners. With respect to obtaining a deeper understanding of the

use and development of formulaic sequences in EFL writing, this study aimed to foreground the voices of EFL learners via retrospective protocols. This may help us to gain insights into how they used FSs in their writing, EFL learners' awareness of FSs and their stance towards the collocational nature of the English language. Consequently, it may be required to effectively incorporate the phraseological nature of English into the curriculum to enhance the writing of EFL learners with respect to phraseological competence. When this phraseological perspective is applied, the learners may see usage patterns of FSs in new contexts, and their knowledge and awareness of these sequences may increase naturally.

Although there are various studies conducted abroad investigating the use and development of the formulaic sequences with learners in different settings (e.g., Ohlrogge, 2009; Paquot and Granger, 2012; O'Donnell et al., 2013; Staples et al., 2013; Elturki, 2015; Duan and Shi, 2021), in Turkey the number of the studies conducted on the use and development of FSs is very limited. Above all, the primary significance of this study comes from the fact that the current study is considered as one of the first longitudinal corpus-based studies examining the use and development of the FSs over time through both group and individual analysis in Turkey.

1.5. Research Questions

The current study addressed the following research questions:

1. What are the most frequent 3- and 4-word sequences found in the longitudinal learner corpus for two semesters of language development?
2. What are the structures and functions of the frequent formulaic sequences?
3. How similar or different are the frequent 3- to 4-word formulaic patterns produced by the learners from those found in native corpus?
4. What do individual participants' inventories of formulaic sequences look like in terms of unique patterns?
5. To what extent, EFL learners become aware of the existence of formulaic sequences in their essays across two semesters?

1.6. Hypothesis

According to Coxhead and Byrd (2007), formulaic sequences “are often repeated and become a part of the structural material used by EFL and ESL writers, making the students' task easier because they work with ready-made sets of words rather than having to create each sentence word by word” (134-135). In line with the statements, the study formulated the following hypothesis:

“If EFL learners are more frequently exposed to different set of formulaic sequence types from their previous language instruction onward, this will be a positive contributing factor to their language development.”

1.7. Assumptions

Considering the fact that those learners previous took prep class language education for a year with heavy emphasis on writing and reading skills, it can be assumed that they would tell the difference between words and word combinations or at least that they would retrieve as a chunk in the writing process instead of treating them as individualized item.

1.8. Definition of Terms

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

Raw frequency: Raw frequency refers merely token counts that is it means how often a feature appears in research data.

Normalized frequency: Normalized frequency refers to a calculation of how often a feature appears in the determined number of words (e.g., per million tokens) when the frequencies are compared across two or more corpora in different sizes.

Word token frequency: It refers the number of individual words in the corpus.

Word type frequency: It refers the number of unique forms of words or types in a word frequency list of the corpus.

1.9. Organization of the Thesis

The current study aims at investigating the use and development of three- and four-word sequences in EFL learners’ written outputs across different time intervals. This thesis consists of five chapters.

Chapter 1, Framework of the Study: This section clearly defines the rationale, aim and direction, and the needs for conducting the study. It also displays why this particular thesis topic is essential to understanding the basic aspects of the research.

Chapter 2, Literature Review: This section is concerned with literature review and offers a review of literature associated with the scope of formulaic sequences such as its definition, importance and acquisition.

Chapter 3, Methodology: This section outlines the method of the research procedure in following subheadings: participants, participant training, corpora, analytical procedures and raters' profile.

Chapter 4, Results and Discussion: This section brings to light the findings of the quantitative and qualitative data. It deals with formulaic sequences and their frequencies, structures and functions in the corpus of non-native learners, and gives an account of the frequency and structural, functional and unique features of formulaic sequences in individual learner's each sub-corpus. Lastly, it presents the results of retrospective protocols. Obtained statistical data are introduced through tables and figures.

Chapter 5, Conclusion and Recommendations: This section is dedicated to the concluding remarks and recommendations for future research.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Phraseology

A thorough analysis of the existing literature has shown that the term phraseology has become one of the focus of language-related studies thanks to its notably wider scope. In the literature, the term phraseology tends to be used to refer to “the study of the structure, meaning, and use of word combinations” (Cowie, 1994: 316) while it is mentioned as the study of word combinations by Howarth (1998). For Gries (2008), phraseology means “the co-occurrence of a form or a lemma of a lexical item and any other kind of linguistic elements (word/grammatical patterns)” (2008: 4), and it “does double duty as a specific term for a particular type of analysis of formulaic language” (Wood, 2015: 2). Hunston (2011) added that it can be described as the tendency of word occurrences likely to be encountered frequently in some settings more than the others. It can be deduced from all these definitions of phraseology that it embodies various kinds of concepts such as idioms, fixed phrases, and collocations (Altenberg, 1998). In recent years, as the role of phraseology has become visible in applied linguistics, the studies on phraseology increased through its potential to illustrate the tendency of words or group of words through the co-selection in conversation and written production (Cheng et al., 2008).

According to Granger and Paquot (2008), phraseology has vague borders within four related disciplines (semantics, morphology, syntax and discourse). Furthermore, they propose two phraseological approaches that constitute certain boundaries. These approaches are the phraseological approach and the distributional or frequency-based approach. It can be remarked that the Collins COBUILD project which is the work of Sinclair is seen as a pioneer for the distributional approach (Granger and Paquot, 2008). This approach “generates a wide range of word combinations, which do not all fit predefined linguistic categories” (Granger and Paquot, 2008: 29).

As noted by various field workers, phraseology has several features in attempt to specify the scope of the field. It is important to note that idiomaticity, fixedness, semantic unity and frequency of co-occurrence are accepted as the main properties of the phraseology, and this is supported by Huang (2014). Likewise, the common properties of phraseology appeared in consequence of the review of several studies by Colson (2008), who named these features as polylexicality, fixedness and idiomaticity. One of these features is idiomaticity and it has the potential to become the major

feature for phraseology. Kennedy (2008) asserted that the studies of phraseology in language learning context are in tendency to highlight the idiomatic sequences (e.g., heavy rain). These sequences show the convention routines of speakers in their productions because they “embrace the conventional rather than the productive or rule-governed side of language” (Altenberg, 1998: 101).

Fixedness is another significant property of phraseology which is emphasized by several researchers such as Sinclair, Cowie and Gries. The term fixedness tends to be used to refer to the lexical flexibility and syntactic substitutability in a language, that is, it displays to what extent a word sequence is tied to each other lexically and syntactically. In the literature, it is asserted that the degree of fixedness may differ in terms of form and meaning. Accordingly, there is a continuum of fixedness (Cowie, 1981), and it is grouped into three categories considering the degree of fixedness as follows: free combinations, restricted collocations, and idioms (Howarth, 1998; Cowie, 1998). It can be deduced that idioms are strictly fixed sequences. For example, the idiom *a piece of cake* means *something that is very easy to do*, and as can be seen that four-word string assigns a different meaning to sequence. While the idioms are fixed combinations, there are many semi-fixed combinations as restricted collocations.

The third one in the feature of phraseology is semantic unity. Sinclair (1991) remarked that “many uses of words and phrases show a tendency to occur in a certain semantic environment” (1991: 112). The sequence *in spite of* is seen as a good example to demonstrate semantic unity that the co-occurrence of *in spite* is high in number more than expected by chance, that is to say, the example of these occurrences are instances of *in spite of* (Gries, 2008).

Frequency of co-occurrence, as another important property to specify phraseology, refers to the number of times a word string occurs in language use. It is clear that the frequency of co-occurrence helps to determine the sequences that should be attached priority in teaching and learning context since “frequency patterns are not accidental, but they are also not explanatory in themselves” (Biber, 2006: 173).

Phraseological patterns are used frequently in language productions and play an important role in the context of language teaching and learning since it can be suggested that they may provide more fluent and accurate production. Granger and Meunier (2008) stated that it “is a key factor in improving learners’ reading and listening comprehension, alongside fluency and accuracy in production” (251). Phraseological tendency occurs “where words tend to go together and make meanings by their combinations” (Sinclair, 2004: 29). Moreover, the importance of the phraseological tendency and its facilitative role in language processing are noticed in both EFL and ESL contexts. Being a major carrier of meaning, the phrase occupies at the centre of a language (Ellis, 2008). In case of the lack of the language patterns in the productions, it can be regarded as “the lack of fluency of a novice or newcomer to that community” (Hyland, 2008: 5). In addition,

phraseological units meet both L1 and L2 learners' needs. That is, "the widespread 'fusion' of expressions which appear to satisfy the individual's communicative needs at a given moment, and are later reused, is one means by which the public stock of formulae and composites is continuously enriched" (Cowie, 1988: 136). To sum up, it can be deduced that phraseological units are of significance for language learners.

2.2. Lexicology

It is widely known that "lexicology is the study of content words, or lexical items" (Halliday, 2004: 3). Towards the end of the twentieth century, there were considerable changes in the understanding of lexicology in terms of both theory and practice. It is affected by the development of the computer and corpus and investigators were able to conduct wide ranging lexical investigations on very large corpora of both written and spoken language as a result of new changes in the field of computer technology. The effect of computer and corpus on lexicology is clear thanks to the Collins COBUILD series of English dictionaries. It is an innovative corpus-based attempt since it has "every citation taken from real-life discourse" and "the way the different meanings of a word are described and classified can be worked out afresh from the beginning (instead of relying on previous dictionary practice) by inspecting how the word is actually used – what other words it collocates with, what semantic domains it is associated with, and so on" (Halliday, 2004: 17).

In language teaching and learning, the lexical approach is "derived from the belief that the building blocks of language communication teaching are not grammar, functions, notions or some other units of planning and teaching, but lexis" (Richards and Rogers, 2001: 132). It focuses on the lexicon and multi-word expressions which are learned and used as single units (Richards and Rogers, 2001). According to Hill (2000), the lexical approach sees a language in larger units as a whole rather than seeing the structures into smaller pieces as individual words. It is claimed that the lexical approach is an amalgamation of applied linguistics and language teaching methodology (Lewis, 2000).

Lexis plays a central role in language acquisition and "the lexical memory load, even for an intermediate learner, is enormous" (Lewis, 2000: 8). Likewise, Schmitt (2000) asserted that lexical knowledge is a major component of communicative competence and plays a pivotal role in second language learning and teaching. Özbay (2015) summarised some stages of lexical knowledge that learners are required to possess. These stages are the followings: production and reception, knowledge and control, breadth and depth, and word combinations, collocations, and phraseology. As seen above, the last stage is related to phraseology that includes FSs. By means of several changes on the lexicon, the term 'lexical item' has been replaced by the concept of 'word' from single to multi-word lexical items also called as phraseology (Qureshi and Akhter, 2019). It is clear that the large-scale body of authentic text or authentic language use is studied via corpus-based studies on

the lexicon. These corpus studies reveal evidence on lexis for both single words and multi-word strings that behave as a single lexeme. Corpus evidence has demonstrated that “very limited number of words do the bulk of the work in language”, and also “words tend to collocate” (Schmitt, 2000: 88).

2.3. Formulaic Language

In spite of the fact that there have been numerous studies about formulaic sequences, a variety of terms were employed to refer to them. In several studies, FSs were adverted beneath diverse names as sentence stems (Pawley and Syder, 1983), multi-word units (Cowie, 1992), prefabs or lexical phrases (Nattinger and DeCarrico, 1992), prefabricated patterns (Granger, 1998), recurrent word combinations (De Cock 1998; Altenberg, 1998), lexical bundles (Biber, Johansson, Leech, Conrad, and Finegan, 1999; Cortes, 2002; Biber and Barbieri, 2007), multiword items (Lewis, 2002), formulas (Jespersen, 1924; Wray, 2002), clusters (Schmitt, Grandage and Adolphs, 2004; Hyland, 2008), formulaic sequences (Schmitt and Carter, 2004; Conklin, and Schmitt, 2008).

The diversity of formulaic sequences is emphasised in several other studies as well (Schmitt and Carter, 2004; Boers et al., 2006; Stengers et al., 2011). Stengers et al. (2011) addressed that FSs are diverse, and include strong collocations, idioms, binomials, standardised similes, proverbs and clichés, and discourse organisers, social routine formulae. FSs “range from simple fillers (e.g., *Sort of*) and functions (e.g., *Excuse me*) over collocations (e.g., *Tell a story*) and idioms (e.g., *Back to square one*) to proverbs (e.g., *Let’s make hay while the sun shines*) and lengthy standardized phrases (e.g., *There is a growing body of evidence that*)” (Boers et al., 2006: 246). It is difficult to identify FSs since they have different properties in terms of length, the purpose of usage, and fixedness. According to Schmitt and Carter’s (2004) formulaic sequences “can be long or short, or anything in between. They are commonly used for different purposes. They can be used to express a message or idea, functions social solidarity, and to transact specific information in a precise and understandable way” (2004: 3).

Wray’s definition of the formulaic sequence asserts that they are the prefabricated word combinations which are “stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (2002: 9). Commenting on formulaic sequence defined by Wray, Xu (2016: 39) states that “the word formulaic is associated with ‘unity’, ‘custom’ and ‘habit’, while sequence indicates that more than one internal unit can be detected, and they do not necessarily have to be words. This definition covers all the possibilities of formulaic linguistic units, thus making reference easier”. Wray’s (2002) working definition on formulaic sequences was employed as a guideline in order to determine which units of language in the essays of learners were likely to be formulaic sequence. There are several studies based on Wray’s definition of formulaic sequences (e.g., Ohlrogge, 2009). According to Duan and Shi (2021), it is

generally agreed that formulaic sequences “are stored in long-term memory and function as single lexical words” (2021: 2), and the following word strings are some of the examples of FSs: *on the other hand, at the same time, it is well known and pay attention to*.

According to Han (2015), there are seven features of FSs that are commonly utilized to find out formulas of spoken and written discourse in L1 and L2. These seven features of formulaic sequences are (2015: 11):

1. varying in grammatical structures (*do homework* vs. *heavy rain*);
2. varying in structural fixedness (*depend on* vs. *pick up*);
3. varying in semantic transparency (*kick the ball* vs. *kick the bucket*);
4. varying in pragmatic function (*I know* vs. *I write*);
5. varying in phonological features (phonological coherence and reduction).
6. varying in frequency counts and association strength in a given corpus or discourse; and
7. being stored and retrieved whole from memory at the time of use and thus psychologically real to native speakers. (Han, 2015:11)

Generally, these multi-dimensional features of formulaic sequences have been referred to by fieldworkers. For instance, the above-mentioned Wray’s definition of FSs has a series of facets that there exist holistic processing and structural fixedness (continuous or discontinuous). Similarly, in the study of Schmitt and Carter (2004), it was indicated formulaic sequences as to whether holistic processing as the field specialists’ focused criteria “which determine whether sequences are known by individual participants, and whether these sequences are formulaic and stored as wholes in the participant’s mental lexicon” (2004: 2). Likewise, ‘institutionalization’, ‘fixedness’ and ‘non-compositionality’ were suggested as some of the criteria of multi-word expressions (Moon, 1997, as cited in Schmitt and Carter, 2004).

Wray (2002: 9) has remarked that “the neutral term formulaic language is too commonly used in the literature to be free of such associations” and he constituted many other terms such as ‘fixed expressions’, ‘prefabricated patterns’, ‘ready-made expressions’ and ‘multi-word units’ for the purpose of specifying aspects of formulaicity.

All things considered, the formulaic sequence is an umbrella term that is “intentionally all-encompassing, covering a wide range of phraseology” (Schmitt and Carter, 2004: 4). As it is discussed above, FSs encompass the variety of terms while they are addressed. In that case, to distinguish certain features in order to define FSs becomes difficult. “Rather it is probably more useful to discuss characteristics which are typical of formulaic sequences, even though every example lexeme might not exhibit each characteristic” (Schmitt and Carter, 2004: 4).

2.3.1. Structural and Functional Characteristics of Formulaic Language

A thorough analysis of the existing literature has revealed that formulaic sequences can be divided into two main categories, namely functional and structural classification. After the n-grams or formulaic sequences have been identified, the investigators have grouped them into their structural and functional categories and register/genre specificity (Greaves and Warren, 2010). The prominent studies in the field of FSs have had these classifications (e.g., Biber et al. 1999; 2004; Hyland, 2008; Conrad and Biber, 2005; Biber and Barbieri, 2007; Cortes, 2004; Simpson-Vlach and Ellis, 2010).

It has been asserted that despite the fact that formulaic sequences are not among of the sorts of the grammatical structures recognized in traditional linguistic theory, the majority of formulaic sequences have “well-defined structural correlates” (Biber et al., 2004: 399) and “strong grammatical correlates” (Cortes, 2004: 400). So, this circumstance helps to categorise them into several primary structural types. Based on these well-defined correlates, the structural classification of formulaic sequences was introduced by Biber et al. (1999) and this classification has been widely employed in many studies for analysing the structures of sequences (e.g., Biber et al., 2004; Cortes, 2004; Cortes, 2006; Chen and Baker, 2010; Salazar, 2011; Adel and Erman, 2012; Fattani, 2018; Bal-Gezegin, 2019). The structural classification three- to four-word sequences in the longitudinal learner corpora followed the taxonomy introduced in the Longman Grammar of Spoken and Written English (LSWE) (in both the conversation and academic prose part) (Biber et al., 1999: 1001-1024) for sub-categories, and Biber et al.’s three major structural types of formulaic sequences provided in 2004 were employed for main categories. That is to say, the researcher reconstituted the final version of the structural categories used in the present study, and it was composed of the mixture of the taxonomies produced in 1999 and 2004. The taxonomy of structural types of formulaic sequences which are created by Biber et al. (2004) mainly embody three categories as follows: it is verb phrase (VP) fragments; it is dependent clause fragments; and it is noun phrase (NP) and prepositional phrase (PP) fragments. In fact, each of these three labels holds several subheadings. For example, in the first type, verb phrase fragments are made up of several sub-categories such as verb phrase (with non-passive verb) (e.g., *is one of the*) and verb phrase (with passive verb) (e.g., *is based on the*). In the second type, dependent clause fragments fall under several sub-categories such as if-clause fragments (e.g., *if you have a*), (verb/adjective+) to-clause fragments (e.g., *to be able to*) and that-clause fragments (e.g., *that there is a*). In the third type, noun phrase and prepositional phrase fragments embrace several sub-categories such as noun phrase with other post-modifier fragment (e.g., *a little bit about*), prepositional phrase expressions (e.g., *at the end of*) and comparative expressions (e.g., *as far as the*).

In addition to the structural classification of formulaic sequences, the investigators also attempted to classify these word sequences functionally. It is a fact that formulaic sequences “represent various categories of meaning and pragmatic characteristics of discourse and conversational structure that exist in many different types of situations” (Nattinger and DeCarrico,

1992: 59). Actually, the functional types shed some light on how the formulaic sequences behaves within the text (Breeze, 2013; Tomankova, 2016). The below mentioned three functional types also may “facilitate accurate understanding and ease fluent discourse production by contributing to the perception of continuity in discourse” (Dontcheva-Navratilova 2012: 41). Based on the meaning and pragmatic characteristics, the functional classification of formulaic sequences was created by Biber et al. (2004) that has been widely used and adapted in many studies for analysing the functions of sequences (e.g., Cortes, 2004; Cortes, 2006; Biber and Barbieri, 2007; Chen and Baker, 2010; Simpson-Vlach and Ellis, 2010; Adel and Erman, 2012; Fattani, 2018; Bal-Gezegin, 2019). The functional types of formulaic sequences which are introduced by Biber et al. (2004) are basically classified into three headings. These are stance expressions, discourse organizers and referential expressions. Stance expressions describe “epistemic evaluations or attitudinal/modality meanings” (Biber and Barbieri, 2007: 270), and these types of expressions are divided into two subheadings, which are: epistemic stance (e.g., *are more likely to*), attitudinal/modality stance (comprise desire (e.g., *if you want to*), obligation/directive (e.g., *you have to do*), intention/prediction (e.g., *I’m not going to*), and ability (e.g., *to be able to*)).

Discourse organizers point “the overall discourse structure and to signal the informational status of statements” (Biber and Barbieri, 2007: 271), and they contain two subheadings as follows: it is topic introduction/focus (e.g., *let’s have a look*); and it has topic elaboration/clarification (e.g., *on the other hand*).

Referential expressions specify “an entity or single out some particular attribute of an entity as especially important” (Biber and Barbieri, 2007: 271), and there are four main subheadings of these expressions. These are identification/focus (e.g., *that’s one of the*), imprecision (e.g., *and things like that*), specification of attributes (comprise quantity specification (e.g., *have a lot of*), tangible framing attributes (e.g., *the size of the*), and intangible framing attributes (e.g., *in terms of the*)), and time/place/text reference (comprise place reference (e.g., *in the United States*), time reference (e.g., *at the same time*), text-deixis (e.g., *shown in figure N*), and multi-functional reference (e.g., *the beginning of the*)).

In essence, the following structural and functional classifications of language formulas which are presented in Tables (Table 1 and Table 2) are created by Biber et al. (1999; 2004) and the current study uses this framework in the classification of formulaic sequences (please see Chapter 4, section 4.2.2. for overall structural and functional taxonomy used in the present study).

Table 1: Structural Classification of Language Formulas

1. Verb phrase fragments	1a. (connector +) 1st/2nd person pronoun + VP fragment	<i>you don't have to, I'm not going to, well I don't know</i>
	1b. (connector +) 3rd person pronoun + VP fragment	<i>it's going to be, that's one of the, and this is a</i>
	1c. Discourse marker + VP fragment	<i>I mean you know, you know it was</i>
	1d. Verb phrase (with non-passive verb)	<i>is going to be, is one of the, have a lot of</i>
	1e. Verb phrase (with passive verb)	<i>is based on the, can be used to, shown in figure</i>
	1f. Yes-no question fragments	<i>are you going to, do you want to, does that make sense</i>
	1g. WH-question fragments	<i>what do you think, how many of you</i>
2. Dependent clause fragment	2a. 1st/2nd person pronoun + dependent clause fragment	<i>I want you to, I don't know if, I don't know why</i>
	2b. WH-clause fragments	<i>what I want to, when we get to</i>
	2c. If-clause fragments	<i>if you want to, if you have a, if we look at</i>
	2d. (verb/adjective+) to-clause fragments	<i>to be able to, to come up with, want to do is</i>
	2e. That-clause fragments	<i>that there is a, that I want to, that this is a</i>
3. Noun phrase and prepositional phrase fragments	3a. (connector+) Noun phrase with of-phrase fragment	<i>one of the things, the end of the, a little bit of</i>
	3b. Noun phrase with other post-modifier fragment	<i>a little bit about, those of you who, the way in which</i>
	3c. Other noun phrase expressions	<i>or something like that, and stuff like that</i>
	3d. Prepositional phrase expressions	<i>at the end of, at the same time</i>
	3e. Comparative expressions	<i>as far as the, greater than or equal</i>

Source: Biber et al., 2004: 381

Table 2: Functional Classification of Language Formulas

1. Stance expressions	A. Epistemic stance	<i>I don't know if, I think it was, the fact that the</i>
	B. Attitudinal/modality stance	
	b1. Desire	<i>I don't want to, if you want to, do you want a</i>
	b2. Obligation/ directive	<i>you have to do, you need to know</i>
	b3. Intention/ Prediction	<i>I'm not going to, are we going to, going to be a</i>
	b4. Ability	<i>to be able to, to come up with, can be used to</i>
2. Discourse organizers	A. Topic introduction/focus	<i>if you look at, let's have a look</i>
	B. Topic elaboration/clarification	<i>I mean you know, on the other hand, as well as</i>

Table 2: (Continue)

3.Referential expressions	A. Identification/ focus	<i>of the things that, is one of the, one of the most</i>
	B. Imprecision	<i>or something like that, and stuff like that</i>
	C. Specification of attributes	
	c1. Quantity specification	<i>have a lot of, how many of you, a little bit of</i>
	c2. Tangible framing attributes	<i>the size of the, in the form of</i>
	c2. Intangible framing attributes	<i>in the case of, in terms of the, on the basis of</i>
	D. Time/ Place/ Text reference	
	d1. Place reference	<i>in the United States, of the United States</i>
	d2. Time reference	<i>at the same time, at the time of</i>
	d3. Text-deixis	<i>shown in figure N, as shown in figure</i>
	d4. Multi-functional reference	<i>the end of the, the beginning of the, the top of</i>

Source: Biber et al., 2004: 384-388

2.4. Significance of Formulaic Sequences and Frequency of FSs

It is claimed that formulaic sequences are widespread in native speakers' language (McCarthy, 1991; Schmitt and Carter, 2004). Similarly, Mel'cuk (1998) asserted that FSs are numerically predominant in language, and they outnumber words roughly ten to one. In addition, FSs, rather than words, are considered as the primary carrier of meaning (Sinclair, 2008). It was found that over half of spoken English discourse - 58.6 percent - and written discourse - 52.3 percent - consist of formulaic sequence and ranges of it (Erman and Warren, 2000). Moreover, formulaic sequences are made up of 25.08 to 32.29 percent of spoken language by native speakers of English, and 16.87 to 17.23 percent of the discourse of non-native speakers (Foster, 2001). Overall, as some of the researchers (Granger, 1998; Howarth, 1998; Foster, 2001) noted, native speakers of a language have more frequent usage of formulaic sequences than non-native speakers. On the other hand, it can be said that the total number of these sequences is more likely to outnumber in language by comparison with single morpheme lexical items (Wong, 2012). Therefore, one of the main issues of research areas in the field of applied linguistics is formulaic sequence studies in the new millennium (Schmitt et al, 2004). It is increasingly evident of the significance of FSs that "convincing explanations of the mechanics of their acquisition must become an essential feature of any model of language acquisition" (Schmitt and Carter, 2004: 14).

If second language learners intend to be successful in writing, a high degree of proficiency in respect to the usage of formulaic sequences is required (Liou and Chen, 2018). According to Martinez and Schmitt (2012), there are four main points emphasising the significance of formulaic language in language use. They are listed the essentialness of formulaic language as follows:

- Formulaic language is ubiquitous in language use;
- Meanings and functions are often realized by formulaic language;
- Formulaic language has processing advantages;
- Formulaic language can improve the overall impression of L2 learners' language production. (Martinez and Schmitt, 2012: 300-301)

The significance of formulaic sequences for L2 becomes apparent under three conditions. The first one is “engaging with native speakers in a genuinely interactive environment” (Wray, 2002: 99). The second one is “the interaction being equal (that is, native speakers being equally motivated to ensure that the non-native speaker understands and reacts to their messages, as the reverse)” (Wray, 2002: 100). The third one is “the non-native speaker being sufficiently confident to pick up and use new forms, even without fully understanding them” (Wray, 2002: 100).

It can be seen that the total number of formulaic sequences takes wide scope in language considering both frequency of occurrence and diversity. In general, the frequency of occurrence of individual words is used as one of the indicators of usefulness in the language (Nation, 2001; O’Keeffe et al., 2007). In the same vein, the relationship between frequency and usefulness may also apply to formulaic language (Martinez, 2011). Most of the items larger than a word act as high-frequency words, and they occur frequently as multi-word units (e.g., *good morning* and *never mind*). The meaning of these sequences is not clear enough to provide a better understanding from the meaning of the single parts of it (e.g., *at once* and *set out*). “If the frequency of such items is high enough to get them into a general service list in direct competition with single words, then perhaps they should be included” (Nation and Waring, 1997: 18). Frequency of occurrence is one of the often-cited criteria. It is assumed that if a sequence occurs frequently in a corpus, it notes that these sequences are conventionalized in language by native speakers, at least to some extent (Schmitt and Carter, 2004). In addition, Wray regards “frequency as a salient, perhaps even a determining, factor in the identification of formulaic sequences” (2002: 25). As noted by Ellis (1997), the fluent usage of FSs by learners is a basic indicator of nativelike proficiency and thus “language learning involves learning sequences of words (frequent collocations, phrases, and idioms) as well as sequences within words” (Ellis, 1997: 130). Similarly, as Kashiha and Chan (2014) point out: a concrete understanding of idiomaticity is achieved by means of frequent use of formulaic sequences. Furthermore, Ellis (2013) submitted that frequency of usage is decisive for learning, memory and perception. It means that the more times it is experienced, the stronger the recall of it, and the more fluently it is attained. Lastly, the relationship between formulaicity and frequency is obvious that when a sequence is used frequently, “that formulaic output is frequently called upon” (Wray and Perkins, 2000: 7). In corpus-based research, either word-focused or sequence-focused can be preferred in an attempt to frequency counts (Wray and Perkins, 2000). It is important to note that the results of the study of Biber and Barbieri (2007) showed that formulaic sequences do not occur together by chance as of corpus frequency analysis, “rather, these word sequences turn out to be consistently functional, indicating

that high frequency is a reflection of pre-fabricated or formulaic status” (2007: 265). Similarly, Conrad and Biber (2005) found that the use of formulaic sequences was not accidental since they have significant discourse functions in the registers.

As noted by various investigators, the importance of FSs for both native and non-native speakers was acknowledged (Schmitt, 2005) and it is asserted that “language production consists of piecing together the ready-made units appropriate for a particular situation and that comprehension relies on knowing which of these patterns to predict in these situations” (Nattinger, 1980: 341). To put it concisely, “language processing is sensitive to formulaicity and collocation” (Ellis, Simpson-Vlach, and Maynard, 2008: 376). Likewise, it is a fact that formulaic sequences are far more than word strings which are tied in compliance with collocational ties (Conklin and Schmitt, 2008). Accordingly, the knowledge of formulaic sequences becomes more of an issue in the writing of L2 learners. It could be suggested that FSs are stored as a single item, and this notion is seen reasonable from both frequency-of-use and psycholinguistic perspective. After all, “retrieving and recognizing such multi-word units would facilitate the level of fluency...” (Conrad and Biber, 2005: 57).

2.5. Acquisition of Formulaic Sequences by L2 Learners

There is considerable evidence from previous studies that formulaic language in both L1 and L2 acquisition plays a significant role (e.g., Weinert, 1995; Wray, 2000; Wood, 2002; Ellis et al. 2008). The process of children’s language acquisition of formulaic language use is based on a usage-based model since they “employ sequences of words taken directly from the input, with productivity arising from the coexistence in a shared representational space of forms that overlap in phonological content or in meaning” (Barnard and Lieven, 2012: 6-7). For usage-based model, frequency is a leading force in the forefront for language acquisition (Wulff, 2019). From the point of usage-based approaches, the units of language are learned from language usage, and the knowledge of language units constitutes the basis of fluent language processing (Ellis and Ogden, 2017). Corpus linguistic evidence of this claim is that “language usage is highly structured and pervaded by collocations and phraseological patterns (...), that is, that lexis, syntax, and semantics are inseparable” (Ellis and Ogden, 2017: 607).

The importance of formulaic language in language acquisition is highlighted by many fieldworkers (Wray, 2000; Ellis, 2002; Wei and Ying, 2011; Millar, 2011). It is the fact that “formulaic language is basic to language development, processing, production and learning” (Wood, 2002: 2). In other words, it is a need to indicate the role of formulaic sequences in language acquisition. Formulaic sequences thanks to their storage as a whole may help to save efforts on language processing. It is proved that formulaic sequences are stored and retrieved as a whole, by this means language processing is relieved of cognitive load (Wei and Yin, 2011). The processing advantage of formulaic sequences is supported in several studies (Pawley and Syder, 1983; Gibbs et

al., 1997; Underwood et al., 2004; Conklin and Schmitt, 2008; Stengers et al, 2011). In the study of Conklin and Schmitt (2008), they investigated processing advantages for formulaic sequences. This study compared reading times of formulaic sequences and non-formulaic equivalents by native and non-native speakers of the language. The findings of the study showed that processing advantage is observed in both L1 and L2 speakers for formulaic sequences over nonformulaic patterns.

In addition to the processing advantage of FSs, it is observed that formulaic sequences play an important role in fluent linguistic production (Wood, 2002; Hyland, 2008; Stengers et al, 2011; Hyland, 2012; Wood, 2010; Davies, 2014). Hyland (2012) remarked that formulaic patterns are widespread in academic language, and they are known as the main component of fluent linguistic production. The use of formulaic sequences between novice and expert usage in written and spoken language can be distinguished. The study of Wood (2006) aimed to investigate the use of formulaic sequences in the development of fluency of language production. The result of this study showed that formulaic sequences have a facilitative role in the development of fluent language production.

As Hill argues: “the largest learning load and the one which is never complete - even for native speakers - is mastering the lexicon” (2000: 68). Cause of the students’ level is mostly intermediate, focusing on the EFL grammar during the lessons leads to slow down or remain the same of the learner proficiency. The fact is that the emphasis on lexis assists learners to improve their competence levels, e.g., intermediate to advance. “It is lexis in general, and collocational competence in particular, which allows students to read more widely, understand more quickly, and speak more fluently” (Hill, 2000: 68). In the same vein, Lewis (2000: 177) declared that “proficiency in a language involves two systems, one formulaic and the other syntactic”, and the former one is referred to as ‘islands of reliability’. Similarly, FSs are called as ‘islands of reliability’ that is Dechert’s (1983) terminology (as cited in Wray, 2002). In brief, it can be said that the use of formulaic sequences assists to perceived as more proficient in language use (Boers et al., 2006). It is believed that mastery of formulaic sequences is beneficial for language users since these sequences may assist to achieve linguistic accuracy. For formulaic sequences establish ‘zones of safety’, the proper use of these patterns “may thus confine the risk of ‘erring’ to the spaces in between the formulaic sequences in one’s discourse” (Boers et al., 2006: 247).

2.6. Corpus Linguistics and Corpus Tools

According to Hidalgo et al. (2007: 32), “a corpus is like a text museum”. The definition of the corpus was remarked by Sinclair that "a collection of naturally-occurring language text, chosen to characterize a state or variety of a language, typically contains many millions of words" (Sinclair, 1991: 171). Corpus is a tool that provides to unearth various dimensions of language use (Reppen, 2010: 31). Conrad’s definition of corpus linguistics is that “the empirical study of language relying on computer-assisted techniques to analyse large, principled databases of naturally occurring

language” (2000: 548). Svartvik stated that “while the manual excerpting of textual data has been the regular means of gathering information for linguistic description, its modern form, which only recently has come to be known by the name of corpus linguistics - the use of large collections of text available in machine-readable form - only dates back to the early 1960s” (1992: 7). Bonelli indicated that “corpus linguistics represents a definite shift towards a linguistics of *parole*; the focus is on ‘performance’ rather than ‘competence’” (2010: 14-15). The purpose of researchers in linguistics is to illustrate language use instead of depicting only common groundworks on linguistic. “The quantitative element (frequency of occurrence) is considered very significant and, depending on the specific approach, is taken to determine the categories of description” (Bonelli, 2010: 15). Conrad suggested that corpus linguistic assists to identify common and uncommon choices; to sum up, it also assists to bring into focus on “the patterns that characterise how a large number of people use the language, rather than basing generalisations on a small set of data or anecdotal evidence or focusing on the accurate/inaccurate dichotomy” (2010: 228). As Lewis (2000) pointed out: it is known that computer corpora and corpus linguistics are seen as substantial tools, and they are ever developing to provide better descriptions of English than they have ever had. To sum up, corpus linguistics assists in research on authentic materials (Allan, 2009). Additionally, it is pointed in the doctoral dissertation of Özbay (2015: 1) that “corpus linguistics presents us with profound changes in the way that we study, teach and learn languages all over the world due to its huge potential to present entirely authentic, genuine, qualitative and quantitative findings related to the nature of language”.

The concerns of representativeness, corpus size and corpus type are regarded as significant issues in corpus linguistics. Biber (1993) stated that “a corpus must be representative in order to be appropriately used as the basis for generalizations concerning a language as a whole” (p. 243). It is not always a fixed criterion across all corpus types, for example, learner corpora “is almost always much more restricted in size as well as type of texts providing their database” (Özbay, 2015: 68). In addition, the size of a corpus may be an influential factor in the notion of representativeness that “the greater the size of corpora, the more representative their nature, the more thorough and more complex analyses” look probable (Biber et al., 1998: 12). It is the fact that ‘the lengthier a corpus, the better’ is a general notion (Biber et al., 1998; Meyer, 2002; Tribble and Jones 1990; Flowerdew 1996; Sinclair 1991). Likewise, according to Sinclair (1991), a corpus requires to include very large number of words. On the other hand, Hunston (2002) suggested that “the feasible size of a corpus is not limited so much by the capacity of a computer to store it, as by the speed and efficiency of the access software” (2002: 25). To sum up, besides very large corpora which may offer large data to create dictionaries and reference books, there are various smaller corpora that may help to conduct a particular research (Hunston, 2002). As it is known, a corpus is designed in compliance with a particular purpose, and this purpose determines the type of the corpus. In her book named ‘Corpora in Applied Linguistics’, Hunston (2002) introduced eight types of corpora as follows: specialized corpora, general corpus, comparable corpora, parallel corpora, learner corpus, pedagogic corpus,

historical and diachronic corpus, and monitor corpus. Among these eight types of corpora, learner corpus which “can be used to identify characteristic patterns in student's writing” (Tognini-Bonelli, 2001: 9) is suitable for the purpose of this thesis. Learner corpora contain a collection of texts produced by language learners that ICLE (International Corpus of Learner English) is the best known among learner corpora, and LOCNESS (Louvain Corpus of Native English Essays) is a comparable corpus as ICLE (Hunston, 2002).

As an applied linguistics approach, corpus linguistics has been assumed one of the foremost methods that examines language today (Anthony, 2013). It is widely accepted that corpus tools are considered as significant because of rapid changing nature of language learning and teaching fields. Generally, the information about frequency of usage and patterning of the texts in corpora is revealed via concordance program software. Computer software tools are of value because they process the corpus and present findings clearly. MonoConc Pro¹ and WordSmith Tools² are among the best eminent in the field. In addition, AntConc⁴ is another toolkit that AntConc is available free online and designed specifically in order to use in the classroom (Anthony, 2004). Furthermore, in order to analyse corpora, there are several programming languages such as Python, Visual Basic, or Perl, as well as software programs. According to Kilgarriff (2005: 264), “language is not random because we speak or write with purposes. We do not, indeed, without computational help are not capable of, producing words or sounds or sentences or documents randomly”; accordingly, it seems to be necessary of a tool so as to perform the analysis of language corpora. Thereupon, Kilgarriff et al. (2004) brought out a tool, namely, Sketch Engine that “a corpus tool which takes as input a corpus of any language (with appropriate linguistic markup), and which then generates, amongst other things, word sketches for the words of that language” (Kilgarriff et al, 2004: 105). This tool serves three core functions, which are: word sketches, concordancing, and thesaurus; as can be seen, in lexicography, Sketch Engine has been run commonly since its developed (Kilgarriff et al, 2014). Nowadays, more sophisticated version of this tool is accessible in order to study on large corpora via several functions.

One of the functions of Sketch Engine is to generate the frequency lists of sequences of tokens, namely n-gram tool which is called FSs. In general, n-grams can be produced by these attributes; word, lemma, tag, lemmas, and part of speech, in fact, word and lemma are the most frequently used to create frequency lists. Lemma which comprises all the forms of words and it is a base word (e.g., *be* is the lemma for *am*, *is*, *are*), that is to say, it can be handle with all instances of words thanks to ‘lemmas’. There are some other filtering options of n-grams. Greaves and Warren stated that “typically, n-grams are grouped together based on the number of words they contain, with the result that two-word n-grams may be referred to as bi-grams, three-word as tri-grams and so on” (2010: 213); indeed, four-word n-grams are inclusive of two and three-word n-grams. Granger and Paquot (2008) stated that “n-gram analysis is a method which allows for the extraction of recurrent continuous sequences of two or more words” (2008: 38).

2.7. Corpus Research on Formulaic Language

The advent of computer technology made it possible to create large corpus, and consequently the number of corpus studies that investigate the use of language increased. Through corpus-based research, many significant findings have emerged from the studies of applied linguistics, and what is more, this type of research allows for the investigation of different linguistic features in both academic written and spoken registers of language users. When it is delved into the literature, it is seen that the majority of the studies tend to focus on the analysis of written language rather than spoken language because the process of constructing written corpora is easier than spoken corpora. For Conrad (1996: 300) states, corpus-based investigations have three common characteristics. These investigations;

- a) are based on principled collections of naturally occurring texts (the corpus),
- b) use computers for both automatic and interactive analyses, and
- c) include both quantitative analyses and functional interpretations in order to describe patterns in language features. (Conrad, 1996: 300)

Table 3 below submits an overview of previous corpus-based studies on formulaic language with different corpora. The core information related to studies such as the aim and findings will be summarised in the following paragraphs.

Table 3: Previous Research on Formulaic Sequences

Author	Year	Participants	Age/Grade/Level of Participants	Themes	Corpus	Corpus Size
Ohlrogge, Aaron	2009	The candidates from 9 first language backgrounds such as Greek (85), Spanish (37) and Portuguese (25)	ages ranged from 13 to 50 years of age	relationship(s) between formulaic language use and L2 writing proficiency	a small corpus of 170 compositions written for English as a Foreign Language (EFL) exam	-
Simpson-Vlach, Rita, and Ellis, Nick C.	2010	-	-	creates an empirically derived, pedagogically useful list of formulaic sequences for academic speech and writing	<i>Target Corpora:</i> MICASE, BNC, Hyland's (2004) research article corpus, and selected BNC files <i>Comparison Corpora:</i> Switchboard (2006) corpus, the FLOB and Frown corpora.	MICASE: 1.7 million words BNC: 431,000 words, Hyland's (2004) research article corpus: 1.2 million words, BNC files: 931,000 words, Switchboard (2006) corpus: 2.9 million words, FLOB and Frown corpora: 1.9 million words
Qi, Yan and Ding, Yanren	2011	Chinese university students in English majors; native speaking college students	age 19 and 22; advanced EFL learners	Use of formulaic sequences in monologues of Chinese EFL learners	Longitudinal Spoken English Corpus of Chinese Learners (LSECCL); American college students' transcriptions of monologues	30,217 word tokens; 6739 word tokens.
Staples, Shelley et al.	2013	ESL writers	low, intermediate, and high	Formulaic sequences and EAP writing development	written responses to items on the TOEFL iBT	249,417 words.
Bestgen, Yves	2017	16 different L1 backgrounds including Spanish, French, etc.	the ages of 16 and 25; intermediate to advanced EFL learners	Formulaic competence, the native-like use of ready-made sequences of words	the First Certificate in English (FCE) examination scripts; ICLE and BNC	460,964 words; 151,448 words
Wang, Ying	2018	L1-English university students in years 3 and 4; the expert writers	university students in years 3 and 4; expert writers	Variability in formulaic sequences with interpersonal functions in L1 novice and expert academic writing	Native Novice Corpus (were drawn from the BAWE corpus); Expert Corpus (is made up of published articles)	Novice Corpus 46,722; Expert Corpus 52,626
Duan, Shiping and Shi, Zhiliang	2021	31 Chinese students of English	advanced learners; between 18 and 20 years old	the use of formulaic sequences (FS) in academic writings	corpus of student academic texts written at five time points between Year 1 and Year 3.	Semester 2 in Year 1: 5267 words; Semester 3 in Year 2: 5938 words; Semester 4 in Year 2: 6989 words; Semester 5 in Year 3: 6582 words; Semester 6 in Year 3: 7575 words

As stated above, Table 3 indicates a list of several corpus studies on formulaic language. In brief, it appears that the findings of the studies emphasize the significance of lexical features in different contexts and themes in language use.

The first study shown in Table 3 is by Ohlrogge (2009) which examines the types of formulaic language, and the relationship between the use of formulaic language and L2 writing proficiency over 170 texts of intermediate proficiency writers, and these productions are written for an EFL proficiency test. In the study, a small corpus of compositions written for a high-stakes English as a Foreign Language (EFL) exam is used. First and foremost, a total of 8 types of formulaic language is found out, and these are; collocations/idioms, phrasal verbs, personal stance markers, transitions, language copied from the prompt, generic rhetorical phrases, and irrelevant biographical information. The results of the study suggest that there are several different relationships between FS use and proficiency level.

The second study, in Table 3, by Simpson-Vlach and Ellis (2010) aims to create an empirically derived and pedagogically useful list of 3-, 4-, and 5-word formulaic sequences for academic speech and writing. They stated that this list comprises FSs “identified as (1) frequent recurrent patterns in corpora of written and spoken language, which (2) occur significantly more often in academic than in non-academic discourse, and (3) inhabit a wide range of academic genres” (2010: 487). They compiled two corpora, included 2.1 million words in each, named as academic speech corpus and academic writing corpus. The academic speech corpus includes MICASE and academic speech files of BNC whereas the academic writing corpus involves Hyland’s (2004) research article corpus and selected files of BNC. For the comparison, the Switchboard corpus and the FLOB and Frown corpora are used. The functional categorisation is applied after the lists are prepared. The findings of Simpson-Vlach and Ellis’ (2010) study suggest that FSs “can be statistically defined and extracted from corpora of academic usage in order to identify those that have both high currency and functional utility” (2010: 508).

The third study shown in Table 3 is by Qi and Ding (2011) which aims to analyse usage patterns of formulaic sequences of university students in prepared monologues during three-year period. In the study, frequency, accuracy and variation are determined as the dimensions of developmental changes. The Longitudinal Spoken English Corpus of Chinese Learners (LSECCL) is used as a learner corpus while the monologues of 15 native speaking college students are used as a native speaker corpus. The findings reveal that students make progress in the use of formulaic sequences across three years. Although there is a significant improvement in terms of variation, there are no significant differences in the sense of frequency and accuracy. Lastly, when they are compared with native speakers, the results show that the non-native learners fall behind the native speakers with regard to frequency and accuracy.

The fourth study shown in Table 3 is by Staples et al. (2013) which investigates the use of FSs by ESL writers across three proficiency levels. A corpus is composed of two written texts on the TOEFL iBT from 480 participants. The findings show that lower-level students use more sequences than higher-level students. Moreover, the results of the functional analysis indicate a similar use of sequences in the sense of stance and discourse organizing across three proficiency levels, and it also displays very few referential expressions in these groups.

The fifth study shown in Table 3 is by Bestgen (2017) which tries to compare the usefulness of formulaic and lexical richness measures when assessing the quality of an L2 learner text. Learner datasets involve the scripts of the First Certificate in English (FCE) examination and the extracted 223 essays the International Corpus of Learner English (ICLE). The British National Corpus (BNC) is used as a reference corpus. The results indicate that “formulaic measures were the best predictors and that they provided a much higher specific contribution to the prediction than single-word lexical measures” (Bestgen, 2017:74).

The sixth study shown in Table 3 is by Wang (2018) which examines the use of interpersonal formulaic sequences that distinguish native novice and expert academic writing. The novice texts of the BAWE corpus and expert corpus that includes the published articles are used in the study. The findings indicate that a wider range of formulaic sequences with interpersonal functions is observed in the texts of novice writers compared to the texts of expert writers.

The last study shown in Table 3 is by Duan and Shi (2021) which aims to explore the use of formulaic sequences in academic writings of EFL learners across different levels of studies. The study has a longitudinal design, and the longitudinal learner corpus is comprised of student academic texts that these texts are written at five time points between first year and third year. The findings of the study show that MI (mutual information), “structure and assumed learners’ proficiency (time points) and their interactions produced significant effects on the development of L2 FS, but function and congruency did not” (2021: 1).

In brief, there are several corpus studies on formulaic sequences in different contexts and purposes (e.g., Buerki, 2016; Grigaliuniene and Jukneviene, 2011). For instance, Buerki’s (2016) study focused on how formulaic sequences fit into a constructionist approach to grammar. In addition, Grigaliuniene and Jukneviene’s (2011) study addressed the issue of formulaic language in the Lithuanian EFL learner speech.

On the other hand, the limited number of studies on FSs had a longitudinal design. Li and Schmitt’s (2009) study is one of the examples of a longitudinal study and they studied on how sequences are acquired and used in the texts of Chinese MA students during an academic writing course. Another example of a longitudinal study on FSs belongs to Schmitt et al. (2004) who

investigated the learning or acquisition of academically-based formulaic sequences over EAP course of the two or three months. In addition, Elturki (2015) conducted the longitudinal study to examine the development of three- to four-word sequences through learner corpus over yearlong written output. In addition to the longitudinal studies, there are several studies on the spoken output. For instance, Adolphs and Durow (2004) investigated the acquisition and the development of usage patterns of formulaic sequences in the spoken output of students over time.

2.8. Native and Non-native Speakers in Applied Linguistics

The contrastive study of language use between native speakers and non-native speakers may contribute to understand the differences in both contexts. It can be said that it is likely to reveal the issues of the language use of non-native speakers thanks to this comparison, and it may provide solutions. So, there is a need of the comparison of native speakers' productions for the purpose of analysis of formulaic sequences that are used by EFL learners.

In his book, *The Native Speaker: Myth and Reality*, Davies (2003) mentioned several aspects in order to discuss native speaker idea as a construct. These components of this construct are: the psycholinguistic aspects, the linguistic aspects, the sociolinguistic aspects, and the communicative competence aspect. Accordingly, Özbay (2015) claimed that "native and non-native speakers of a language will exhibit some variations in the way they use words, grammar rules and sentence structures and so on" (2015: 76). Therefore, the native speaker is considered to be suitable for the comparison when the researchers intend to ascertain the differences of language use between native and non-native groups. In line with this notion, there are several studies comparing formulaic sequences of native speakers and non-native speakers (e.g., Ma, 2009; Chen and Baker, 2010; Adel and Erman, 2012; Karabacak and Qin, 2013; Salazar, 2014; Güngör and Uysal, 2016; Taşkaya, 2019). In line with Özbay's statement, Güngör and Uysal (2016) conducted a study to compare the structural and functional characteristics of sequences used in L1 and L2 research articles in English. They concluded that the majority of sequences used by native English writers were noun phrases and prepositional phrase fragments, namely phrasal structures whereas L1 Turkish writers utilized clausal or verb-phrase structures. Another study conducted by Adel and Erman (2012) investigated usage patterns of sequences in advanced learner writing by comparing native and non-native speakers of English. The study results showed that native learners used more varied sequences than non-native counterparts.

Kashiha and Chan (2015) conducted a study to find out the use of four-word formulaic expressions in classroom discussions between native and non-native speakers. The results of the study showed that native speakers used more formulaic expressions than Malaysian non-native speakers. Likewise, Chen and Baker (2010) carried out a study on recurrent FSs. In their study, these sequences were retrieved from both published academic texts of native speakers and academic

writing texts of L1 and L2 students. The findings of the study revealed that while FSs were used surprisingly similar in the academic writing texts of native and non-native students, these sequences were used widely and more varied in published academic texts. Similarly, these findings were supported by other studies which investigate the use of FSs among native and non-native language users in written language (Ma, 2009; Karabacak and Qin, 2013; Taşkaya, 2019). As studies in the literature suggest that there is a difference between native and non-native speakers of English in terms of language use.



CHAPTER THREE

3. METHODOLOGY

3.1. Introduction

This is a longitudinal study including both quantitative and qualitative data collection methods. The methodology chapter includes the research design of the present thesis. The first section presents the outline of methodology; the second section describes participants' demographic information; the third section involves information about longitudinal learner corpora and native corpus and the fourth section covers analytical procedures of gathered data. The study was conducted on several steps: first, the longitudinal learner corpora were compiled. Then, the formulaic sequences were extracted into two groups across two semesters, and these extracted FSs were categorised in structural and functional terms. Next, individual analysis was administered with 8 learners. Their essays were investigated structurally and functionally, and their unique sequences were determined. In the final stage, the retrospective protocol was carried with purposively selected six learners. The detailed stages of this longitudinal study are elaborated in the following sections. The study by Elturki (2015), who investigated the use of three- to four-word sequences in a longitudinal learner corpus composed of yearlong written data, inspired this current study.

3.2. Participants

This longitudinal study was conducted with freshmen students from Karadeniz Technical University Department of English Language and Literature in Trabzon, Turkey, and all participants were native speakers of Turkish. The total number of participants consisted of 85 ELL students 59,99 % of them were females, (n=51) and 39,99 % of them were males, (n=34) and whose ages range between 18-30. These students studied one-year intensive English preparatory class before their bachelor's degree. For all the participants, the medium of instruction throughout their undergraduate education was English. The selection of the participants was made through purposive sampling method. The present study compared native and non-native written corpora. The data for non-native corpora were gathered during two-semester-observation of Academic Writing Courses in the fall and spring term of the 2018-2019 academic year. Participants received 4 hours of classroom instruction per week, 4 weeks a month. All participants prepared an untimed essay in the beginning of the fall term. Their proficiency levels in English were scaled based on their written outputs which were held in the first week of the fall term. In accordance with the proficiency levels of learners, two

groups of students were determined. Group 1 was composed of lower level learners while Group 2 was composed of higher level learners, as noted Table 4 below. The demographic information of participants was given in the following table (Table 4).

Table 4: Demographic Information of Participants

	Variables	Number	Percent %
Gender	Female	51	%59,99
	Male	34	%39,99
Writing Proficiency Exam	Less than 64	42	%49,41
	64 and above	43	%50,58

3.3. Participant Training

In the current research, participants of longitudinal learner corpora received in-class training for two hours for two consecutive semesters. Participant training was followed by the instructor of academic writing course. Throughout the training period, the instructor met with the students on a weekly basis for 120 minutes during which he taught the basics and nature of academic writing as well as a number of FSs. After each training week, each participant wrote an untimed essay which was produced in accordance with essay topics determined by the instructor. That is, during participant training sessions, the instructions were offered, and data were collected through learners' written outputs. In the following week/s after each training, a lesson was leaved to implement for the feedback session. As for data collection, ten untimed argumentative essays were elicited from each participant. Throughout the training period, the instructor followed the course schedule on the weekly basis. This training plan is submitted in Table 5.

Table 5: Further Training on the Weekly Basis

	Overview of Lessons
Treatment 1 - Essay 1	Getting to know each other. Introduce components of the writing process. Introduction to part of an essay. Process of writing essay. Narrowing down the topic. Writing thesis statement. Assignment of first essay topic. Teacher feedback to essays.
Treatment 2 - Essay 2	Generating ideas for the body and organizing them. Sample essay feedback. Brainstorming. Drafting. Introducing content and lay up design. Outline and detailed plan of argumentation. Introducing content and lay up design. Argument structure. Identification of formulaic language. Examples of various types of formulaic language. Assignment of second topic. Teacher feedback to essays.

Table 5: (Continue)

	Overview of Lessons
Treatment 3 - Essay 3	<p>Argument structure. Getting familiarized with academic vocabulary and phrases. Introduction/Thesis-Claim. Body paragraphs 1, 2 and 3. Conclusion/Result. English academic style and language. Formal style. Cautious writing. Academic vocabulary. Topic sentence, supporting sentence, concluding sentence. Introducing content and lay up design. Assignment of third topic. Teacher feedback to essays.</p>
Treatment 4 - Essay 4	<p>Argument structure. Delivery of the list of useful essay words and phrases. Introduction/thesis-claim. Body paragraphs 1, 2 and 3. Conclusion/result. Word choice, academic word and word combinations in the essays. Verbs, collocations, and logical connectors. Using transitional words in an argumentative essay. Coherence. Cohesion. Unity. Linking paragraphs. Four ways to link paragraphs. Repeating key words or ideas from the thesis statement. Referring to words or ideas from preceding paragraph. Using transitional expressions. Using transitional sentences. Introducing content and lay up design. Assignment of fourth topic. Teacher feedback to essays.</p>
Treatment 5 - Essay 5	<p>Argument structure. Introduction/Thesis-Claim. Body paragraphs 1, 2 and 3. Conclusion/result. Various grammar features in writing. The use of adjective clauses. The use of noun clauses, and adverbial clauses. Reduction and sentence starters. Mechanics such as punctuation, spelling etc. Introducing content and lay up design. Assignment of fifth topic. Teacher feedback to essays.</p>
Treatment 6 - Essay 6	<p>Argument structure. Learning about academic writing conventions and integrating into essay writing. Note-making paraphrasing and summarizing. Combining sources. Using other people's ideas in your writing. Delivery of the list of transition words. Coherence and cohesion. Introducing content and lay up design. Assignment of sixth topic. Teacher feedback to essays.</p>

Table 5: (Continue)

	Overview of Lessons
Treatment 7 - Essay 7	<p>Argument structure. Using transitional words in an argumentative essay. Purpose and content of summary. Stages of summarising. Effective paraphrasing. Paraphrasing techniques. Purpose of references and citation. Main reference system. Use of quotations. Organising the references. Learning about academic writing conventions and integrating into essay writing. Avoiding plagiarism. Using other people's ideas in your writing. Referencing properly. References and quotations. Restatement and repetition. Revising essays and proofreading. Learning academic phrases that can be used in arguments. Delivery of the lists of common connectives. Paraphrase and summary. Introducing content and lay up design. Assignment of seventh topic. Teacher feedback to essays.</p>
Treatment 8 - Essay 8	<p>Stylistic matters. Learning about academic writing conventions and integrating into essay writing. Avoiding plagiarism. Using other people's ideas in your writing. Referencing properly. References and quotations. Restatement and repetition. Revising essays and proofreading. Introducing content and lay up design. Correct paper formatting. Learning how to integrate paraphrases and quotations into an essay. Improving writing style. Assignment of eighth topic. Teacher feedback to essays.</p>
Treatment 9 - Essay 9	<p>Stylistic matters. Learning about academic writing conventions and integrating into essay writing. Avoiding plagiarism. Using other people's ideas in your writing. Referencing properly. References and quotations. Restatement and repetition. Revising essays and proofreading. Correct paper formatting. Introducing content and lay up design. Assignment of ninth topic. Teacher feedback to essays.</p>
Treatment 10 - Essay 10	<p>Stylistic matters. Learning about academic writing conventions and integrating into essay writing. Avoiding plagiarism. Using other people's ideas in your writing. Referencing properly. References and quotations. Restatement and repetition. Correct paper formatting. Revising essays and proofreading. Introducing content and lay up design. Assignment of tenth topic Teacher feedback to essays.</p>

3.4. Corpora

3.4.1. Longitudinal Learner Corpora

This current study includes the comparison of native and non-native corpora in terms of formulaic sequences. Non-native corpora were compiled from the argumentative essays of EFL writers ($n=85$) and include 824 essays written by tertiary level students.

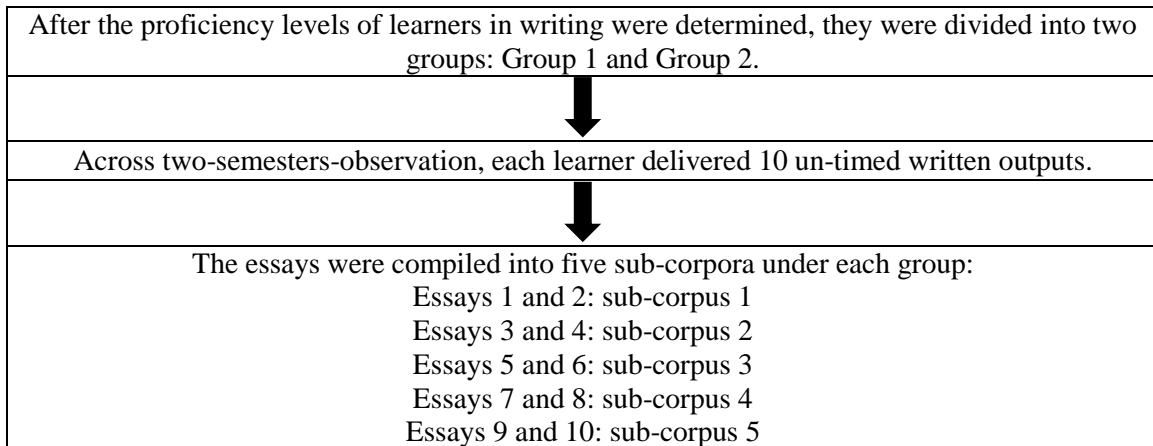
In the study, the longitudinal learner corpus is compiled on a weekly basis for two consecutive semesters (one corpus from at least four weeks of training) observation of students' argumentative essays from 85 participants yielding 824 texts a total of 644,063 tokens. The length of the essays ranges from 248 to 1198 words. During two-semester-observation, each student delivered ten un-timed written assignments. Participants' argumentative essays were delivered as both hard copy and Word documents. Learners sent word documents to the instructor by electronic mail. After then all documents were collected, these word documents were converted into text file documents with AntFile Converter by the researcher.

After the proficiency levels of learners in writing were determined, they were divided into two groups. In each group, these essays were compiled into five sub-corpora as follows; first and second essays were compiled into sub-corpus 1, third and fourth essays into sub-corpus 2, fifth and sixth essays into sub-corpus 3, seventh and eighth essays into sub-corpus 4, and ninth and tenth essays into sub-corpus 5. The profiles of longitudinal learner corpora in two groups are submitted in Table 6. The compilation of the longitudinal learner corpora proceeded through the several steps summarized in Figure 1. The learner corpora consist of in-class writing tasks on ten topics. These topics of essays during weeks are listed (please see Appendix 2).

Table 6: The Profiles of Longitudinal Learner Corpora in Two Groups

	Sub-corpus	<i>N</i> Participants	<i>N</i> Texts	<i>N</i> Tokens	<i>N</i> Words
Group 1	Sub-corpus 1	42	84	47,909	42,582
	Sub-corpus 2	42	81	51,573	45,799
	Sub-corpus 3	42	83	63,780	56,930
	Sub-corpus 4	42	80	72,065	64,632
	Sub-corpus 5	42	77	82,652	73,884
Group 2	Sub-corpus 1	43	85	48,800	43,536
	Sub-corpus 2	43	86	56,748	50,448
	Sub-corpus 3	43	84	65,920	58,772
	Sub-corpus 4	43	83	77,833	69,610
	Sub-corpus 5	43	81	79,612	71,291

Figure 1: The Process of Compiling Longitudinal Learner Corpora



3.4.2. Native Written Corpus: LOCNESS

In the studies which are compared native and non-native corpus, a reference corpus enables to create the norm and standard of native speaker. That is, the comparisons of native speaker use, and learner use demand a control native corpus (Altenberg and Granger, 2001). The native corpus LOCNESS (The Louvain Corpus of Native English Essays), which is taken as the reference and control corpus of the current study, involves 322 texts produced by native speakers of English who were between 17 and 23 years of age. LOCNESS comprises of British pupils' A level essays, British university students' essays and American university students' essays. To prefer a reference corpus that is similar for learner corpora in terms of text types, sizes, participants ages and topic may be significant. LOCNESS as a control native speaker corpus would be representative enough; therefore, it would be considerably comparable to the learner corpora. According to Granger and Tyson (1996), there is a requirement to select a control native speaker corpus that is "exactly the same type of writing" in order to compare with learner writing (1996: 19). The detailed information about native and non-native corpora is provided in the following table (Table 7).

Table 7: The Profiles of Native Written Corpus and Longitudinal Learner Corpora

	Longitudinal Learner Corpora	LOCNESS
Tokens	644,063	360,685
Texts	824	282+
L1	Turkish	American English, British English
Genre	Argumentative	Argumentative

3.5. Analytical Procedures

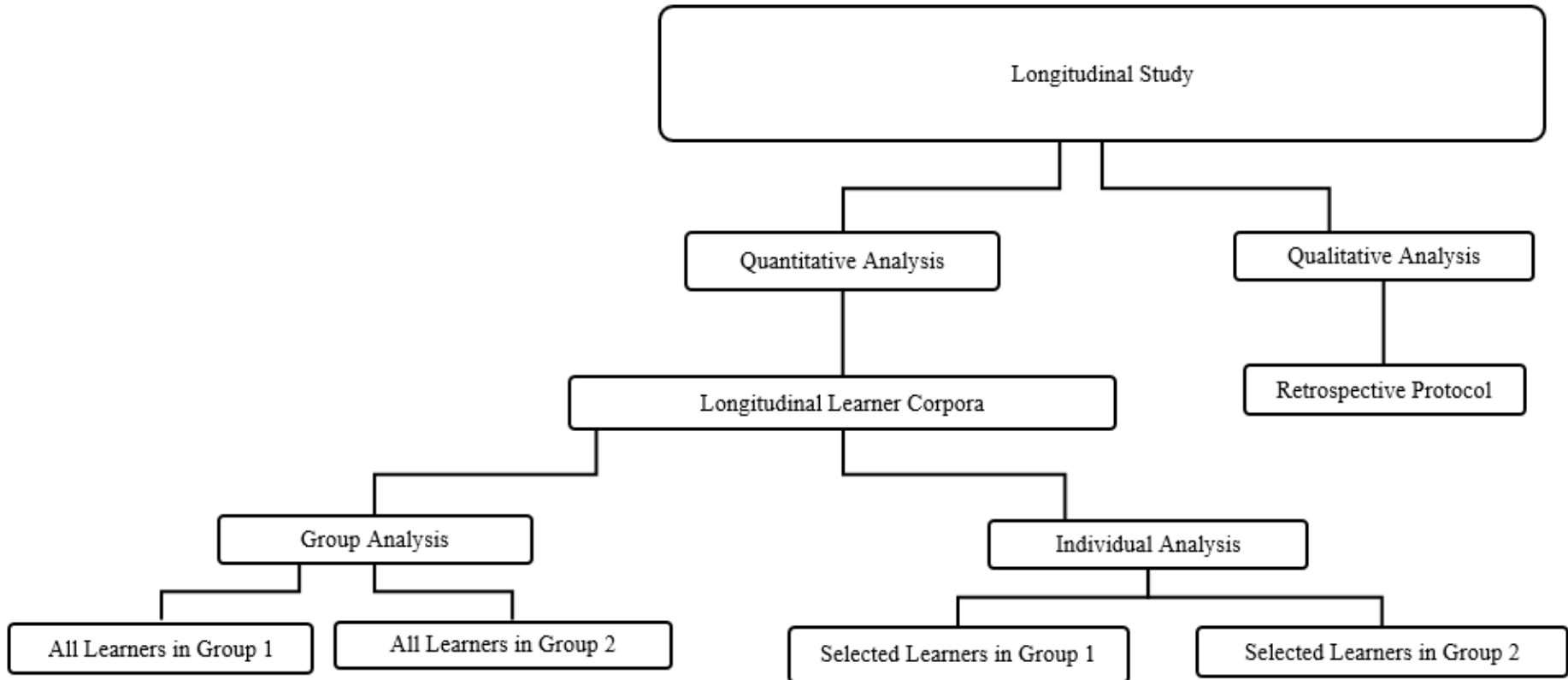
One of the most important developments in corpus studies is the emergence of learner corpora, which are the electronic collections of naturally occurring texts written by ESL and EFL learners

(Granger, 2003). Furthermore; “the release of a learner corpus such as the ICLE marks the beginning of a new stage in the evolution of learner corpus research. The time has come to use the resource on a wider scale in both SLA and ELT” (Granger, 2003: 544). As a component of Corpus linguistics, learner corpora assist “to arrive at conclusions regarding the lexical development of the learners” (Özbay, 2015: 5). The compilation of learner corpora which contain L2 learners’ essays through timed or untimed tasks and its analysis have become the focus of corpus research in recent times. Frequency-based (Nesselhauf, 2004) or the distributional (Evert, 2004) approach is used during the analysis phase of corpus research, especially in the studies of learner corpora. The frequency-based approach was applied to examine learner data in terms of formulaic sequences and to compare between learner corpora and native-speaker reference corpus.

The results of the data analysis are shown in figures and tables. The information related to each table or figure is displayed before the table, and the explanation of each table or figure follows. In Figure 2, the detailed study design is given that quantitative analysis includes corpus-based analysis of the formulaic sequences in the longitudinal learner corpora. On the other hand, qualitative analysis comprised of the retrospective protocol. The results of each data collection procedure are given under Results and Discussion chapter.

In this research, the data analysis of learner corpora was carried out twofold, involving both group analysis and individual analysis on the usage patterns of formulaic sequences of the students’ argumentative essays. As in the nature of longitudinal studies, a group of participants was followed over a period of time. The essays of non-native corpora were examined through the group analysis as the first step of the process. The current study tries to find out collective trends via group analysis. In the second part of the study, individual analysis was done and the development and the use of three- and four-word formulaic sequences were analysed. In individual analysis, six learners from each group were chosen. That design of the study is displayed in Figure 2.

Figure 2: Research Design



At the beginning of the 1st term of 2018-2019 academic year, the participants of the current study wrote an untimed essay which was scored by three instructors with reference to rating criteria as an analytic scoring (please see Appendix 1) adapted from Brooks (2013). “Analytic scoring is a type of rating scale where a candidate’s performance (for example in writing) is analysed in terms of various components (for example organization, grammar, spelling, etc.) and descriptions are given at different levels for each component” (Wang, 2009: 39). In thesis, that rubric settled analytic guidelines to assess how adequate it is at content, organization, vocabulary, language use, and mechanics layers. So as to measure interrater reliability for quantitative data, Shrout and Fleiss (1979) offered that an interclass correlation would be functional on account of its high flexibility. In the present study, an interclass correlation would be functional particularly given that this study has the same number of scores for every participant, and the instructors are the same for all the participants that these are prerequisites to using the interclass correlation to measure interrater reliability. IBM SPSS, version 26 which was released in 2019, analysis software was used to calculate interrater reliability that it is a way with the intention of finding out the degree of consistency over marks to the same set of scripts among different raters. As mentioned before, interrater reliability in order to determine consistency among three raters was calculated via intraclass correlation coefficient. The interrater reliability coefficient among the three raters was 0.70, which represented fair agreement.

3.5.1. Identification of Formulaic Sequences

For the first research question, the frequency analysis was used to examine 3- and 4-word sequences into longitudinal learner corpora for two semesters of language development. In the field of corpus linguistics, many investigators use automated corpus (concordance) tools to determine the number of formulaic sequences to be included in the studies, which are carried out on the basis of frequency counts that these outputs display the number of times these formulaic sequences occur within the texts or more comprehensively within the corpus of texts. In line with these corpus studies, the frequency analysis for three- and four-word formulaic sequences was undertaken with Sketch Engine which is available for free online corpus tool. While the data were examined via online corpus tool, a frequency cut-off point and minimum and maximum length of word strings were incorporated as substantial two measures. As can be seen in studies on FSs that there is no consensus with regards to frequency cut-offs and lengths. The cut-off point was determined at 2 and above which depends on corpus size of learner corpora. Minimum and maximum length of word strings were identified 3- and 4-word sequences. As a first step, three- to four-word sequences were extracted via N-gram function which is the tool to generate sequences in Sketch Engine. After automatic retrieval of 3- to 4-word sequences using Sketch Engine, several guidelines were followed: excluding the context-dependent sequences or topic specific sequences, the treatment of overlapping sequences and the treatment of contractions. Firstly, context-dependent sequences that are combinations containing the topic of the texts such as *depend on technology* and *addict to technology* were detected and removed

from the lists by researcher. The process of excluding the context-dependent sequences was employed by many researchers (e.g., Chen and Baker, 2010; Salazar, 2011; Fattani, 2018). Secondly, some three-word sequences were repeated in four-word combinations, for instance, three-word sequences such as ‘*all over the*’ as well as ‘*over the world*’ occurred in a longer sequence ‘*all over the world*’. The treatment of overlapping was applied on the condition that the ‘complete overlap’ which refers to two three-word sequences which are derived from a single four-word sequences was detected. As given above, ‘*all over the*’ as well as ‘*over the world*’ both occur two times, coming from the longer combinations ‘*all over the world*’. In here, Chen and Baker’s (2010) approach was followed to guard against inflated results that they claimed that four-word sequences “which are actually part of a longer expression and yet, as a result of automatic retrieval, the longer expression is split into two or three shorter units” (Chen and Baker, 2016: 855). Lastly, the treatment of contractions (e.g., *don’t*) as a separate word (e.g., *do not*) was followed. For example, the sequence ‘*they don’t*’ was counted as three words.

3.5.2. Structure and Function of Formulaic Sequences

The second research question aimed to bring out the structure and function of the frequent formulaic sequences in accordance with structural and functional classifications of language formulaic sequences which were submitted by Biber et al. (1999) and Biber et al. (2004). In order to obtain three- to four-word sequences from longitudinal learner corpora, the researcher used N-gram function according to which the researcher can set the frequency cut-off point and minimum and maximum length of word strings in Sketch Engine. Following this, formulaic sequences were identified by the frequency analysis as three-word and four-word sequences’ lists which were generated via Sketch Engine for all groups, and their categorisation in terms of structure and function were created over all levels.

3.5.3. EFLs’ Formulaic Sequences Compared to Native Speaker

The third research question set out shared formulaic sequences between non-native learner corpora and native corpus in order to see how the frequent 3- to 4-word formulaic patterns produced similarly or differently by the learners from those found in native English speakers’ corpora. In other words, this study sought answer to the question of whether there is a relationship between the frequent shared formulaic sequences among each sub-corpus of both groups and LOCNESS. With the intent of comparison, the written corpus LOCNESS was employed. Using frequency analysis, the frequently used top 100-word strings from this corpus were formed. The comparison process of learner corpora and reference corpora required normalization. After extracting three- to four-word sequences lists from Sketch Engine, raw frequencies were converted to normalized frequencies. The normalization of raw frequency was done by dividing raw frequency with total number of words, and then multiplying the result to 50,000 (an arbitrary value). For instance, XXXXX (raw frequency)

/ XXXXX (tokens) x 50,000 = XXXX (normalized frequency). In the process of calculation of normalized frequencies, IBM SPSS Statistics 26 was used. This helped to see what the shared formulaic sequences in the two different learner corpora were. After the normalized frequency was calculated, \log_{10} transformations of each normalized frequency were applied via using SPSS LOG10 function in order to meet the assumption of normality. The transformation of frequency variables to log scores ensures the criterion of normal distribution that correlation coefficient assumes and to provide a linear display. Osborne (2002) stated that “data transformation is a viable option for improving normality of a variable” (2002:1), that is to say, logarithm transformation of variables is applied to satisfy the assumption of normality.

Pearson correlation coefficient using the statistical software package IBM SPSS (version 26) was performed in order to find out whether there is a relationship between the frequent common sequences among in sub-corpus of group one and sub-corpus of group two and native written corpus. The findings in the following chapter report the strength and significance of the relationship between the frequencies of the formulaic sequences produced by EFL learners in each sub-corpus (dependent variables) and their corresponding ones detected in the written reference corpora (i.e., independent variables). This enabled us to see how similar or different the formulaic sequence patterns of EFLs from those of native speaker learners were across the different times of language development. Findings of the correlation test were also presented graphically using scatter plots generated through IBM SPSS 26.

3.5.4. Individual EFLs’ Inventories of Formulaic Sequences

As it is stated earlier, the second part of the process was individual analysis and it focused on the use and development of formulaic sequences across two semesters. For the purpose of individual analysis, eight learners (four learners of group 1 and four learners of group 2) were chosen to analyse the essays of learners in a detailed way. In order to dig up how frequently formulaic sequences were used by individual learners, the most frequent FSs that appeared in these learners’ texts were extracted using Sketch Engine. In the second phase of individual analysis, structural and functional classification were applied. Lastly, the unique patterns in these selected argumentative essays were identified. It is aimed to find out uniqueness of formulaic sequences usage individually. There is a criterion to determine whether a formulaic sequence sample is a unique one for an individual learner. The criterion for unique formulaic sequence is that it must be used by an individual learner, and the unique sequence should not appear in both longitudinal learner corpora and LOCNESS. In Table 8, the demographic information about each of the 8 learners is given.

Table 8: Participants' Demographic Information of Individual Analysis

Group	Learner*	Gender	Age	Grades of First Essay
Group 1	GİZEM	Female	30	61
	ELİF	Female	20	63,33
	ARDA	Male	20	62
	LALE	Female	19	50,33
Group 2	BETÜL	Female	28	79,33
	ALİ	Male	19	81,33
	ASLI	Female	20	79,66
	ZEYNEP	Female	20	64

* All the participants in the individual analysis are given pseudonyms.

3.5.5. EFL Inventories of Formulaic Sequences: Retrospective Protocol

The fifth research question aimed to investigate whether EFL learners become aware of the existence of formulaic sequences in their essays and across two semesters or not. In order to fulfil that aim, retrospective protocols were made with the 6 subjects, and in the process analysis, verbatim protocols were transcribed to identify codes.

As stated before, the retrospective protocol was another data collection method in this longitudinal study. After the individual analysis was conducted, the retrospective protocol was undertaken with some of the participants in individual analysis in two groups to understand the participant's thoughts regarding FSs. In retrospective protocol, there were a total of six learners from both groups. The retrospective protocols were conducted at the spring term of 2018-2019. The participants answered the Turkish version of protocol questions (please see Appendix 5 and Appendix 6). The protocols took about 30 minutes each and lasted a week. The data collection procedure for each protocol followed some steps. Firstly, the list of the common formulaic sequences created by Simpson-Vlach and Ellis (2010) was introduced, and the participants were given their own essays to read and evaluate in terms of formulaic sequence use. In this phase, they read their previous ten essays with formulaic sequence samples that were previously underlined and highlighted on the essays. Lastly, the participants were asked to report the reasons why they used these formulaic sequences in their essays, and also several other questions were asked for further information about their perceptions of difficulties, the importance of word selection, their familiarity, attention, previous exposure and awareness towards formulaic sequences. The retrospective protocol took place on an online platform, namely Zoom which is a cloud-based video and audio-conferencing service. During the sessions, the meetings were recorded to be analysed later. Below is the timeline (Table 9) for each protocol session.

Table 9: Timeline of the Protocol Sessions with Each Sample

	Participants	June 21	June 22	June 23
Group 1	P1	*		
	P2		*	
	P3			*
Group 2	P4	*		
	P5		*	
	P6			*

3.6. Raters' Profile

The raters were non-native of English teachers. Two raters were from a state university, the other one was from a foundation university in Black Sea region. They had all tertiary level teaching experience as well as they had several grades of teaching experiences ranging from English Language Schools to Public Schools. The raters had 22, 12 and 11 years of experience, consecutively. All of the raters had academic writing teaching experience and two of them were from Turkey and one of them was from Iran. Two but three raters held PhD degrees in Applied Linguistics. Furthermore, one of them held a Certificate of English Language Teaching for Adults (CELTA). Finally, 2 of the raters were female and one of them was male. A summary of the demographic information related to the raters is presented in Table 10.

Table 10: Demographic Information of Raters

Variables	Categories	Total
Gender	Female	2
	Male	1
Teaching Experience in Turkey (years)	Lower 15 years	2
	Above 15 years	1
Teaching Experience in Academic Writing Course	Lower 10 years	2
	Above 10 years	1
Degree	PhD	2
	MA	1
Nationality	Turkish	2
	Foreign	1
Qualification	CELTA	1

CHAPTER FOUR

4. RESULTS AND DISCUSSION

4.1. Introduction

The aim of this study was to broaden current knowledge of formulaic sequence development and usage patterns in EFL learners with a focus on longitudinal learner corpora and native corpus; accordingly, the Group and Individual Analysis of the corpora during the analytical phase was carried out. This section reveals the findings of Group Analysis and Individual Analysis, and this process follows the research questions, which comprise; analysis of 3- and 4-word formulaic sequences, structural and functional analysis, comparison of native and non-native formulaic sequences. In addition, EFL inventories of formulaic sequences are displayed by way of retrospective protocol.

4.2. Group Analysis

4.2.1. Analysis of 3- and 4-Word Formulaic Sequences

To introduce the total number of token and type frequency of three- to four-word FSs in both groups meets one of the demands of the first research question, and Table 11 submits these numbers. So as to find out frequently used three- and four-word formulaic sequences in longitudinal learner corpora; in the first phase, the lists of top 10 most repeated FSs which are excerpted from the top 100 frequent three- to four-word FSs were established in Table 12 and Table 13. In the second phase, the lists of shared most frequent 3- to 4-word FSs across both group one and group two are tendered in Table 14 for group one and Table 15 for group two.

Table 11: Total Number of Token and Type Frequency of 3- to 4-Word Sequences in Group 1 and Group 2

3- and 4-word sequence	Sub-corpus 1		Sub-corpus 2		Sub-corpus 3		Sub-corpus 4		Sub-corpus 5	
	tokens	types	tokens	types	tokens	types	tokens	types	tokens	types
Group 1	47,909	3,314	51,573	3,989	63,780	4,812	72,065	7,406	82,652	8,013
Group 2	48,800	4,706	56,748	4,797	65,920	5,136	77,833	7,763	79,612	6,691

Table 12 below illustrates the top 10 frequent formulaic sequences over time in Group 1, and they are listed in order, depending on their normalized frequency.

Table 12: Top 10 Frequent Formulaic Sequences over Time in Group 1

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
one of the	38	39,66	one of the	39	37,81	one of the	49	38,41	one of the	69	47,87	one of the	26	37,41
a lot of	28	29,22	a lot of	36	34,90	a lot of	49	38,41	the most important	52	36,08	according to the	19	27,34
there is a	25	26,09	in terms of	26	25,21	in order to	41	32,14	the end of	39	27,06	in the world	18	25,90
day by day	21	21,92	it is a	23	22,30	in the world	31	24,30	according to the	39	27,06	I believe that	17	24,46
should be given	17	17,74	the number of	23	22,30	because of the	31	24,30	it is not	38	26,37	in the same	16	23,02
to sum up	17	17,74	in order to	22	21,33	in terms of	29	22,73	in the future	34	23,59	it is a	15	21,58
there is no	17	17,74	according to the	21	20,36	the most important	29	22,73	a lot of	34	23,59	as much as	14	20,14
it is not	16	16,70	first of all	20	19,39	it is a	23	18,03	there is a	32	22,20	in terms of	14	20,14
they do not	15	15,65	want to have	19	18,42	first of all	21	16,46	in my opinion	29	20,12	I do not	13	18,71
have to do	14	14,61	is one of the	18	17,45	day by day	19	14,89	the quality of	29	20,12	there is no	13	18,71

As can be seen from Table 12, the most frequent sequence in all sub-corpora was *one of the*. When the list of three- to four-word formulaic sequences was retrieved through Sketch Engine, it was found that the most frequent three- to four-word formulaic sequence was *one of the* in each sub-corpus of Group 1 and this finding is concurrent with the findings of the previous research. When looking at the length of FSs, it was found that nine out of 10 were three-word sequences, with only 1 four-word sequences in sub-corpus 2 of Group 1, whereas all of sequences in sub-corpus 1, 3, 4 and 5 were three-word sequences. Table 12 reveals that the most frequent three- and four-word formulaic sequences in sub-corpus 1 were *one of the*, *a lot of* and *there is a* with the frequencies of 39,66, 29,22 and 26,09, respectively. In sub-corpus 2, *one of the*, *a lot of* and *in terms of* with the frequencies of 37,81, 34,90 and 25,21, respectively, were the frequently occurring FSs. *One of the*, *a lot of* and *in order to* were the most common FSs in sub-corpus 3 with the frequencies 38,41, 38,41 and 32,14, respectively, whereas *one of the*, *the most important* and *the end of* were the most popular FSs in sub-corpus 4 with the following frequencies: 47,87, 36,08 and 27,06, respectively. In sub-corpus 5, *in the world*, *one of the* and *cannot be* were the most recurrent formulaic sequences with the frequencies 90,74, 42,35 and 25,41, respectively. On the other hand, the least frequent formulaic sequences as to be seen in the table were *have to do*, *is one of the*, *day by day*, *the quality of* and *there is no* with a frequency of 14,61, 17,45, 14,89, 20,12 and 19,36, respectively.

Table 13 below demonstrates the top 10 frequent formulaic sequences over time in Group 2, and they are listed in order, depending on their normalized frequency.

Table 13: Top 10 Frequent Formulaic Sequences over Time in Group 2

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
one of the	35	35,86	a lot of	42	37,01	one of the	65	49,30	one of the	84	53,96	in terms of	34	47,14
a lot of	29	29,71	one of the	40	35,24	a lot of	41	31,10	the most important	59	37,90	one of the	34	47,14
there is no	28	28,69	in terms of	40	35,24	in order to	40	30,34	in the future	54	34,69	there is a	28	38,82
they do not	23	23,57	the number of	37	32,60	in the world	30	22,75	it is not	49	31,48	it is not	24	33,28
in order to	22	22,54	in order to	28	24,67	is one of the	28	21,24	the quality of	49	31,48	between +noun and +noun	23	31,89
to sum up	21	21,52	to sum up	24	21,15	to sum up	27	20,48	a lot of	37	23,77	that there is	22	30,50
we do not	21	21,52	in my opinion	23	20,27	one of the most	25	18,96	in my opinion	34	21,84	there is no	22	30,50
is one of the	19	19,47	to go to	23	20,27	in terms of	25	18,96	there is no	34	21,84	according to the	22	30,50
day by day	18	18,44	they do not	21	18,50	of the most	25	18,96	in terms of	33	21,20	should not be	21	29,12
it is not	17	17,42	it is a	20	17,62	according to the	23	17,45	is the most	32	20,56	in the world	21	29,12

It is apparent from Table 13 that the most frequent three- to four-word formulaic sequence was *one of the* in each sub-corpus of Group 2, which was also identified as a frequent formulaic sequence in the literature. This was also in line with the findings of Group 1. When looking at the length of FSs, the table above demonstrated that nine out of 10 were three-word sequences, with only 1 four-word sequences in sub-corpus 1 of Group 2. In sub-corpus 3, eight out of 10 are three-word sequences, with only 2 four-word sequences whereas all of sequences in sub-corpus 2, 4 and 5 were three-word FSs. Table 13 shows that the most frequent three- and four-word formulaic sequences in sub-corpus 1 were *one of the*, *a lot of* and *there is no* with the frequencies of 35,86, 29,71 and 28,69, respectively, and two of the FSs were the same in Group 1 with different frequencies. In sub-corpus 2, *one of the*, *a lot of* and *in terms of* with the frequencies of 37,01, 35,24 and 35,24, respectively, were the frequently occurring FSs, and they were the same in Group 1 with different frequencies. *One of the*, *a lot of* and *in order to* were the most common FSs in sub-corpus 3 with the frequencies 49,30, 31,10 and 30,34, respectively, while *one of the*, *the most important* and *in the future* were the most popular FSs in sub-corpus 4 with the following frequencies: 53,96, 37,90 and 34,69, respectively. In sub-corpus 5, *in the world*, *one of the* and *there is no* were the most recurrent formulaic sequences with the frequencies 67,20, 52,13 and 31,40, respectively. On the other hand, the least frequent formulaic sequences as to be seen in the table were *it is not*, *it is a*, *according to the*, *is the most* and *in order to* with a frequency of 17,42, 17,62, 17,45, 20,56 and 21,98, respectively. These least frequent formulaic sequences in Group 2 were totally different from the least common ones in Group 1.

Table 14 lists most frequent three- to four-word sequences that are shared across 5 sub-corpora of Group 1.

Table 14: Shared Frequent 3- and 4-Word Formulaic Sequences over Time in Group 1

3-4 Formulaic Sequences	Normed Freq Sub-corpus 1	Normed Freq Sub-corpus 2	Normed Freq Sub-corpus 3	Normed Freq Sub-corpus 4	Normed Freq Sub-corpus 5
one of the	39,66	37,81	38,41	47,87	42,35
a lot of	29,22	34,90	38,41	23,59	18,15
according to the	11,48	20,36	14,89	27,06	25,41
the most important	11,48	16,48	22,73	36,08	9,68
in order to	11,48	21,33	32,14	16,65	16,33
in terms of	14,61	25,21	22,73	11,79	16,33
it is a	14,61	22,30	18,03	17,35	18,75
there is a	26,09	15,51	12,54	22,20	14,52
as a result	14,61	17,45	14,89	18,73	19,96
it is not	16,70	9,69	7,84	26,37	24,20
in the world	7,31	11,63	24,30	7,63	90,74
there is no	17,74	11,63	10,19	15,26	19,36
day by day	21,92	15,51	14,89	10,41	8,47
I believe that	9,39	11,63	8,62	15,96	19,36
in my opinion	14,61	10,66	9,41	20,12	13,91
they do not	15,65	10,66	10,19	13,88	13,31
there are many	11,48	16,48	7,06	17,35	16,94
to sum up	17,74	16,48	13,33	9,71	7,86
the fact that	12,52	10,66	14,89	9,71	21,17
that it is	14,61	12,60	14,89	11,79	12,70
of the most	11,48	10,66	13,33	11,79	13,91
one of the most	10,44	9,69	11,76	11,10	13,31

In Table 14, in 5 sub-corpora of Group 1, two of the (*to sum up, day by day*) shared formulaic sequences tend to be less frequent later over time. Some of the formulaic sequences that were less frequent in sub-corpus 1 tended to become more frequent in later sub-corpora such as *one of the, according to the, in order to, it is a* and *in the world*. In addition, the frequency analyses demonstrated a fluctuation in the normalized frequency scores of the majority of shared formulaic sequences. For example, *according to the* was less frequent in sub-corpus 1, increased in sub-corpus 2, and then dropped its frequency in sub-corpus 3, and its fluctuation continued in both sub-corpus 4 and 5. Table 14 demonstrates that *one of the* was the most frequently occurring FSs among shared ones in each sub-corpus with the frequencies of 39,66, 37,81, 38,41, 47,87 and 42,35, respectively. *A lot of* was one of the most popular ones among shared ones in each sub-corpus with the frequencies of 29,22, 34,90, 38,41, 23,59 and 18,15, respectively. The least frequent formulaic sequences as to be seen in the table was *one of the most* in each sub-corpus with the frequencies of 10,44, 9,69, 11,76, 11,10 and 13,31, respectively.

Table 15 lists most frequent three- to four-word sequences that are shared across 5 sub-corpora of Group 2.

Table 15: Shared Frequent 3- and 4-Word Formulaic Sequences over Time in Group 2

3-4 Formulaic Sequences	Normed Freq Sub-corpora 1	Normed Freq Sub-corpora 2	Normed Freq Sub-corpora 3	Normed Freq Sub-corpora 4	Normed Freq Sub-corpora 5
one of the	35,86	35,24	49,30	53,96	52,13
a lot of	29,71	37,01	31,10	23,77	19,47
there is no	28,69	7,93	12,89	21,84	31,40
they do not	23,57	18,50	15,93	14,13	12,56
in order to	22,54	24,67	30,34	11,56	21,98
to sum up	21,52	21,15	20,48	10,28	11,30
is one of the	19,47	13,22	21,24	7,71	13,19
it is not	17,42	9,69	16,69	31,48	28,26
there is a	16,39	15,86	15,93	14,78	30,77
as a result	16,39	13,22	14,41	7,71	16,96
in terms of	16,39	35,24	18,96	21,20	30,77
it is a	16,39	17,62	15,17	16,06	16,96
on the other hand	15,37	11,45	9,10	16,06	13,82
that it is	14,34	7,05	11,38	19,27	15,70
according to the	13,32	14,10	17,45	15,42	23,24
most of the	13,32	14,10	9,86	12,85	6,91
because of the	12,30	15,86	12,89	16,06	20,10
I strongly believe that	12,30	17,62	11,38	8,35	12,56
in my opinion	11,27	20,27	9,86	21,84	15,07
the fact that	9,22	7,05	11,38	16,06	25,75

In Table 15, in 5 sub-corpora of Group 2, two of the (*a lot of, they do not*) shared formulaic sequences tended to be less frequent later over time. Some of the formulaic sequences that were less frequent in sub-corpus 1 tended to become more frequent in later sub-corpora such as *one of the, in terms of, according to the* and *the fact that*. As can be seen from the table, the frequency analyses demonstrated a fluctuation in the normalized frequency scores of the majority of shared formulaic sequences. This was in line with the findings of Group 1. For instance, *a lot of* was less frequent in sub-corpus 1, increased in sub-corpus 2, and then dropped its frequency through sub-corpus 3, 4 and 5. While the sequence that it is was more frequent in sub-corpus 1, there was a steady decrease in sub-corpus 2,

and then increased through sub-corpus 3 and 4. Table 15 illustrates that *one of the* was the most frequently occurring FSs among shared ones in each sub-corpus with the frequencies of 35,86, 35,24, 49,30, 53,96 and 52,13, respectively. *A lot of* was one of the most recurrent FSs among shared ones in each sub-corpus with the frequencies of 29,71, 37,01, 31,10, 23,77 and 19,47, respectively. These two most frequently occurring formulaic sequences were the same with Group 1 with different frequencies. The least frequent formulaic sequences as to be seen in the table was *the fact that* in each sub-corpus with the frequencies of 9,22, 7,05, 11,38, 16,06 and 25,75, respectively.

In short, when it was retrieved and analysed the top 10 shared three- to four-word formulaic sequences in Group 1 and Group 2 through N-gram feature of Sketch Engine, it was found that the most frequent three- to four-word formulaic sequences were *one of the*, *a lot of*, *in terms of*, *it is not*, *in order to*, *the most important*, *there is no*, *there is a*, *according to the* and *it is a*. The top 10 formulaic sequences common in both spoken and written academic language in the Academic Formulas List (AFL) developed by Simpson-Vlach and Ellis (2010) are *in terms of*, *at the same time*, *from the point of view*, *in order to*, *as well as*, *part of the*, *the fact that*, *in other words*, *the point of view of*, and *there is a*. Three sequences (*in terms of*, *in order to* and *there is a*) in the top 10 list were common in this study.

4.2.2. Structural and Functional Analysis

This part evaluates the structure and function of sequences in longitudinal learner corpora by EFLs. With the use of Biber's (Biber et al., 1999; 2004) structural and functional classifications, the forms and functions of three- and four-word formulaic sequences used by L1-Turkish university students were compared. These structural classifications are given in Tables 16, 17 and 18 for both groups. These functional classifications are also given in Tables 19, 20 and 21 for both groups. Moreover, while Figures 5 and 6 outline the distribution of the type-frequency of the structural categories of the top 100 frequent formulaic sequences across each sub-corpus of Group 1 and Group 2, Figures 7 and 8 display the distribution of functional categories of both groups over time.

4.2.2.1. Structural Analysis

The structural classification of the 100 most frequent three- and four-word sequences in the longitudinal learner corpora formulaic sequences list followed the taxonomy introduced in the Longman Grammar of Spoken and Written English (LSWE) (in both the conversation and academic prose part) (Biber et al., 1999: 1001-1024) for sub-categories, and Biber et al.'s three major structural types of formulaic sequences provided in 2004 were employed as the main categories. That is to say, the researcher rearranged the final version of the structural categories employed in this study, and it was composed of the mixture of the taxonomies produced in 1999 and 2004.

Formulaic sequences extracted from each sub-corpus were categorized on their structural characteristics based on the taxonomy introduced earlier in chapter two. In this classification, formulaic sequences were firstly grouped as FSs that incorporate *verb phrase fragments*, *dependent clause fragments*, *noun phrase and prepositional phrase fragments* and *other expressions* that were suggested by Biber et al. (1999) as formulas “that do not fit neatly into any of the other categories” (1999: 1024). Verb phrase fragments have following subcategories such as: *Personal pronoun + verb phrase (+complement-clause fragment)*, *Verb phrase with active verb*, *Anticipatory it +verb phrase/adjective phrase*, *Passive verb + prepositional phrase fragment*, *Copula be + noun phrase/adjective phrase*, *Pronoun/noun phrase + be (+ . . .)*. Dependent clause fragments have the following subcategories such as: FSs with *wh-clause fragments*, *(verb phrase +) that-clause fragment*, *(verb/adjective +) to-clause fragment*, *Adverbial clause fragment*, *If-clause fragments*. Noun phrase and prepositional phrase fragments have the following subcategories such as: *Quantifier expressions*, *Noun phrase with of-phrase fragment*, *Noun phrase with other post-modifier fragment*, *Other noun phrase expressions*, *Prepositional phrase with embedded of-phrase fragment*, *Other prepositional phrase fragment*, *Comparative expressions*. In the category of *other expressions*, there is no subcategory. As mentioned in Chapter 2, this structural framework has been widely employed and applied to formulaic sequences to investigate their basic structural patterns and to understand their characteristics in EFL learners’ essays. Based on these classifications, formulaic sequences identified in each sub-corpus of longitudinal learner corpora were structurally categorized as submitted in Tables 16, 17 and 18 below.

Table 16 illustrates the structure of the top 100 frequently occurring three- to four-word formulaic sequences in sub-corpus 1 and sub-corpus 2 of Group 1 and Group 2. It is clearly seen that some of the sub-categories had no type in the sub-corpus 1 and 2 of longitudinal learner corpora. So, it can be said that these sub-categories were the similar in two groups, and there were also slight differences with regards to the sub-categories.

Table 16: Structure of the Top 100 Frequent Formulaic Sequences in Sub-corpus 1 and Sub-corpus 2 of Group 1 and Group 2

Structural Types	Sub-corpus 1		Sub-corpus 2	
	Group 1	Group 2	Group 1	Group 2
Personal pronoun + verb phrase (+ complement-clause fragment)	<i>we can see, we need to, you want to, we have to, they have to, they want to, I mentioned above, they do not + verb, we do not + verb, they cannot + verb, we cannot + verb, we want to, and we can, you cannot + verb, I believe that,</i>	<i>I strongly believe that, they need to, they have to, we have to, we can see, they do not + verb, we do not + verb, they cannot, we cannot, we do not need, people do not, as you can, you can see, as I said, we can say that, I believe that,</i>	<i>they want to, I strongly believe that, I believe that, you want to, they will be, I firmly believe that, they do not + verb, they cannot + verb, and they can, we need to, we can say, I mentioned above, I think that, we can understand, as I said, as I mentioned, because they have,</i>	<i>I strongly believe that, and they can, I firmly believe that, they can be, we have to, they do not + verb, they cannot, they grow up, as you can, they have to,</i>
Verb phrase with active verb	<i>want to do, have to do, should not be, do not do, do not have, not want to, be aware of, know how to,</i>	<i>know how to, not want to, do not know how, not need to, do not have, cannot do, do not do, should not be, cannot be,</i>	<i>want to have, go to the, want to be, should go to, not want to, do not know, do not want, cannot be, will not be, will be better, learn how to,</i>	<i>want to have, learn how to, do not want to, know how to, do not have, do not know, cannot be, will be better, will not be, have you ever,</i>
FSs with wh-clause fragments	<i>which is given, when they do, who do not + verb, a person who, which are given,</i>	<i>how to use, the reason why, who does not, a person who,</i>	<i>who is a, when they are, who are still,</i>	<i>who is a, who are still too, when they are, while they are,</i>
Quantifier expressions	<i>a lot of, all of the, some of the, most of us,</i>	<i>a lot of, all of us,</i>	<i>a lot of, some of them,</i>	<i>a lot of, some of them,</i>
Noun phrase with of-phrase fragment	<i>one of the, one of the most, the use of, one of the reasons, the field of, the sense of, because of the, because of this,</i>	<i>one of the, one of the most, the use of, of the world, the sense of, the number of, because of the,</i>	<i>one of the, the number of, of the world, the rate of, one of the most, the use of, the importance of, because of the,</i>	<i>one of the, the number of, the importance of, the rate of, one of the most, the use of, of the country, the beginning of, the risk of, the lack of, the development of, one of them, because of the, because of their,</i>
Noun phrase with other post-modifier fragment	<i>the fact that,</i>	<i>the fact that,</i>	<i>the fact that,</i>	<i>the fact that,</i>
Other noun phrase expressions	<i>the most important,</i>	<i>the most important,</i>	<i>due to the, the best way, the most important, very important for,</i>	<i>due to the, the best option, the most important,</i>
Prepositional phrase with embedded of-phrase fragment	<i>in terms of,</i>	<i>in terms of, my point of view, in front of, the end of, as a result of,</i>	<i>in terms of, as a result of,</i>	<i>in terms of, as a result of,</i>
Other prepositional phrase fragment	<i>in my opinion, on the other, in this way, in the first, in this case, on the other hand, in the first place, in the future, in the world, with each other, of the most, most of the, in addition to, in order to, for this reason,</i>	<i>on the other hand, in this way, in my opinion, with each other, in many ways, of the most, most of the, at the same time, in order to, in addition to, in other words,</i>	<i>in order to, in the future, in the world, in my opinion, about this issue, in our country, in recent years, in the past, at the same time, on the other hand, of the most, on the contrary, for this reason,</i>	<i>in our country, on the other hand, in the future, in the world, in my opinion, in the country, in this way, in the past, most of the, of the most, at the same time, of the most important, in order to, on the contrary, in other words, in addition to, for this reason,</i>

Table 16: (Continue)

Structural Types	Sub-corpus 1		Sub-corpus 2	
	Group 1	Group 2	Group 1	Group 2
Anticipatory it + verb phrase/adjective phrase	<i>it is a, it is an, it is not, it should be, it is the, it is important, it is an undeniable,</i>	<i>it is a, it is not, it should be, it is an, as it is, it can be, but it is,</i>	<i>it is a, and it is, it is the, it is not,</i>	<i>it is a, it can be, it should be, and it is, it is not, it is an,</i>
Passive verb + prepositional phrase fragment	<i>are given to, should be given,</i>			
Copula be + noun phrase/adjective phrase	<i>is one of the, is necessary for, is the most, is not a, cannot be, is important for,</i>	<i>is one of the, is the most, be aware of, is the most important</i>	<i>is one of the, is the most, is very important for, is important for, are a lot of, is necessary for, is the best,</i>	<i>is one of the, is very important, is the most, is the best, are the best, is not a,</i>
(verb phrase +) that-clause fragment	<i>that it is, that there are, an undeniable fact that,</i>	<i>that it is, that we can, can say that, that people have, is no doubt that, that they are, that they can,</i>	<i>that it is, that there is, think that the, can say that, they think that,</i>	<i>that it is, can be said that,</i>
(verb/adjective +) to-clause fragment	<i>be able to, to do their, not to be,</i>	<i>be able to, to do their, to have a, have to do, to understand the, to begin with, to use it, to do it,</i>	<i>to have a, to go to, to be a, be able to, are going to, too + adjective + to, too + adjective + to go,</i>	<i>to have a, to go to, be able to, need to be, best way to, to look after, too + adjective + to, too + adjective + to go, still too + adjective + to,</i>
Adverbial clause fragment	<i>day by day, first of all, the same time,</i>	<i>day by day, first of all, in the past,</i>	<i>first of all, day by day,</i>	<i>first of all, day by day,</i>
Pronoun/noun phrase + be (+ . . .)	<i>there is a, there is no, there are many, there are some, this is the, there will be,</i>	<i>there is no, there is a, there is no doubt, there are many,</i>	<i>there is a, there is no, there are many, there are a lot, there are some, there are lots of,</i>	<i>there is a, there is no, there are many,</i>
Other expressions	<i>as well as, he or she, his or her, as a result, according to the, last but not least, thanks to the, as long as, to sum up,</i>	<i>to sum up, his or her, as a result, according to the, as well as, last but not least, after a while, as long as, thanks to the,</i>	<i>he or she, according to the, as a result, last but not least, as well as, such as + noun, his or her, to sum up,</i>	<i>his or her, he or she, as well as, according to the, as a result, last but not least, according to a, such as + noun, to sum up,</i>
If-clause fragments	<i>if they are, if it is,</i>			
Comparative expressions		<i>more and more, as much as,</i>		

In Table 16, while eighteen structural categories of formulaic sequences were produced in sub-corpus 1 of Group 1, seventeen structural categories were found in sub-corpus 1 of Group 2. VP-based structures incorporated 6 sub-categories in sub-corpus 1 of Group 1 while they incorporated 5 sub-categories in sub-corpus 1 of Group 2. As shown in Table 16, the most notable sub-category in the VP-based structure was *personal pronoun + verb phrase (+ complement-clause fragment)*, including 15 types in sub-corpus 1 of Group 1 while including 16 types in sub-corpus 1 of Group 2. This sub-category held the highest number of sequences (e.g., *you want to* and *we do not need*) in this structure. The other sub-categories in the VP-based pattern were *verb phrase with active verb* and *anticipatory it + verb phrase/adjective phrase*, which had similar types in terms of FSs types (8 types) in Group 1. Whereas *verb phrase with active verb* was the third commonly used sub-category, comprising 9 types while *anticipatory it + verb phrase/adjective phrase* was the fourth commonly used sub-category, which comprised of 7 types in Group 2. The sequences included *do not have* and *it is a*, respectively. Another two sub-categories in the VP-based pattern were verb phrase with *copula be + noun phrase/adjective phrase* and *pronoun/noun phrase + be (+ . . .)*, which had similar numbers in terms of FSs types (6 types) in Group 1 while 4 types in Group 2. Regarding these categories, *is one of the* and *there is a* were examples found within sub-corpus 1 of Group 1 and Group 2. *Passive verb + prepositional phrase fragment* was the least common sub-structural categories in Group 1 whereas there was no sequence in this sub-category in Group 2.

Noun phrase and prepositional phrase fragments incorporated 6 sub-categories in sub-corpus 1 of Group 1 while they incorporated 7 sub-categories sub-corpus 1 of Group 2. As shown in Table 16 that the most notable sub-category in the NPPP-based structure was *other prepositional phrase fragment* which included a variety of examples with different frequencies and ranges, including 15 types in sub-corpus 1 of Group 1 while including 11 types in sub-corpus 1 of Group 2. An example was the sequence *on the other hand*. *Noun phrase with of-phrase fragment* (e.g., *one of the* and *because of the*) was the second commonly used sub-category, including 8 types in Group 1 while 7 types in Group 2. It is important to demonstrate that the other sub-categories were in the low proportions of used sequences.

Dependent clause fragments incorporated 5 sub-categories in sub-corpus 1 of Group 1 while they incorporated 4 sub-categories sub-corpus 1 of Group 2. In the sub-categories of dependent clause fragments, FSs *with wh-clause fragments* comprised of a larger proportion of sequence types than all the other sub-categories of *dependent clause fragments* category in Group 1 and the other three sub-categories are *(verb phrase +) that-clause fragment*, *(verb/adjective +) to-clause fragment* and *adverbial clause fragment*, which had similar proportions in terms of FSs types (3 types) in Group 1. In Group 2, *(verb phrase +) that-clause fragment* (including 7 types) and *(verb/adjective +) to-clause fragment* (including 8 types) comprised of a larger proportion of sequence types than all the other sub-categories of *dependent clause fragments* category.

Other expressions included the lowest overall proportion of formulaic sequences compared to VP-based, DC-based and NPPP-based structures, including 9 types in the sub-corpus 1 of both groups.

For sub-corpus 2, in Table 16, while eighteen structural categories of formulaic sequences were produced in sub-corpus 2 of Group 1, sixteen structural categories were found in sub-corpus 2 of Group 2. VP-based structures incorporated five sub-categories in sub-corpus 2 of both groups. As shown in Table 16 that the most attention-grabbing sub-category in the VP-based structure was *personal pronoun + verb phrase (+ complement-clause fragment)*, including 17 types in sub-corpus 2 of Group 1 while including 10 types in sub-corpus 2 of Group 2. The other sub-category in the VP-based pattern was *verb phrase with active verb*, which comprised of 11 types in Group 1 and this category comprised of 10 types in Group 2. Whereas *copula be + noun phrase/adjective phrase* was the third commonly used sub-category which comprised of 7 types, *pronoun/noun phrase + be (+ . . .)* was the fourth commonly used sub-category, which comprised of 6 types in sub-corpus 1 of Group 1. In group 2, *anticipatory it + verb phrase/adjective phrase* and *copula be + noun phrase/adjective phrase* had similar proportions in terms of formulaic sequence types. It is important to remark that there was no sequence in *passive verb + prepositional phrase fragment* category.

Noun phrase and prepositional phrase fragments incorporated 6 sub-categories in sub-corpus 2 of both groups. As given in Table 16 that the sub-category *other prepositional phrase fragment* was made up of the most commonly used one, including 13 types in sub-corpus 2 of Group 1 while including 17 types in sub-corpus 2 of Group 2. The subcategory *noun phrase with of-phrase fragment* which comprised of 8 types is the second mostly used one in sub-corpus 2 of Group 1 while this category comprised of 14 types in sub-corpus 2 of Group 2. *Quantifier expressions, noun phrase with other post-modifier fragment* and *prepositional phrase with embedded of-phrase fragment* were in the low proportions of used sequences in both groups.

Dependent clause fragments incorporated 4 sub-categories in sub-corpus 2 of both groups. In the sub-categories of dependent clause fragments, *(verb/adjective +) to-clause fragment* comprised of a larger proportion of sequence types than all the other sub-categories of *dependent clause fragments* category in both groups. The shared examples in both groups were the sequence *to have a* and *to go to*. Whereas *(verb phrase +) that-clause fragment* was the second commonly used one in Group 1, *wh-clause fragment* was the second commonly used one in Group 2. The last sub-category was *adverbial clause fragment*, which had similar proportions (2 types) in both groups.

Similar to the sub-corpus 1, *other expressions* included the lowest overall proportion of formulaic sequences compared to VP-based, DC-based and NPPP-based structures, including 8 types in Group 1 and 9 types of sequence types in Group 2.

Table 17 presents the structure of the top 100 frequently occurring three- to four-word formulaic sequences in sub-corpus 3 and sub-corpus 4 of Group 1 and Group 2. Similar findings were observed in these sub-corpora that some of the sub-categories had no type in the sub-corpus 3 and 4 of longitudinal learner corpora. Thus, it can be asserted that there were also slight differences with regards to the sub-categories of two groups.



Table 17: Structure of the Top 100 Frequent Formulaic Sequences in Sub-corpus 3 and Sub-corpus 4 of Group 1 and Group 2

Structural Types	Sub-corpus 3		Sub-corpus 4	
	Group 1	Group 2	Group 1	Group 2
Personal pronoun + verb phrase (+ complement-clause fragment)	<i>they need to, we need to, we have to, they want to, I mentioned above, we are all, I firmly believe that, they do not + verb, they are not, I believe that, I strongly believe that, I think that, when we look at, they will be, as I mentioned above,</i>	<i>I strongly believe that, I firmly believe that, we need to, they need to, they do not + verb, they cannot + verb, they are not,</i>	<i>I do not agree, they want to, we need to, I can say that, but I think, I do not + verb, they do not + verb, we cannot, I cannot, we do not + verb, I do feel that, I have several issues, I think it, person cannot find, I agree with, while I do feel, this issue may seem, I believe that, I firmly believe that, he thinks that, he says that,</i>	<i>he said that, I believe that, I do not agree, he fails to mention, I firmly believe that, I think that, I strongly believe that, he claims that, they want to, he thinks that, we look at, I do feel that, I do not + verb, they do not + verb, we cannot + verb, people do not + verb, we do not + verb, we should not + verb, he says that, as I mentioned,</i>
Verb phrase with active verb	<i>do not know, should not be,</i>	<i>not want to, make it possible, know how to, will be more, do not have, do not want, do not know, should not be,</i>	<i>look at the, do not agree with, fails to mention, not find anything to, going to be, cannot find anything, find anything to support, does not have, will not be, does not mean, agree with him, is looking hard enough, do not know, may seem as,</i>	<i>look at the, do not agree with, has failed to, will be a, cannot be, does not mean, should not be, may seem as,</i>
FSs with wh-clause fragments	<i>people who are, when they are, when I was, the people who, people who have, no matter how, what to do,</i>	<i>who is a, who live in, which is a, what is the, when I was, people who are, who do not,</i>	<i>who is a</i>	<i>who is a, when it comes to, while I do feel,</i>
Quantifier expressions	<i>a lot of, most of their, some of them, all of the,</i>	<i>a lot of, so many people, a lot of people, most of their, some of them,</i>	<i>a lot of,</i>	<i>a lot of,</i>
Noun phrase with of-phrase fragment	<i>one of the, of the world, one of the most, the importance of, the development of, the benefits of, a part of, of the people, because of the,</i>	<i>one of the, of the world, one of the most, the rules of, of the country, the number of, the importance of, point of view, a part of, the development of, one of these, one of them, all of them, because of the,</i>	<i>one of the, the end of, the development of, the end of the, one of the most, of the world, the field of, the quality of,</i>	<i>one of the, the end of, the development of, one of the most, point of view, of the people, the quality of, because of the,</i>
Noun phrase with other post-modifier fragment	<i>the fact that,</i>	<i>the fact that,</i>	<i>the fact that,</i>	<i>the fact that,</i>
Other noun phrase expressions	<i>the most important, due to the,</i>	<i>the most important,</i>	<i>the most important, the most important thing,</i>	<i>the most important, the most important criteria,</i>
Prepositional phrase with embedded of-phrase fragment	<i>in terms of, in front of,</i>	<i>in terms of, as a result of, in front of,</i>	<i>in terms of,</i>	<i>in terms of,</i>

Table 17: (Continue)

Structural Types	Sub-corpus 3		Sub-corpus 4	
	Group 1	Group 2	Group 1	Group 2
Other prepositional phrase fragment	<i>in the world, with each other, in the same, in the country, in our country, in my opinion, in this way, for people to, in this case, around the world, in the past, in the future, of the most, most of the, in order to, in addition to, for this reason,</i>	<i>In the world, with each other, in the same, in the country, in my opinion, in this way, on the other hand, all over the world, of the most, most of the, at the same time, of the most important, in order to, in addition to, for this reason,</i>	<i>in my opinion, in the future, thoughts about this issue, his thoughts about, on the other, about this topic, at first appearance, in the world, on the other hand, of the most, most of the, in order to,</i>	<i>in my opinion, in the future, in the world, about this issue, at first appearance, in the field of, his thoughts about, of the most, most of the, in many ways, on the other hand, in order to, in addition to, on the contrary, for this reason,</i>
Anticipatory it + verb phrase/adjective phrase	<i>it is a, it is not, it is important, it can be, it is the, but it is, it is very, it is important to, and it is,</i>	<i>it is a, it should be, it is important, it is not, it can be, it is a fact, it is very, it will be,</i>	<i>it is not, it is a, and it is, it would be, it can be, it will be, think it is, but it is,</i>	<i>it is not, it is a, it will be, it can be, it does not, because it is, it should be,</i>
Passive verb + prepositional phrase fragment	<i>can be seen</i>		<i>based on my,</i>	<i>based on my,</i>
Copula be + noun phrase/adjective phrase	<i>is one of the, is very important, is the most, is important to, is important for, are a lot of, is good for,</i>	<i>is one of the, is a very, is the most, is not only, is important for, is a fact that,</i>	<i>is one of the, is the most, is not a, is the most important, is going to, cannot be,</i>	<i>is one of the, is the most, is not a, is the most important, is going to, strongest supporting claims is, is not the,</i>
(verb phrase +) that-clause fragment	<i>that it is, that they have, of the fact that,</i>	<i>that it is, that they are,</i>	<i>that it is, that there is, to say that, can say that, fails to mention that,</i>	<i>that it is, that there is, think that the, fails to mention that, people think that, that it will, can say that, to say that, does not mean that, is that the,</i>
(verb/adjective +) to-clause fragment	<i>to be a, to go out,</i>	<i>to each other, to create a, be able to, need to be, to be a, to have a, to make people,</i>	<i>be able to, to buy a, will be able to,</i>	<i>be able to, to be a, to look at,</i>
Adverbial clause fragment	<i>day by day, first of all,</i>	<i>first of all, in the past, in the future, day by day,</i>	<i>day by day,</i>	<i>for a long time, day by day,</i>
Pronoun/noun phrase + be (+ . . .)	<i>there is a, there is no, and they are, of them are, there are many, there are some, there are a lot,</i>	<i>there is a, there is no, there will be, there are many,</i>	<i>there is a, there is no, there are many, there are some, this is a, this is the, because they are, because it is,</i>	<i>there is a, there is no, this is a, there will be,</i>
Other expressions	<i>as a result, according to the, last but not least, according to a, according to their, as well as, as long as, to sum up,</i>	<i>as a result, according to the, according to a, as well as, last but not least, he or she, all in all, to sum up,</i>	<i>as a result, according to the, according to my, to sum up,</i>	<i>according to a, according to the, according to my, as a result, he or she, to sum up,</i>
If-clause fragments	<i>if we want,</i>		<i>if someone is, if it is,</i>	<i>if it is</i>
Comparative expressions	<i>as much as,</i>		<i>more important than,</i>	

In Table 17, while nineteen structural categories of formulaic sequences were produced in sub-corpus 3 of Group 1, sixteen structural categories were found in sub-corpus 3 of Group 2. Table 17 shows that in terms of n-gram types, in the sub-category of VP-based *personal pronoun + verb phrase (+ complement-clause fragment)* was the most common structure, including 15 types in sub-corpus 3 of Group 1 whereas this category was the second commonly used one (including 7 types) in sub-corpus 3 of Group 2. It is clear that the sub-categories *verb phrase with active verb* and *anticipatory it + verb phrase/adjective phrase* were the most common structure, which have similar percentages (8 types) in sub-corpus 3 of Group 2. The sequences included *do not have* and *it is very*, respectively.

Noun phrase and prepositional phrase fragments incorporated 7 sub-categories in sub-corpus 3 of Group 1 while 6 sub-categories in sub-corpus 3 of Group 2. As presented in Table 16 that the sub-category *other prepositional phrase fragment* was made up of the most commonly used one, including 17 types in sub-corpus 3 of Group 1 while including 15 types in sub-corpus 3 of Group 2. Examples of these prepositional phrases were: *in the world*, *with each other*, *in the same* and *in the country*. As sub-corpus 2 of both groups, the subcategory *noun phrase with of-phrase fragment* which comprised of 9 types was the second mostly used one in sub-corpus 3 of Group 1 while this category comprised of 14 types in sub-corpus 3 of Group 2. In the same vein with the sub-corpus 1 and 2, *quantifier expressions*, *noun phrase with other post-modifier fragment*, *other noun phrase expressions* and *prepositional phrase with embedded of-phrase fragment* were in the low proportions of used sequences in both groups.

Dependent clause fragments incorporated 5 sub-categories in sub-corpus 3 of Group 1 while 4 sub-categories in sub-corpus 3 of Group 2. In the sub-categories of dependent clause fragments, while FSs with *wh-clause fragments* comprised of a larger proportion of sequence types than all the other sub-categories of *dependent clause fragments* category in Group1, with *wh-clause fragments* and *(verb/adjective +) to-clause fragment* have similar proportions (7 types) and examples were expressions such as *who is a*, *the people who* and *be able to*. Whereas *(verb phrase +) that-clause fragment and adverbial clause fragment* were in the lower proportions in both groups. The last sub-category was *if-clause fragments* in Group 1, including 1 type while there was no sequence in this sub-category of Group 2.

Similar to the sub-corpus 1 and 2, *other expressions* included the lowest overall proportion of formulaic sequences compared to VP-based, DC-based and NPPP-based structures, including 8 types in both groups.

For sub-corpus 4, in Table 17, while nineteen structural categories of formulaic sequences were produced in sub-corpus 4 of Group 1, eighteen structural categories were found in sub-corpus 4 of Group 2. Table 17 demonstrates that in terms of n-gram types, in the sub-category of VP-based *personal pronoun + verb phrase (+ complement-clause fragment)* was the most common structure, including 21 types in sub-corpus 4 of Group 1 while 20 types in sub-corpus 4 in Group 2. Examples of this sub-

category were: *I agree with*, *I have several issues*, *I believe that* and *I do not agree*. The sub-category *verb phrase with active verb* was the second most commonly used one, including 14 types in sub-corpus 4 of Group 1 and 8 types in sub-corpus 4 in Group 2. In the Group 2, *verb phrase with active verb* and *pronoun/noun phrase + be (+. . .)* had similar proportions (8 types). The other sub-categories were in the lower proportions. Actually, the least used one was *passive verb + prepositional phrase fragment*, including 1 type and an example is *based on my*.

In the same vein with sub-corpus 3, *noun phrase and prepositional phrase fragments* incorporated 7 sub-categories in sub-corpus 4 of Group 1 while 6 sub-categories in sub-corpus 4 of Group 2. As shown in Table 17 that the sub-category *other prepositional phrase fragment* was made up of the most commonly used one, including 12 types in sub-corpus 4 of Group 1 while including 15 types in sub-corpus 4 of Group 2. Examples of these prepositional phrases were: *in my opinion*, *in the future*, *in the world*, *about this issue* and *at first appearance*. As sub-corpus 2 and 3 of both groups, the subcategory *noun phrase with of-phrase fragment* which comprised of 8 types was the second mostly used one in sub-corpus 4 of both groups. In the same vein with the sub-corpus 1, 2 and 3, *quantifier expressions*, *noun phrase with other post-modifier fragment*, *other noun phrase expressions* and *prepositional phrase with embedded of-phrase fragment* were in the low proportions of used sequences in both groups. Whereas the sub-category *comparative expression* appeared with the lowest proportion in Group 1, it did not place in Group 2.

Dependent clause fragments incorporated 5 sub-categories in sub-corpus 4 of both groups. In the sub-categories of dependent clause fragments, it is important to remark that *(verb phrase +) that-clause fragment* appeared first time as the most commonly used one in this category in both groups. This category comprised of 5 types in group 1 and 10 types in Group 2 and the examples were the expressions such as *to say that*, *can say that*, *fails to mention that* and *people think that*. With *wh-clause fragments*, *(verb/adjective +) to-clause fragment*, *adverbial clause fragment* and *if-clause fragments* were in the lower proportions of sequence types in both groups.

Similar to the sub-corpus 1, 2 and 3, *other expressions* included the lowest overall proportion of formulaic sequences compared to VP-based, DC-based and NPPP-based structures, including 4 types in Group 1 and 6 types in Group 2.

Table 18 below illustrates the structure of the top 100 frequently occurring three- to four-word formulaic sequences in sub-corpus 5 of Group 1 and Group 2. Similar results were noticed in this sub-corpus that some of the sub-categories had no type in the sub-corpus 5 of longitudinal learner corpora, and there were also slight differences with regards to the sub-categories of two groups.

Table 18: Structure of the Top 100 Frequent Formulaic Sequences in Sub-corpus 5 of Group 1 and Group 2

Structural Types	Group 1	Group 2
Personal pronoun + verb phrase (+ complement-clause fragment)	<i>I do not agree, I believe that, I firmly believe that, he says that, and they can, I do not, they do not, they cannot, they are not, I think that, he claims that, I cannot, we look at, when we look,</i>	<i>I do not agree, I believe that, I firmly believe that, he says that, he fails to mention, I think that, I agree with, we can see, he thinks that, I strongly believe that, they do not, I do not, they cannot, they are not, I cannot, they want to, we cannot, we do not, and they are,</i>
Verb phrase with active verb	<i>do not agree with, should not be, do not have, will not be, cannot find, do not think, look at the, people do not, cannot be,</i>	<i>have the same, do not agree with, should not be, cannot be, does not mean, not want to, can be a, do not have, do not want to, will be a, will not be,</i>
FSSs with wh-clause fragments	<i>who is a, people who are,</i>	<i>who is a, who is the, when it comes to,</i>
Quantifier expressions	<i>a lot of, some of the,</i>	<i>a lot of,</i>
Noun phrase with of-phrase fragment	<i>one of the, the number of, one of the most, the problem of, because of the, the end of, the rate of, the use of,</i>	<i>one of the, the number of, of the world, because of their, because of the, of the most, one of the most, the problem of, the rate of,</i>
Noun phrase with other post-modifier fragment	<i>the fact that,</i>	<i>the fact that,</i>
Other noun phrase expressions	<i>the most important,</i>	<i>the most important,</i>
Prepositional phrase with embedded of-phrase fragment	<i>in terms of, as a result of,</i>	<i>in terms of, as a result of,</i>
Other prepositional phrase fragment	<i>in the world, about this issue, on the other hand, in the same, around the world, about this topic, in this way, at the same time, of the most, in my opinion, in order to, for this reason, the reason for, in addition to, in his article, in the future, in the work, most of the, of the world, on the contrary, people in the,</i>	<i>in the world, about this issue, on the other hand, in the same, around the world, most of the, in my opinion, on the contrary, in order to, in other words, over the world,</i>
Anticipatory it + verb phrase/adjective phrase	<i>it is a, and it is, it is not, it can be, because it is, it does not,</i>	<i>it is a, it is not, but it is, and it is, it can be, it does not,</i>
Passive verb + prepositional phrase fragment	<i>based on my,</i>	<i>based on my, was published in,</i>
Copula be + noun phrase/adjective phrase	<i>is one of, is not a, be the answer, is not the,</i>	<i>is one of the, is not a, is not the,</i>
(verb phrase +) that-clause fragment	<i>that it is, that there are, that they are, that there is, can say that, is that the, claims is that, to say that,</i>	<i>that it is, that there is, fails to mention that, that they are, that there is a, that there is no, think that the,</i>
(verb/adjective +) to-clause fragment	<i>to be a, be able to, a solution to the, the answer to the, the only way to, to increase the,</i>	<i>need to be, to be a, a solution to the, be a solution to, be able to, to have a,</i>
Adverbial clause fragment	<i>day by day,</i>	<i>at the same time, day by day, in the future,</i>
Pronoun/noun phrase + be (+ . . .)	<i>there is a, there is no, there are many, because they are, this is a,</i>	<i>there is a, there is no, there are many, this is a, because they are, there are some,</i>
Other expressions	<i>as well as, as a result, according to the, according to a, to sum up, due to the, thanks to the,</i>	<i>as well as, as a result, according to the, according to a, to sum up, according to my, according to his, as long as, in this way,</i>
Comparative expressions	<i>as much as,</i>	

In Table 18, while eighteen structural categories of formulaic sequences were produced in sub-corpus 5 of Group 1, seventeen structural categories were found in sub-corpus 5 of Group 2. Table 18 represents that in terms of n-gram types, as each of the sub-corpus of both groups in the sub-category of VP-based *personal pronoun + verb phrase (+ complement-clause fragment)* was the most common structure, including 14 types in sub-corpus 5 of Group 1 while 19 types in sub-corpus 5 in Group 2. Examples of this sub-category were: *I do not agree, I believe that, I firmly believe that, he says that, he fails to mention, I think that, I agree with, he claims that* and *we can see*. As sub-corpus 1, 2 and 4, except for sub-corpus 3, the sub-category *verb phrase with active verb* was the second most commonly used one, including 8 types in sub-corpus 5 of Group 1 and 11 types in sub-corpus 5 in Group 2 and the examples were the patterns such as *have the same, do the same, do not agree with* and *cannot get*. The sub-category *anticipatory it + verb phrase/adjective phrase* had similar proportions (6 types) in both groups. The other sub-categories were in the lower proportions.

In the same vein with sub-corpus 3 and 4, *noun phrase and prepositional phrase fragments* incorporated 7 sub-categories in sub-corpus 5 of Group 1 while 6 sub-categories in sub-corpus 5 of Group 2. As given in Table 18 that as each of the sub-corpus of both groups, the sub-category *other prepositional phrase fragment* comprised the most commonly used one, including 21 types in sub-corpus 5 of Group 1 while including 11 types in sub-corpus 5 of Group 2. Examples of these prepositional phrases were: *in the world, in our country, reason for this, about this issue, on the other hand, in the same, around the world*. As sub-corpus 2, 3 and 4 of both groups, the subcategory *noun phrase with of-phrase fragment* which comprised of 8 types in Group 1 and 9 types in Group 2 was the second mostly used one in sub-corpus 5 of both groups. In the same vein with the sub-corpus 1, 2, 3 and 4, *quantifier expressions, noun phrase with other post-modifier fragment, other noun phrase expressions* and *prepositional phrase with embedded of-phrase fragment* were in the low proportions of used sequences in both groups. While the sub-category *comparative expressions* appeared with the lowest proportion (1 types) in Group 1, it does not place in Group 2.

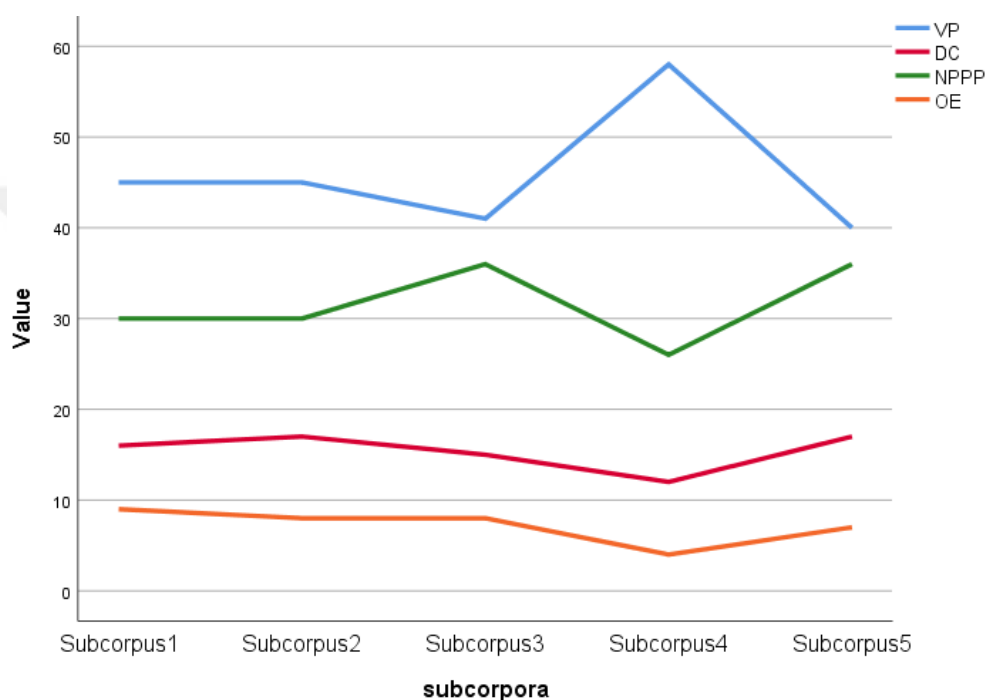
Dependent clause fragments incorporated 5 sub-categories in sub-corpus 5 of Group 1 and 4 sub-categories in sub-corpus 5 in Group 1. In the sub-categories of dependent clause fragments, it was important to state that as in sub-corpus 4 of both groups, *(verb phrase +) that-clause fragment* was the most commonly used one in this category in both groups. This category comprised of 7 types in group 1 and 10 types in Group 2 and the examples were the expressions such as *that it is, that there is, fails to mention that, that they are, that there is a, that there is no* and *feel that they*. Similar to other sub-corpora, *with wh-clause fragments, (verb/adjective +) to-clause fragment, adverbial clause fragment* and *if-clause fragments* were in the lower proportions of sequence types in both groups.

Similar to the sub-corpus 1, 2, 3 and 4 *other expressions* included the lowest overall proportion of formulaic sequences compared to VP-based, DC-based and NPPP-based structures, including 7 types in

both groups. The examples were: *as well as*, *as a result*, *according to the*, *according to a*, *to sum up* and *but also in*.

The overall distribution of three- and four-word formulaic sequences across structural categories in Group 1 can be seen in Figure 3 below. This figure provides a chance to see how the distribution of type frequency of main categories changes over time in Group 1.

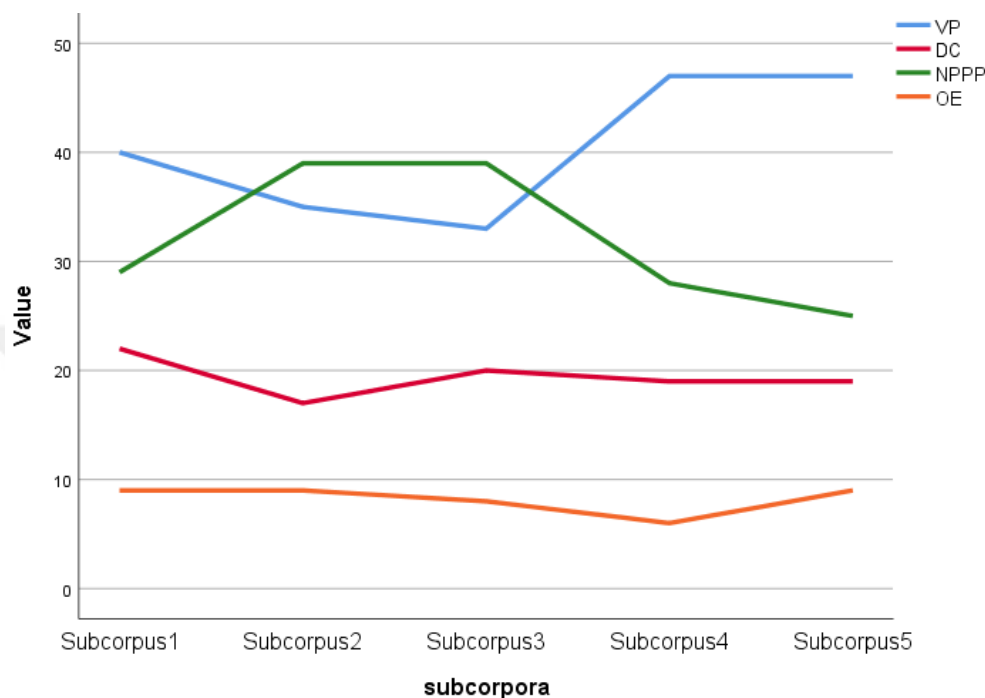
Figure 3: The Distribution of the Type Frequency of the Structural Categories of the Top 100 Frequent Formulaic Sequences across 5 Sub-corpora of Group 1



Based on Figure 3, it seems that there were some fluctuations in the usage of the items in the several structural categories in Group 1. As illustrated in Figure 3, the majority of the formulaic sequences of Group 1 took place in the category of *verb phrase fragments*. According to Biber et al. (2004), the structural types of formulaic sequences in the academic text are different from the structure of FSs in conversation. The findings of their study showed that nearly 90 percent of all formulaic sequences in conversation contain verb phrase fragments. It can be clearly seen from tables that *noun and prepositional phrase fragments* were the second mostly used structural types in each corpus of Group 1. Similarly, the results of the study conducted by Hyland (2008) reported that the *noun and prepositional phrase fragments* were the most common structure overall in the corpus of research articles, doctoral dissertations and master’s theses in four disciplines. Biber et al. (2004) found that the majority of the formulaic sequences in academic prose “consist of noun phrase expressions (e.g., *the nature of the*) or a sequence that bridges across two prepositional phrases (e.g., *as a result of*)” (2004: 382). While *dependent clause fragments* were the third mostly used structural types, the category of other expressions was less used ones and showed a falling tendency.

The overall distribution of three- and four-word formulaic sequences across structural categories in Group 2 can be observed in Figure 4 below. This figure gives a chance to see how the distribution of type frequency of main categories changes over time in Group 2.

Figure 4: The Distribution of the Type Frequency of the Structural Categories of the Top 100 Frequent Formulaic Sequences across 5 Sub-corpora of Group 2



Based on Figure 4, it seems that there were some fluctuations in the usage of the items in the several structural categories in Group 2. The majority of the formulaic sequences found were made up of *verb phrase fragments* and this was followed by *noun phrase and prepositional phrase fragments*. The three- and four-word formulaic sequences were found to vary in their structures, even though most of them are *verb phrase fragments*. This corresponded to what Fattani (2018) found with FSs in instructors' materials: "in terms of N-gram types, the VP- and PP-based forms are the most common structures, accounting for 34% and 33%" (2018: 113). It is important to note that the similar usage proportion of the category of *verb phrase fragments* and *noun and prepositional phrase fragments* found in Group 2 was seen compared to Group 1. The results of sub-corpus 2 and 3 are consistent with the findings of Drouhamane (2016) who investigated the usage patterns of four- and seven-word FSs in the writing of non-native writers taking the Canadian Academic English Language (CAEL) assessment that the investigator found that noun phrase and prepositional phrase fragments were predominant in academic prose. This is in agreement with the findings in earlier studies such as Biber et al. (1999); Cortes (2004) and Hyland (2008). As seen in Figure 6, the next category was dependent clause fragments and the formulaic sequences in the category of other expressions were not as common as other structural types of FSs.

On the other hand, by including far more phrase fragments than clause fragments, the findings of the analysis of the longitudinal learner corpora validated the notion that academic prose is more phrasal than clausal in structure, and this notion is attributed to the main informational purpose of academic prose (Biber and Gray, 2013). The formulaic sequences found in longitudinal learner corpora did not represent complete structural units as seen in the tables, such as *one of the*, *think that the*, *as a result of*, *and it is*, and *does not mean*, etc. This finding is in line with previous studies (Biber et al., 1999; Drouhamane, 2016) suggesting that most formulaic sequences in academic prose are also incomplete structural units. This finding was also supported by Cortes (2004) that they “are usually not complete structural units, but rather fragmented phrases or clauses with new fragments embedded” (2004: 400).

4.2.2.2. Functional Analysis

In addition to analysing syntactic structures, functional analysis of frequent formulaic sequences was employed in all sub-corpora of two groups. They were grouped under four functional categories which were adopted from Biber et al. (1999; 2004). Biber et al. (2004) distinguished between three primary functional types: *stance expressions*, *discourse organizers* and *referential expressions*. The majority of the FSs of all sub-corpora of two groups functioned as *referential expressions*. The types of *referential expressions* of the frequent formulaic sequences fluctuated over time in both groups.

Table 19 below displays the function of the top 100 three- to four-word formulaic sequences in sub-corpus 1 and sub-corpus 2 of Group 1 and Group 2.

Table 19: Function of the Top 100 Frequent Formulaic Sequences in Sub-corpus 1 and Sub-corpus 2 of Group 1 and Group 2

Functional Types	Sub-corpus 1		Sub-corpus 2	
	Group 1	Group 2	Group 1	Group 2
1. Stance expressions	<i>we can see, according to the, I mentioned above, I believe that, know how to, the fact that, an undeniable fact that, be aware of, you want to, they want to, we want to, want to do, not want to, we need to, we have to, they have to, have to do, should not be, should be given, it should be, is necessary for, we do not + verb, be able to, they cannot + verb, we cannot + verb, you cannot + verb, cannot be, and we can, in my opinion,</i>	<i>I strongly believe that, we can say that, we can see, as you can, you can see, as I said, I believe that, according to the, know how to, do not know how, the fact that, how to use, be aware of, can say that, we have to, they have to, we do not need, have to do, should not be, it should be, we do not + verb, be able to, they cannot + verb, we cannot + verb, cannot do, cannot be, in my opinion,</i>	<i>I strongly believe that, we can say, according to the, I mentioned above, I believe that, I firmly believe that, I think that, they think that, we can understand, learn how to, the fact that, do not know, is the best, can say that, think that the, as I said, as I mentioned, you want to, they want to, want to have, not want to, want to be, do not want, we need to, should go to, is necessary for, they will be, will not be, are going to, will be better, be able to, they cannot, and they can, cannot be, in my opinion,</i>	<i>I strongly believe that, I firmly believe that, according to the, according to a, the fact that, do not know, as you can, learn how to, know how to, have you ever, can be said that, is the best, are the best, want to have, do not want to, they have to, we have to, will be better, will not be, be able to, they cannot + verb, and they can, they can be, cannot be, in my opinion,</i>
2. Discourse organizers	<i>in the first, in the first place, first of all, on the other, on the other hand, but it is, as well as, as a result, in addition to, in order to, to sum up, last but not least, thanks to the, for this reason, as long as,</i>	<i>first of all, on the other hand, but it is, as well as, as a result, in addition to, in order to, to sum up, last but not least, after a while, as a result of, in other words, as long as,</i>	<i>first of all, on the other hand, as well as, as a result, in order to, to sum up, last but not least, due to the, for this reason, on the contrary, such as + noun, too +adj+ to, because they have, as a result of, too +adj+ to go,</i>	<i>first of all, on the other hand, as well as, as a result, in order to, to sum up, last but not least, due to the, for this reason, in addition to, too+adj+ to, too+adj+ to go, still too+adj+ to, such as +noun, as a result of, on the contrary, in other words,</i>
3. Referential expressions	<i>one of the, is one of the, this is the, one of the most, which is given, which are given, one of the reasons, it is a, it is an, it is not, it is the, it is important, is important for, that it is, that there are, there is a, there is no, the most important, because of this, because of the, is not a, who do not + verb, do not do, do not have, they do not + verb, a person who, with each other, it is an undeniable, to do their, not to be, there will be, he or she, his or her, of the most, most of the, a lot of, there are many, there are some, all of the, some of the, most of us, is the most, in terms of, in this case, the use of, in this way, the field of, the sense of, in the world, day by day, in the future, the same time, if it is, if they are, are given to, when they do,</i>	<i>one of the, is one of the, one of the most, it is a, it is not, that it is, that we can, that people have, there is a, there is no, there is no doubt, in this way, the most important, because of the, who does not, do not do, do not have, they do not, to do their, his or her, the reason why, with each other, a person who, it is an, as it is, it can be, that they are, that they can, thanks to the, of the most, most of the, a lot of, there are many, more and more, as much as, all of us, the number of, is the most, is the most important, in terms of, the use of, my point of view, the sense of, in many ways, of the world, in front of, day by day, in the past, at the same time, the end of, is no doubt that, to have a, to understand the, to begin with, to use it, to do it, not want to, they need to, not need to, people do not,</i>	<i>one of the, is one of the, one of the most, it is a, and it is, it is not, it is the, that it is, that there is, there is a, there is no, the most important, very important for, because of the, they do not + verb, he or she, who is a, is very important for, is important for, to be a, go to the, who are still, about this issue, his or her, of the most, a lot of, some of them, there are many, there are a lot, there are some, the number of, the rate of, is the most, are a lot of, there are lots of, in terms of, the use of, the best way, the importance of, of the world, in our country, in the world, day by day, in the future, when they are, in recent years, in the past, at the same time, to have a, to go to,</i>	<i>one of the, is one of the, one of the most, it is a, and it is, it is not, that it is, there is a, there is no, the most important, one of them, because of the, is not a, they do not + verb, who is a, who are still too, is very important, do not have, of the most important, it can be, it should be, need to be, his or her, he or she, while they are, because of their, it is an, of the most, most of the, a lot of, some of them, there are many, the number of, the rate of, is the most, in terms of, the use of, the best option, the importance of, the risk of, the lack of, the development of, best way to, in this way, in our country, in the world, of the country, in the country, day by day, in the future, when they are, at the same time, in the past, the beginning of, they grow up, to look after, to have a, to go to,</i>

For sub-corpus 1, in Table 19 above, when we look at the overall analysis of formulaic sequences across functional categories, it is seen that each type of *stance expressions* and *discourse organizers* occurred in sub-corpus 1 of Group 1 and Group 2. It is also revealed that nearly each type of *referential expressions* occurred in sub-corpus 1 of Group 1 and Group 2 except for the category of *imprecision*, *tangible framing attributes* and *text-deixis* while *multi-functional reference* did not occur in sub-corpus 1 of Group 1.

For sub-corpus 2, in Table 19, when we look at the overall analysis of formulaic sequences across functional categories, it is seen that each type of *stance expressions* and *discourse organizers* occurred in sub-corpus 2 of Group 1 and Group 2. It is also revealed that nearly each type of *referential expressions* occurred in sub-corpus 1 of Group 1 and Group 2 except for the category of *imprecision*, *tangible framing attributes* and *text-deixis* while *multi-functional references* did not occur in sub-corpus 2 of Group 1.

Table 20 illustrates the function of the top 100 three- and four-word frequent formulaic sequences in sub-corpus 3 and sub-corpus 4 of Group 1 and Group 2.

Table 20: Function of the Top 100 Formulaic Sequences in Sub-corpus 3 and Sub-corpus 4 of Group 1 and Group 2

Functional Types	Sub-corpus 3		Sub-corpus 4	
	Group 1	Group 2	Group 1	Group 2
1.Stance expressions	<i>I firmly believe that, according to the, I mentioned above, I believe that, I strongly believe that, I think that, the fact that, do not know, can be seen, is good for, of the fact that, as I mentioned above, according to a, according to their, they want to, they need to, we need to, we have to, should not be, if we want, they will be, in my opinion,</i>	<i>according to the, according to a, I strongly believe that, I firmly believe that, the fact that, do not know, know how to, not want to, do not want, they need to, we need to, should not be, to have a, be able to, they cannot + verb, in my opinion,</i>	<i>I can say that, but I think, I do feel that, I think it, according to the, according to my, I believe that, I firmly believe that, can say that, the fact that, do not know, going to be, they want to, we need to, I do not agree, I agree with, is going to, we do not + verb, I do not, do not agree with, will not be, agree with him, it will be, be able to, we cannot, I cannot, cannot be, person cannot find, will be able to, he thinks that, in my opinion,</i>	<i>I believe that, I firmly believe that, I think that, I strongly believe that, can say that, I do feel that, the fact that, as I mentioned, his thoughts about, think that the, to be a, according to a, according to the, according to my, they want to, should not be, we should not, I do not agree, is going to, we do not, I do not, do not agree with, will be a, be able to, we cannot, cannot be, he thinks that, people think that, in my opinion,</i>
2.Discourse organizers	<i>first of all, but it is, as a result, in order to, to sum up, last but not least, due to the, for this reason, as well as, as long as, in addition to,</i>	<i>first of all, as a result, in order to, to sum up, last but not least, for this reason, as well as, in addition to, on the other hand, all in all, as a result of,</i>	<i>look at the, because they are, because it is, while I do feel, On the other, on the other hand, as a result, in order to, to sum up,</i>	<i>look at the, we look at, to look at, while I do feel, because it is, when it comes to, on the other hand, in order to, to sum up, in addition to, for this reason, strongest supporting claims is, as a result, on the contrary,</i>
3.Referential expressions	<i>one of the, is one of the, one of the most, it is a, it is not, it is the, it is important, it is very, that it is, and it is, there is a, there is no, the most important, because of the, they are not, they do not, people who are, people who have, what to do, , for people to, it can be, it is important to, is very important, is important to, is important for, that they have, to be a, and they are, of them are, with each other, in the same, to go out, of the most, most of the, a lot of, there are many, there are some, is the most, we are all, most of their, some of them, are a lot of, there are a lot, as much as, all of the, in terms of, in this way, the benefits of, the development of, the importance of, a part of, in this case, in the world, of the world, in the country, in our country, in front of, around the world, day by day, in the future, in the past, when they are, when I was, when we look at, the people who, of the people, no matter how,</i>	<i>one of the, is one of the, one of the most, it is a, it is not, it is important, it is very, that it is, that they are, there is a, there is no, there will be, it is a fact, the most important, because of the, they are not, they do not, who live in, people who are, it should be, it can be, is important for, to be a, with each other, to each other, in the same, do not have, who is a, which is a, of the most important, all over the world, it will be, is a fact that, he or she, who do not, is not only, need to be, what is the, one of these, one of them, all of them, is a very, of the most, most of the, a lot of, there are many, is the most, most of their, some of them, will be more, so many people, the number of, a lot of people, a part of, in terms of, in this way, the development of, the importance of, point of view, the rules of, in the world, of the world, of the country, in the country, in front of, in the future, in the past, at the same time, day by day, make it possible, to make people, when I was, to create a,</i>	<i>one of the, is one of the, one of the most, it is a, it is not, and it is, it would be, it can be, that it is, that there is, there is a, there is no, the most important, the most important thing, thoughts about this issue, is the most important, is not a, does not have, they do not, who is a, does not mean, his thoughts about, but it is, this is a, this is the, of the most, most of the, a lot of, there are many, there are some, is the most, cannot find anything, more important than, the development of, the field of, In terms of, based on my, about this topic, the quality of, of the world, in the world, day by day, in the future, the end of, the end of the, this issue may seem, may seem as, fails to mention, not find anything to, find anything to support, think it is, he says that, if someone is, if it is, is looking hard enough, at first appearance, fails to mention that, to buy a, to say that, I have several issues,</i>	<i>one of the, is one of the, one of the most, it is a, it is not, it can be, it does not, that it is, that there is, there is a, there is no, the most important, the most important criteria, is the most important, is not a, is not the, they do not, who is a, does not mean, it will be, it should be, because of the, people do not, that it will, does not mean that, is that the, this is a, there will be, he or she, of the most, most of the, a lot of, in many ways, is the most, the development of, In terms of, based on my, about this issue, in the field of, point of view, at first appearance, the quality of, in the world, day by day, in the future, for a long time, the end of, he said that, if it is, he claims that, to say that, has failed to, he says that, fails to mention that, he fails to mention, of the people, may seem as,</i>

For sub-corpus 3, in Table 20, when we look at the overall analysis of formulaic sequences across functional categories, it is seen that each type of *stance expressions* and *discourse organizers* occurred in sub-corpus 3 of Group 2 while each type of *stance expressions* occurred in sub-corpus 3 of Group 1 except for the category of ability, and also each type of *discourse organizers* occurred in sub-corpus 3 of Group 1. And it is also revealed that nearly each type of *referential expressions* occurred in sub-corpus 3 of Group 1 and Group 2 except for the category of *imprecision* and *text-deixis*. It is important to note that *tangible framing attributes* appeared in sub-corpus 3 of Group 2 for the first time. While *multi-functional references* did not occur in sub-corpus 3 of Group 2, they did not occur in Group 1.

For sub-corpus 4, in Table 20, when we look at the overall analysis of formulaic sequences across functional categories, it is seen that each type of *stance expressions* and *discourse organizers* occurred in sub-corpus 4 of Group 1 and Group 2. And it is also revealed that nearly each type of *referential expressions* occurred in sub-corpus 4 of Group 1 and Group 2 except for the category of *imprecision*, *tangible framing attributes* and *text-deixis*. It is clear that *multi-functional references* occurred in both groups.

Table 21 illustrates the function of the top 100 frequent three- to four-word formulaic sequences in sub-corpus 5 of Group 1 and Group 2.

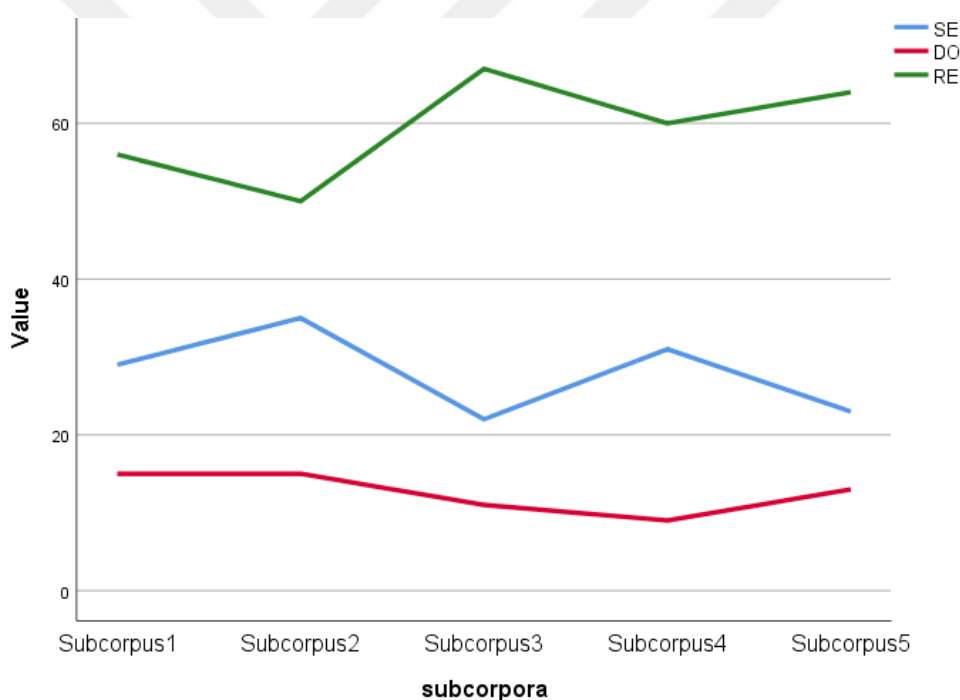
Table 21: Function of the Top 100 Frequent Formulaic Sequences in Sub-corpus 5 of Group 1 and Group 2

Functional Types	Group 1	Group 2
	3- and 4- FSs	3- and 4- FSs
1.Stance expressions	<i>according to the, according to a, I believe that, I firmly believe that, I think that, the fact that, can say that, it can be, do not think, to say that, should not be, I do not agree, I do not, do not agree with, will not be, and they can, they cannot, cannot be, to be a, be able to, cannot find, I cannot, in my opinion,</i>	<i>according to the, according to a, according to my, according to his, I believe that, I firmly believe that, I think that, we can see, I strongly believe that, the fact that, think that the, do not have, not want to, do not want to, they want to, should not be, need to be, I do not agree, I agree with, I do not, do not agree with, we do not, will be a, will not be, they cannot, cannot be, to be a, be able to, can be a, I cannot, we cannot, he thinks that, in my opinion,</i>
2.Discourse organizers	<i>look at the, because they are, on the other hand, as well as, as a result, in order to, to sum up, reason for this, the reason for, as a result of, in addition to, on the contrary, thanks to the,</i>	<i>because they are, on the other hand, as well as, as a result, in order to, to sum up, in other words, on the contrary, when it comes to, as a result of, as long as, in this way,</i>
3.Referential expressions	<i>one of the, one of the most, it is a, and it is, it is not, is one of, that it is, that there are, that they are, that there is, is that the, there is a, there is no, the most important, because of the, they are not, do not have, they do not, who is a, in the same, is not a, of the most, a solution to the, be the answer, because it is, claims is that, due to the, he claims that, in the work, is not the, it does not, most of the, people do not, people in the, people who are, the answer to the, the only way to, this is a, to increase the, we look at, when we look, as much as, a lot of, there are many, the number of, some of the, based on my, in terms of, about this issue, the problem of, about this topic, in this way, the rate of, the use of, in the world, around the world, of the world, day by day, at the same time, in the future, in his article, the end of, he says that,</i>	<i>one of the, it is a, it is not, is not a, is one of the, that it is, that there is, that they are, that there is a, that there is no, there is a, there is no, because of their, they are not, they do not, this is a, have the same, who is a, who is the, in the same, is not the, most of the, a solution to the, and it is, was published in, be a solution to, to have a, it can be, it does not, of the most, one of the most, the most important, there are some, and they are, a lot of, there are many, the number of, the rate of, based on my, in terms of, about this issue, the problem of, in the world, of the world, over the world, at the same time, day by day, in the future, he says that, he fails to mention, fails to mention that, but it is, does not mean, around the world,</i>

In Table 21, when we look at the overall analysis of formulaic sequences across functional categories, it is seen that each type of *stance expressions* occurred except for the category of *desire* in sub-corpus 5 of Group 1 while each type of *stance expressions* occurred in sub-corpus 5 of Group 2. It can be seen that there was no formulaic sequence in the category of *topic introduction/focus* while *topic elaboration/clarification* occurred in sub-corpus 5 of both groups. And it is also revealed that nearly each type of *referential expressions* occurred in sub-corpus 5 of Group 1 and Group 2 except for the category of *imprecision, tangible framing attributes, text-deixis* and *multi-functional reference* whereas time references did not occur in sub-corpus 5 of Group 2.

The overall distribution of three- and four-word formulaic sequences across functional categories in Group 1 can be seen in Figure 5 below. The following figure gives a chance to see how the distribution of type frequency of main categories changes over time in Group 1.

Figure 5: The Distribution of the Type Frequency of the Functional Categories of the Top 100 Frequent Formulaic Sequences across 5 Sub-corpora of Group 1

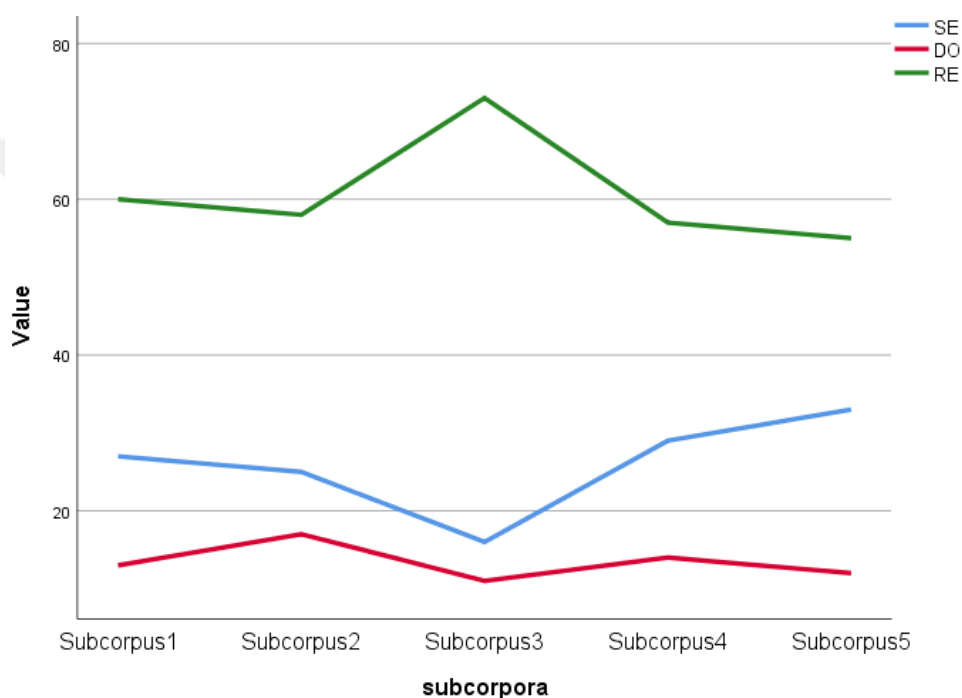


Based on the figure it seems that there were some fluctuations in the usage of the items in the several functional categories in Group 1. As can be seen, the majority of the formulaic sequences of Group 1 functioned as *referential expressions*. The types of the *referential expressions* decreased by sub-corpus 2 and 4 while *stance expressions* increased. In *stance expressions* category, the types of the *stance expressions* increased until sub-corpus 3. It is seen that in sub-corpus 3, whereas the types of *stance expressions* were less used in sub-corpus 3, the types of *referential expressions* are more popularly used. In the types of *discourse organizers*, there was a gradually decline nearly in all sub-corpora of Group 1 and they also were less used, compared to the other two functional categories. All

in all, the types of the functional categories fluctuated over time. These fluctuations can be given to the fact that the writers had various knowledge of *referential expressions* that differ from each other in writing.

The overall distribution of three- to four-word formulaic sequences across functional categories in Group 2 can be seen in Figure 6 below. The following figure provides a chance to see how the distribution of type frequency of main categories changes over time in Group 2.

Figure 6: The Distribution of the Type Frequency of the Functional Categories of the Top 100 Frequent Formulaic Sequences across 5 Sub-corpora of Group 2



The findings of Group 2 are in line with Group 1 that there were some fluctuations in the usage of the items in the several functional categories in Group 2 as shown in Figure 6. The majority of the formulaic sequences of Group 2 functioned as *referential expressions*. The types of the *referential expressions* decreased by sub-corpus 4 and 5 while *stance expressions* increased. In *stance expressions* category, the types of the *stance expressions* decreased until sub-corpus 3 after then there was continuous increase. In contrast to the continuous decline of *discourse organizers* in Group 1, the types of the *discourse organizers* fluctuated over time and they also were less used, compared to the other two functions.

As revealed in the type distribution of both groups over time (see Figure 5 and Figure 6), it seems that *referential expressions* were highly frequent while *discourse organizers* were less frequent. That is, what it is found were similar proportions for *referential expressions* in the two groups, and a greater proportion of *stance expressions* and a smaller proportion of *discourse organisers*. Similar to the

findings of the current study, Bal-Gezegin (2019) reported that *referential expressions* constituted the largest part (%75) followed by *discourse organizers* (%15) and *stance expressions* (%8) in the academic writing of L1 Turkish speakers of English. Similarly, Adel and Erman (2012) revealed that the largest part of proportion functioned as *referential expressions* in advanced learner writing both by L1 speakers of Swedish and native speakers. When compared to other studies with a focus on formulaic sequences, it was revealed that the distribution of functional category varies. In Chen and Baker's (2016) study, they found out that the majority of the formulaic sequences functioned as *discourse organizers* in L2 student writing. In the same vein, Staples et al. (2013) lighted upon that in written responses across three proficiency levels in the TOEFL iBT more than half of the formulaic sequences were in the category of *discourse organizers*. In brief, this study produced findings which corroborate the findings of a great deal of the previous work in this field.

4.2.3. Comparison of Native and Non-native Formulaic Sequences

The third research question centres upon the comparison of longitudinal learner corpora and the reference written (LOCNESS) corpus. "By gathering instances of the usage of learners and comparing these with normative model corpora, the language of learners can be explored in a much more profound way than the previous work on error analysis was able to do" (Bonelli, 2010: 25). Using frequency analysis, the 100 frequent formulaic sequences and their raw and normalized frequencies in LOCNESS were first extracted through Sketch Engine (please see Appendix 3). Next, to measure the strength and significance level between the shared FSs in each sub-corpus of the longitudinal learner corpora and LOCNESS, a Pearson correlation analysis was conducted.

For each sub-corpus of Group 1, the findings of the Pearson correlation demonstrated the followings: (1) Table 22 shows that there was a significant and positive relationship, moderate in strength between LOCNESS and sub-corpus 1 of Group 1 ($r = .408$, $N = 30$, $p = < 0.05$). (2) No relationship existed between sub-corpus 2 of Group 1 and LOCNESS ($r = .260$, $N = 33$) (Table 23). (3) There was a significant and positive relationship, moderate in strength between sub-corpus 3 of Group 1 and LOCNESS ($r = .542$, $N = 33$, $p = 0.01$) (Table 24). (4) There was a significant and positive relationship between sub-corpus 4 of Group 1 and LOCNESS ($r = .381$, $N = 33$, $p = 0.05$) and the correlation was of moderate strength (Table 25). (5) There was a significant and positive relationship between sub-corpus 5 of Group 1 and LOCNESS ($r = .351$, $N = 42$, $p = 0.05$) and the correlation was of moderate strength (Table 26). The following scatterplots summarize the results of correlation between each sub-corpus of Group 1 and LOCNESS (Figure 7).

Table 22: Correlations between Sub-corpus 1 of Group 1 and LOCNESS

		LOCNESS	SUB_CORPUS1ofG1
LOCNESS	Pearson Correlation	1	,408*
	Sig. (2-tailed)		,025
	N	30	30
SUB_CORPUS1ofG1	Pearson Correlation	,408*	1
	Sig. (2-tailed)	,025	
	N	30	30

*. Correlation is significant at the 0.05 level (2-tailed).

Table 23: Correlations between Sub-corpus 2 of Group 1 and LOCNESS

		LOCNESS	SUB_CORPUS2ofG1
LOCNESS	Pearson Correlation	1	,260
	Sig. (2-tailed)		,144
	N	33	33
SUB_CORPUS2ofG1	Pearson Correlation	,260	1
	Sig. (2-tailed)	,144	
	N	33	33

Table 24: Correlations between Sub-corpus 3 of Group 1 and LOCNESS

		LOCNESS	SUB_CORPUS3ofG1
LOCNESS	Pearson Correlation	1	,542**
	Sig. (2-tailed)		,001
	N	33	33
SUB_CORPUS3ofG1	Pearson Correlation	,542**	1
	Sig. (2-tailed)	,001	
	N	33	33

** . Correlation is significant at the 0.01 level (2-tailed).

Table 25: Correlations between Sub-corpus 4 of Group 1 and LOCNESS

		LOCNESS	SUB_CORPUS4ofG1
LOCNESS	Pearson Correlation	1	,381*
	Sig. (2-tailed)		,029
	N	33	33
SUB_CORPUS4ofG1	Pearson Correlation	,381*	1
	Sig. (2-tailed)	,029	
	N	33	33

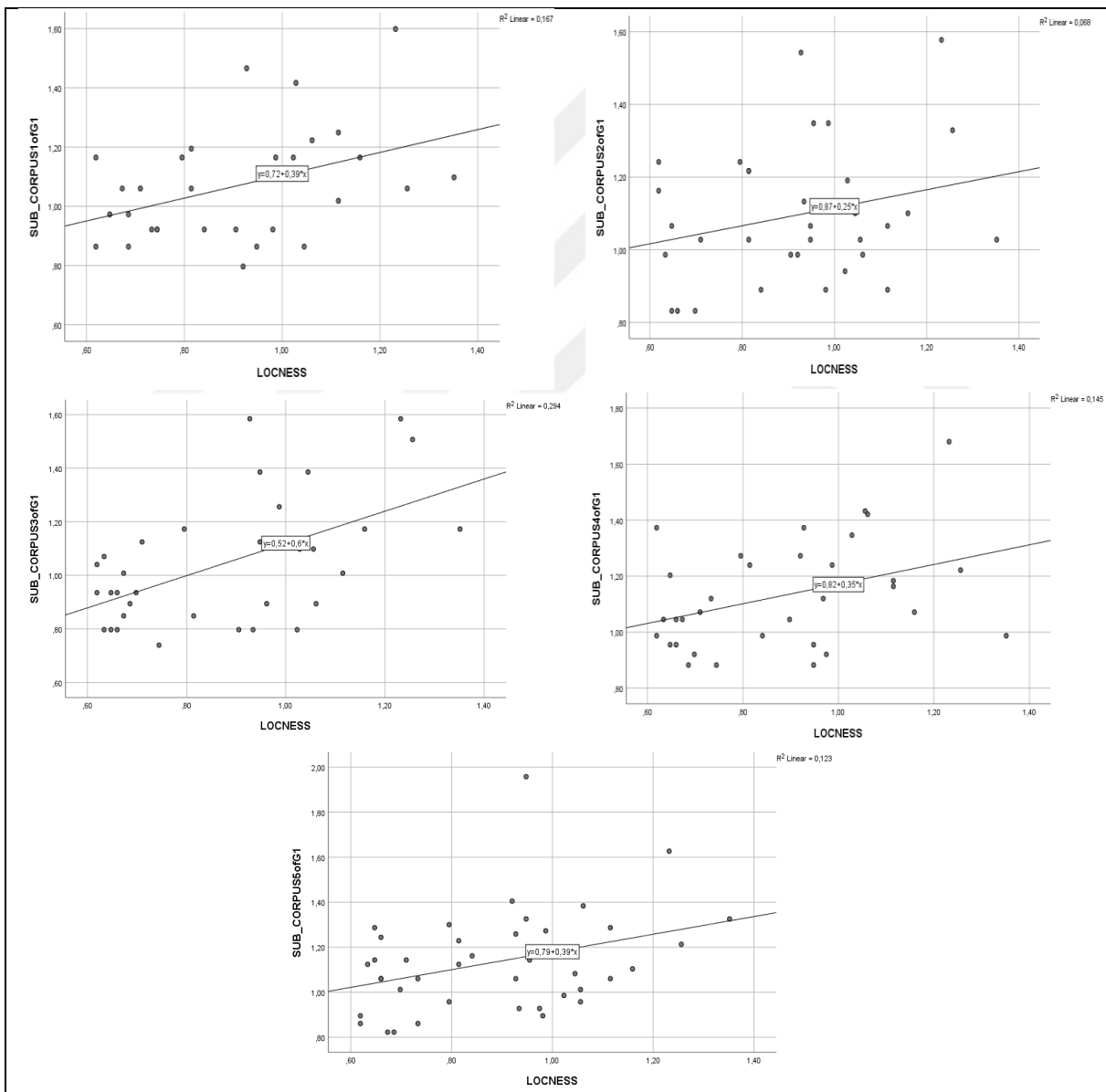
*. Correlation is significant at the 0.05 level (2-tailed).

Table 26: Correlations between Sub-corpus 5 of Group 1 and LOCNESS

		LOCNESS	SUB_CORPUS5ofG1
LOCNESS	Pearson Correlation	1	,351*
	Sig. (2-tailed)		,023
	N	42	42
SUB_CORPUS5ofG1	Pearson Correlation	,351*	1
	Sig. (2-tailed)	,023	
	N	42	42

*. Correlation is significant at the 0.05 level (2-tailed).

Figure 7: Scatter Plots of the Relationship between Frequency Scores in LOCNESS and Each Sub-corpus of Group 1 in Longitudinal Learner Corpora



For each sub-corpus of Group 2, the findings of the Pearson correlation displayed the followings: (1) No relationship existed between sub-corpus 1, 2, 3 and 4 of Group 2 and LOCNESS ($r = .268$, $N = 35$; $r = .165$, $N = 31$; $r = .272$, $N = 32$ and $r = .338$, $N = 33$, respectively) (Table 27, 28,29 and 30). (2) There was a significant and positive relationship between sub-corpus 5 of Group 2 and LOCNESS ($r = .537$, $N = 38$, $p = < 0.001$) and the correlation was of moderate strength (Table 31). These findings demonstrate that most of the sub-corpora of Group 2 hold several formulaic sequences that make them distinct and unique different from native written corpus (LOCNESS). The following scatterplots summarize the results of correlation between each sub-corpus of Group 2 and LOCNESS (Figure 8).

Table 27: Correlations between Sub-corpus 1 of Group 2 and LOCNESS

		LOCNESS	SUB_CORPUS1ofG2
LOCNESS	Pearson Correlation	1	,268
	Sig. (2-tailed)		,119
	N	35	35
SUB_CORPUS1ofG2	Pearson Correlation	,268	1
	Sig. (2-tailed)	,119	
	N	35	35

Table 28: Correlations between Sub-corpus 2 of Group 2 and LOCNESS

		LOCNESS	SUB_CORPUS2G2
LOCNESS	Pearson Correlation	1	,165
	Sig. (2-tailed)		,376
	N	31	31
SUB_CORPUS2G2	Pearson Correlation	,165	1
	Sig. (2-tailed)	,376	
	N	31	31

Table 29: Correlations between Sub-corpus 3 of Group 2 and LOCNESS

		LOCNESS	SUB_CORPUS3G2
LOCNESS	Pearson Correlation	1	,272
	Sig. (2-tailed)		,132
	N	32	32
SUB_CORPUS3G2	Pearson Correlation	,272	1
	Sig. (2-tailed)	,132	
	N	32	32

Table 30: Correlations between Sub-corpus 4 of Group 2 and LOCNESS

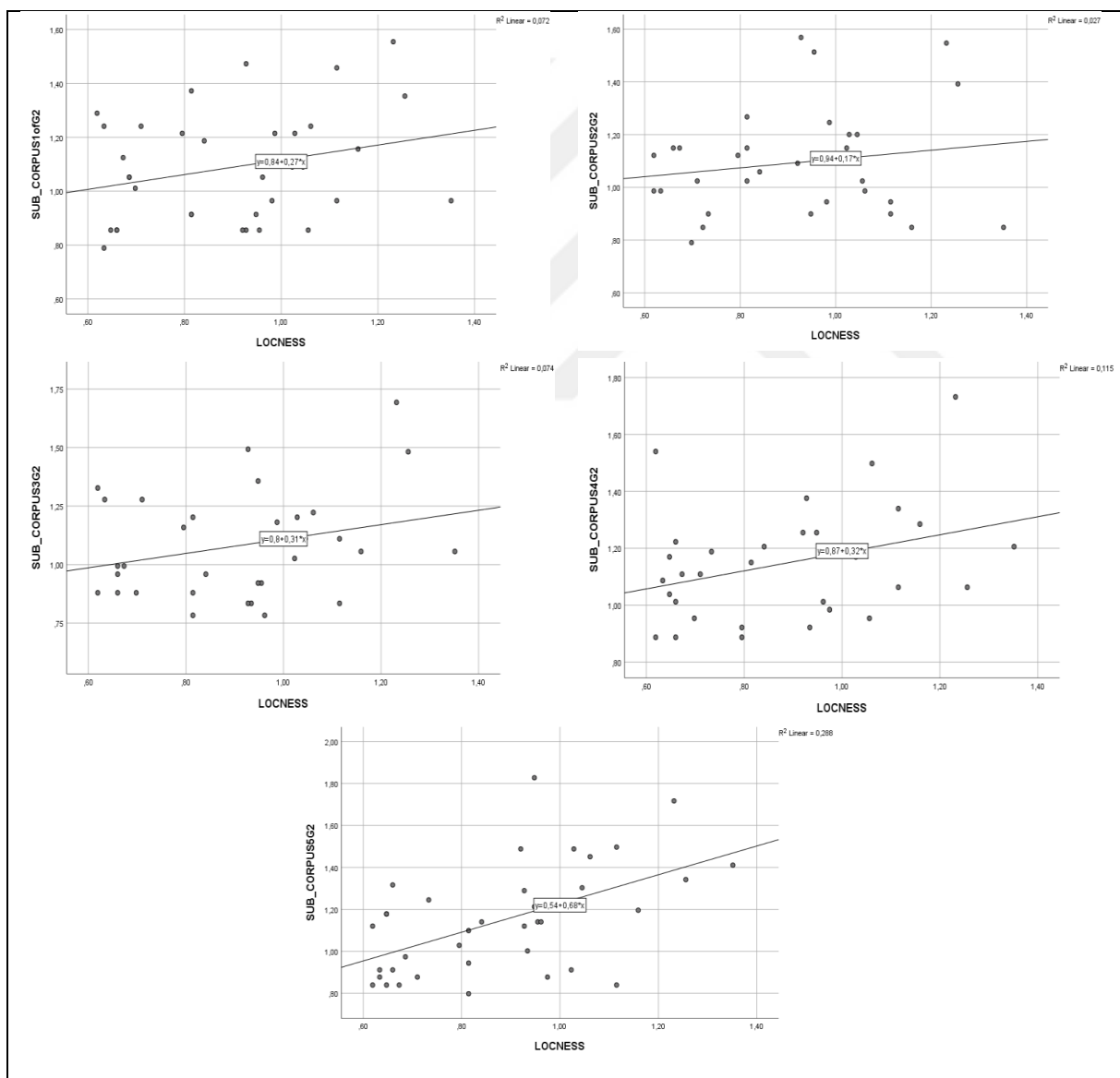
		LOCNESS	SUB_CORPUS4G2
LOCNESS	Pearson Correlation	1	,338
	Sig. (2-tailed)		,054
	N	33	33
SUB_CORPUS4G2	Pearson Correlation	,338	1
	Sig. (2-tailed)	,054	
	N	33	33

Table 31: Correlations between Sub-corpus 5 of Group 2 and LOCNESS

		LOCNESS	SUB_CORPUS5G2
LOCNESS	Pearson Correlation	1	,537**
	Sig. (2-tailed)		,001
	N	38	38
SUB_CORPUS5G2	Pearson Correlation	,537**	1
	Sig. (2-tailed)	,001	
	N	38	38

** . Correlation is significant at the 0.01 level (2-tailed).

Figure 8: Scatter Plots of the Relationship between Frequency Scores in LOCNESS and Each Sub-corpus of Group 2 in Longitudinal Learner Corpora



4.3. Individual Analysis: EFL Inventories of Formulaic Sequences

This section provides an in-depth analysis of formulaic sequences used by individual learners. It attempts to find out regarding how individual EFLs' inventory of formulaic sequences develops over time and the structure and function of the FSs produced by each learner. As mentioned earlier, eight participants from the same L1 background were chosen (see Table 8 in Chapter 3). As it was done in the analytical procedures of the Group Analysis, the first step was extracting the most frequent formulaic sequences for each sub-corpus that occurred twice or more times across two semesters. In the frequency analysis, Sketch Engine was used. The second step was classifying the structure and function of the frequent formulaic sequences. The last step was exploring the unique formulaic sequences used by individual participants.

As mentioned in the methodology section of this thesis, the criterion to identify a sequence as unique is that it is required to be used by an individual learner. The presence of word-for-word and partially shared usage of sequences by other participants in individual analysis, longitudinal learner corpora and LOCNESS means that the uniqueness criterion is not satisfied, or these sequences are not regarded as a unique pattern. A partially shared formulaic sequence is a sequence including part of the other sequences (longer or same length); for example, the sequence *do not agree with* comprises *do not agree*, and so this pattern is not considered as a unique sequence because of partially shared by other corpora.

4.3.1. Gizem's Inventory of Formulaic Sequences

The top 20 frequently occurring formulaic sequences produced by Gizem across 5 sub-corpora of the longitudinal learner corpus are presented in Table 32. Sub-corpus 1 consisted of 1,233 tokens and 10 formulaic sequence types. Sub-corpus 2 consisted of 1,494 tokens and 9 formulaic sequence types. Sub-corpus 3 consisted of 1,664 tokens and 18 formulaic sequence types. Sub-corpus 4 consisted of 2,077 tokens and 33 formulaic sequence types. Sub-corpus 5 consisted of 3,058 and 38 formulaic sequence types.

Table 32: Top 20 Frequent Formulaic Sequences Produced by Gizem across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
as long as	2	81,10	noun+ who are	3	100,40	in my opinion	3	90,14	I cannot	4	96,29	in the world	6	96,81
and quality of	2	81,10	noun+ better than	2	66,93	a lot of	3	90,14	to buy a	3	72,22	some of the	5	80,67
data which are	2	81,10	most of these	2	66,93	and it is	3	90,14	there is a	3	72,22	I do not	3	48,4
far from the	2	81,10	because of the	2	66,93	to live in	3	90,14	we need to	3	72,22	have shown that	3	48,4
it should not be	2	81,10	in rural areas	2	66,93	one of the	3	90,14	but at least the	2	48,15	in higher positions	3	48,4
of our age	2	81,10	it is a	2	66,93	know each other better	2	60,10	and durability of a	2	48,15	the formation of	3	48,4
it may be	2	81,10	there are many	2	66,93	he claims that	2	60,10	for long term	2	48,15	the number of	3	48,4
the sense of	2	81,10	she can ask	2	66,93	you know what	2	60,10	cannot help	2	48,15	in the same time	3	48,4
which are learned in	2	81,10	shows that the	2	66,93	how many of you	2	60,10	are more important than	2	48,15	be able to	2	32,27
will turn into	2	81,10				there are a	2	60,10	it is not	2	48,15	by changing the	2	32,27
						to get outside	2	60,10	it will be	2	48,15	can be seen	2	32,27
						to know each other	2	60,10	made by human	2	48,15	and that in some	2	32,27
						we should take	2	60,10	of a product	2	48,15	according to their	2	32,27
						you have to	2	60,10	pay a lot	2	48,15	have not been able	2	32,27
						to protect their	2	60,10	in case of	2	48,15	I think that	2	32,27
						to teach the	2	60,10	for a product	2	48,15	if they are	2	32,27
						an opportunity to	2	60,10	according to reports	2	48,15	in my opinion	2	32,27
						a part of	2	60,10	states that these	2	48,15	in some parts of	2	32,27
									that it is	2	48,15	and this is a	2	32,27
									the end of	2	48,15	of the world it	2	32,27

4.3.1.1. Structures of Formulaic Sequences in Gizem's each Sub-corpus

All the structural formulaic sequence types produced by Gizem in her subsequent essays are presented in Table 33.

Table 33: Structure of Formulaic Sequences in Gizem's each Sub-corpus

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
Personal pronoun + verb phrase (+ complement-clause fragment)		<i>she adds that this, she can ask,</i>	<i>he claims that, you know what, we should take, you have to,</i>	<i>I cannot, we need to, they do not,</i>	<i>smb+ claims that, I do not, I think that, they change their,</i>
Verb phrase with active verb	<i>will turn into,</i>		<i>know each other better,</i>	<i>cannot help, pay a lot, will be possible,</i>	<i>should not be, can be seen, have not been able, increase the number, reduce the number of, may be beneficial to, transferred to the, will increase the,</i>
FSs with wh-clause fragments	<i>data which are, which are learned in,</i>	<i>noun+ who are,</i>	<i>how many of you,</i>	<i>how can be, which has no,</i>	
Quantifier expressions			<i>a lot of,</i>		
Noun phrase with of-phrase fragment	<i>and quality of, of our age, the sense of,</i>	<i>most of these, because of the,</i>	<i>one of the, a part of,</i>	<i>of a product, and durability of a, the end of, because of their,</i>	<i>some of the, the formation of the number of, in some parts of, of the world it, one of the,</i>
Other noun phrase expressions				<i>the most important criteria,</i>	
Prepositional phrase with embedded of-phrase fragment				<i>a product for, for a product,</i>	
Other prepositional phrase fragment		<i>in rural areas,</i>	<i>in my opinion</i>	<i>but at least the, for long term, in case of, in my opinion,</i>	<i>in my opinion, in the world, in the same time,</i>
Anticipatory it + verb phrase/adjective phrase	<i>it should not be, it may be,</i>	<i>it is a,</i>	<i>and it is,</i>	<i>it is not, it will be, it can be,</i>	<i>it may be beneficial.</i>
Passive verb + prepositional phrase fragment				<i>made by human,</i>	
Copula be + noun phrase/adjective phrase				<i>are more important than, is not everything, is not the,</i>	<i>is a serious, are at the,</i>
(verb phrase +) that-clause fragment		<i>shows that the,</i>		<i>states that these, that it is,</i>	<i>that they cannot, have shown that, and that in some,</i>

Table 33: (Continue)

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
(verb/adjective +) to-clause fragment			<i>to live in, to get outside, to know each other, to protect their, to teach the, an opportunity to,</i>	<i>to buy a, the ways to, to make a,</i>	<i>an important role to, to be a, to their skills, to be taken, be able to,</i>
Pronoun/noun phrase + be (+ . . .)	<i>the knowledge is,</i>	<i>there are many,</i>	<i>there are a,</i>	<i>there is a,</i>	<i>and this is a, there is no,</i>
Other expressions	<i>as long as,</i>			<i>according to reports,</i>	<i>according to their, by changing the</i>
If-clause fragments					<i>if they are,</i>
Comparative expressions	<i>far from the,</i>	<i>noun+ better than,</i>		<i>the lower is the,</i>	<i>in higher positions,</i>

In sub-corpus 1, Gizem produced 7 types of formulaic sequence structures. The most utilized structural type was *noun phrase with of-phrase fragment*. She produced 3 different types of this structure with a total of 6 token frequencies, followed by 2 types of *anticipatory it + verb phrase/adjective phrase* with a total of 4 token frequencies and 2 types of *FSs with wh-clause fragments* with a total of 4 token frequencies.

In sub-corpus 2, Gizem produced 8 types of formulaic sequence structures. The most popular structural types were *personal pronoun + verb phrase (+ complement-clause fragment)* and *noun phrase with of-phrase fragment*. She produced 2 different types of these structures with a total of 4 token frequencies.

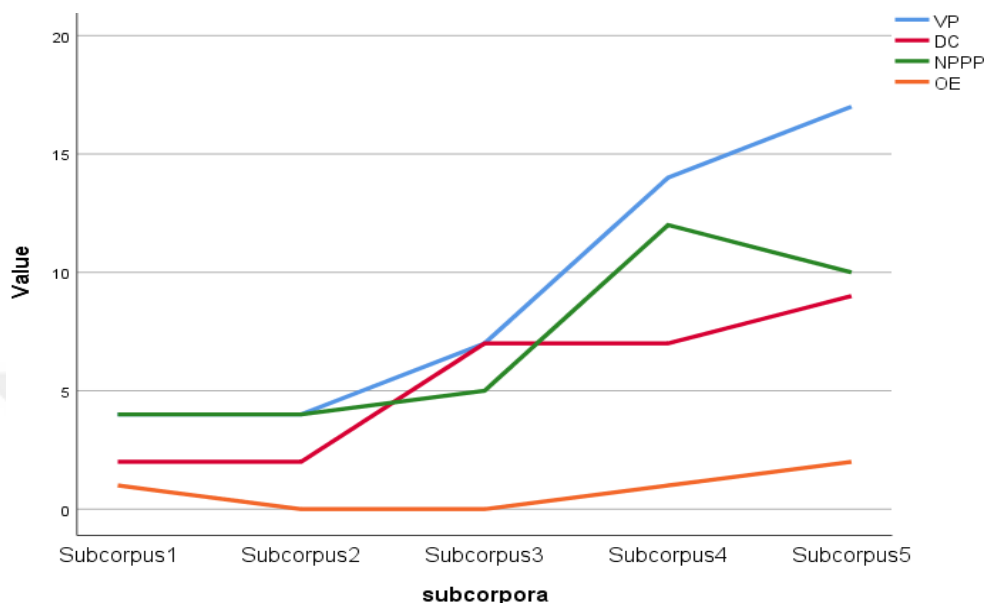
In sub-corpus 3, Gizem produced 9 types of formulaic sequence structures. The most frequent structural type was *(verb/adjective +) to-clause fragment*. She produced 6 different types of this structure with a total of 13 token frequencies, followed by 4 types of *personal pronoun + verb phrase (+ complement-clause fragment)* with a total of 8 token frequencies.

In sub-corpus 4, Gizem produced 15 types of formulaic sequence structures. Similar to sub-corpus 1, the most utilized structural types were *noun phrase with of-phrase fragment* and *other prepositional phrase fragment*. She produced 4 different types of these structures with a total of 8 token frequencies, followed by 3 types of *personal pronoun + verb phrase (+ complement-clause fragment)*, *verb phrase with active verb* and *anticipatory it + verb phrase/adjective phrase*.

In sub-corpus 5, Gizem produced 12 types of formulaic sequence structures. the most frequent structural type was *verb phrase with active verb*. She produced 8 different types of this structure with a total of 16 token frequencies, followed by 6 types of *noun phrase with of-phrase fragment* with a total of 17 token frequencies.

The distribution of structural types of formulaic sequences produced by Gizem across 5 sub-corpora is shown in Figure 9 below.

Figure 9: Distribution of Structural Types of Formulaic Sequences Produced by Gizem across 5 Sub-corpora



4.3.1.2. Functions of Formulaic Sequences in Gizem’s each Sub-corpus

Tables 34 present the different functional types of the formulaic sequences in each sub-corpus.

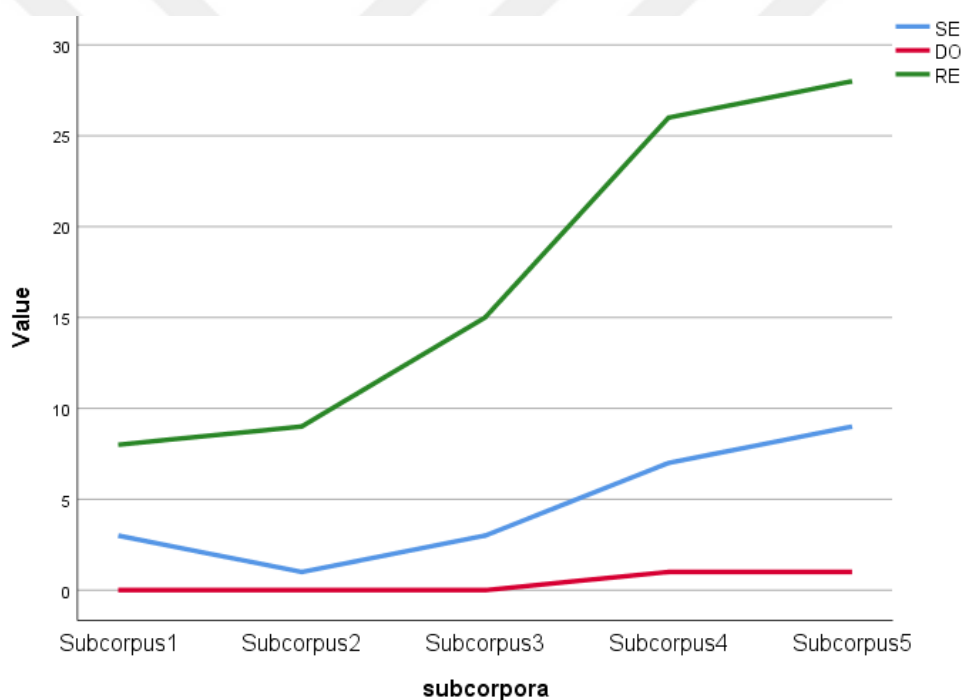
Table 34: Function of Formulaic Sequences in Gizem’s each Sub-corpus

Functional Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4-FSs	3- and 4-FSs	3- and 4-FSs	3- and 4-FSs	3- and 4-FSs
1. Stance expressions	<i>it should not be, it may be</i> <i>Will turn into,</i>	<i>she can ask,</i>	<i>we should take, you have to, in my opinion,</i>	<i>how can be, according to reports, we need to, will be possible, I cannot, cannot help, in my opinion,</i>	<i>according to their, should not be, I do not, be able to, can be seen, have not been able, I think that, in my opinion, to be a,</i>
2. Discourse organizers				<i>in case of,</i>	<i>in the same time,</i>
3. Referential expressions	<i>data which are, which are learned in, far from the, and quality of, of our age, the sense of, as long as, the knowledge is,</i>	<i>noun+ who are, most of these, because of the, it is a, noun+ better than, there are many, in rural areas, she adds that this, shows that the,</i>	<i>and it is, there are a, how many of you, a lot of, one of the, a part of, he claims that, you know what, know each other better, to live in, to get outside, to know each other, to protect their, to teach the, an opportunity to,</i>	<i>they do not, because of their, the most important criteria, a product for, for a product, but at least the, it is not, it can be, it will be, are more important than, is not everything, is not the, that it is, the ways to, there is a, of a product, and durability of a, for long term, the end of, the lower is the, made by human, states that these, to buy a, to make a, pay a lot, which has no,</i>	<i>that they cannot, an important role to, to their skills, have shown that, in higher positions, some of the, the formation of, the number of, in the world, by changing the, and that is some, if they are, in some parts of, and this is a, of the world it, one of the, increase the number, reduce the number of, is a serious, it may be beneficial, may be beneficial to, there is no, they change their, transferred to the, will increase the, smb+ claims that, to be taken, are at the,</i>

Referential expressions were the most recurrent functional type in each sub-corpus. In sub-corpus 1, while Gizem produced 8 *referential expressions* and 3 *stance expressions*, there were no *discourse organizers*. In sub-corpus 2, she produced similar types of functional categories as sub-corpus 1. In sub-corpus 3, the type of *referential expressions* (15 types) increased but *stance expressions* (3 types) remained a similar amount. In sub-corpus 4, there was a steady increase in the types of both *stance expressions* and *referential expressions* (7 types and 25 types respectively), and *discourse organizers* appeared with one type. In sub-corpus 5, while she produced 9 *stance expressions*, 28 *referential expressions* and 1 *discourse organizers*.

The distribution of functional types of formulaic sequences produced by Gizem across 5 sub-corpora is illustrated in Figure 10 below.

Figure 10: Distribution of Functional Types of Formulaic Sequences Produced by Gizem across 5 Sub-corpora



4.3.1.3. Unique Formulaic Sequences in Gizem’s each Sub-corpus

In sub-corpus 1, Gizem shared about 50% of formulaic sequence types with learner corpora and about 20% with LOCNESS on a word-for-word or partial basis. The following five formulaic sequence types were unique to Gizem: *data which are, far from the, of our age, which are learned in and will turn into.*

In sub-corpus 2, she shared approximately 78% of formulaic sequence types with learner corpora and about 44% with LOCNESS. Only two (*noun+ better than* and *in rural areas*) formulaic sequences were unique to her.

In sub-corpus 3, about 56% of formulaic sequence types were common with learner corpora while about 17% of FSs were common with LOCNESS. *You know what, know each other better, to know each other better, we should take, how many of you, to get outside, to protect their* and *to teach the* were the unique sequences to Gizem.

In sub-corpus 4, she shared almost 58% of formulaic sequence types with learner corpora and about 24% with LOCNESS. The following fourteen types of sequences were unique to her: *and durability of a, for long term, but at least the, pay a lot, for a product, is not everything, to make a, which has no, a product for, made by human, in case of, the lower is the, the ways to* and *will be possible*.

In sub-corpus 5, approximately 66% of formulaic sequence types were common with learner corpora, and about 29% of FSs were common with LOCNESS. *In higher positions, the formation of, may be beneficial to, by changing the, and that in some, in some parts of, reduce the number of, is a serious, are at the, they change their, to their skills, transferred to the* and *an important role to* were unique to her.

More than half of the formulaic sequences produced by Gizem were shared on a word-for-word or partial basis with longitudinal learner corpora whereas the proportion of shared sequences were less with native corpus. On the other hand, the analysis of unique sequences indicated that there was a gradual increase in the type and frequency of unique sequences from sub-corpus 1 to sub-corpus 5.

4.3.2. Elif's Inventory of Formulaic Sequences

The top 20 frequently occurring formulaic sequences produced by Elif across 5 sub-corpora of the longitudinal learner corpus are presented in Table 35. Sub-corpus 1 consisted of 1,204 tokens and 10 formulaic sequence types. Sub-corpus 2 consisted of 1,294 tokens and 26 formulaic sequence types. Sub-corpus 3 consisted of 1,494 tokens and 14 formulaic sequence types. Sub-corpus 4 consisted of 1,794 tokens and 29 formulaic sequence types. Sub-corpus 5 consisted of 3,183 and 66 formulaic sequence types.

Table 35: Top 20 Frequent Formulaic Sequences Produced by Elif across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
in this case	3	124,58	the fact that	3	115,92	due to the	4	133,87	of the person	6	167,22	in the world	13	204,21
for the students	3	124,58	due to the	3	115,92	on the contrary	3	100,40	the end of	4	111,48	I think that	7	109,96
can say that	2	83,06	day by day	2	77,28	inside than outside	2	66,93	the most important	3	83,61	he claims that	5	78,54
he or she	2	83,06	slowing down the	2	77,28	is difficult to	2	66,93	when people are	3	83,61	the increase of	5	78,54
it has a	2	83,06	cannot adapt the	2	77,28	is that the	2	66,93	end of the	3	83,61	the inadequacy of	4	62,83
they do not	2	83,06	due to the fact	2	77,28	more time inside than	2	66,93	due to the	2	55,74	more nutritious than	4	62,83
think that it	2	83,06	at least three	2	77,28	can be seen	2	66,93	of the person in	2	55,74	is not important	4	62,83
most of the	2	83,06	firmly believe that	2	77,28	the busy pace of	2	66,93	failed to present	2	55,74	that with the	4	62,83
according to the	2	83,06	they try to	2	77,28	the importance of	2	66,93	fails to mention	2	55,74	I believe that	3	47,13
a lot of	2	83,06	rate of the	2	77,28	the value of	2	66,93	as the end of	2	55,74	they think that	3	47,13
			that focused on	2	77,28	of nature on	2	66,93	he mentioned that	2	55,74	on the other hand	3	47,13
			have you ever thought	2	77,28	who live in	2	66,93	the price of	2	55,74	as a result of	3	47,13
			the difference between	2	77,28	one of the	2	66,93	when they are	2	55,74	who is a	3	47,13
			the pace of	2	77,28	in final consideration	2	66,93	on my research	2	55,74	as well as	3	47,13
			the rate of	2	77,28				one of the	2	55,74	he says that	3	47,13
			I firmly believe that	2	77,28				regardless of the	2	55,74	that it is	3	47,13
			the reduction rate of	2	77,28				I do feel that	2	55,74	that there is	3	47,13
			issues such as	2	77,28				in the article	2	55,74	I do not think	3	47,13
			to increase population	2	77,28				status of the person	2	55,74	the rate of	3	47,13
			which I mentioned above	2	77,28				such as the	2	55,74	in terms of	3	47,13

4.3.2.1. Structures of Formulaic Sequences in Elif's each Sub-corpus

All the structural formulaic sequence types produced by Elif in her each sub-corpus are presented in Table 36.

Table 36: Structure of Formulaic Sequences in Elif's each Sub-corpus

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
Personal pronoun + verb phrase (+ complement-clause fragment)	<i>they do not,</i>	<i>they try to, I firmly believe that, which I mentioned above, they cannot,</i>		<i>when people are, he mentioned that, when they are, I do feel that, while I do feel,</i>	<i>I think that, he claims that, I believe that, they think that, he says that, I do not think, as they grow early, he argues that, he thinks that, I do not agree, you may not, they are not so, they cannot, while I do,</i>
Verb phrase with active verb		<i>slowing down the, cannot adapt the, have you ever thought,</i>		<i>failed to present, fails to mention,</i>	<i>have a place in, have to get permission, take care of, should be at, spend more time, will not be, gives an example, has the same, do not agree with, do not have, will be solved,</i>
FSs with wh-clause fragments		<i>who are cared by, who are too young, who is Turkish, an who is,</i>	<i>who live in,</i>	<i>who is a, who is the,</i>	<i>who is a, people who are, who is an,</i>
Quantifier expressions	<i>a lot of</i>				<i>so many people, a lot of,</i>
Noun phrase with of-phrase fragment		<i>the pace of, rate of the, the rate of, the reduction rate of,</i>	<i>the busy pace of, the importance of, the value of, one of the,</i>	<i>of the person, the end of, end of the, of the person in, as the end of, the price of, one of the, regardless of the, status of the person, the end of the, the standard of,</i>	<i>the rate of, percent of the, the increase of, the inadequacy of, one of the, instead of exploiting,</i>
Noun phrase with other post-modifier fragment		<i>the fact that,</i>			
Other noun phrase expressions				<i>the most important,</i>	
Prepositional phrase with embedded of-phrase fragment					<i>in terms of, as a result of,</i>
Other prepositional phrase fragment	<i>in this case, for the students, most of the,</i>	<i>at least three, the difference between,</i>	<i>on the contrary, of nature on, in final consideration,</i>	<i>on my research, in the article,</i>	<i>in the world, on the other hand, on this subject,</i>

Table 36: (Continue)

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
Anticipatory it + verb phrase/adjective phrase	<i>it has a,</i>			<i>it is not a, it is not,</i>	<i>it is an, and it is, it is absolutely, it is a, it will not be,</i>
Passive verb + prepositional phrase fragment			<i>can be seen,</i>		
Copula be + noun phrase/adjective phrase			<i>is difficult to,</i>	<i>is not a,</i>	<i>is not important, are a profit for, are not high, are not the,</i>
(verb phrase +) that-clause fragment	<i>can say that, think that it,</i>	<i>firmly believe that, that focused on, you ever thought that,</i>	<i>is that the</i>	<i>to claim that,</i>	<i>that there is, that with the, that it is, do not think that, done was that, that are so important,</i>
(verb/adjective +) to-clause fragment		<i>to increase population,</i>			<i>a solution to the, a solution to this, way to solve,</i>
Pronoun/noun phrase + be (+ . . .)					<i>there are very serious, there is a,</i>
Adverbial clause fragment		<i>day by day,</i>			
Other expressions	<i>he or she, according to the,</i>	<i>due to the, due to the fact, issues such as,</i>	<i>due to the, inside than outside, more time inside than,</i>	<i>due to the, such as the,</i>	<i>as well as, according to the, even if they,</i>
Comparative expressions					<i>less salary than, more nutritious than,</i>

In sub-corpus 1, Elif produced 10 types of formulaic sequence structures. The most utilized structural type was *other prepositional phrase fragment*. She produced 3 different types of this structure, followed by 2 types of *(verb phrase +) that-clause fragment* and *other expressions*.

In sub-corpus 2, Elif produced 10 types of formulaic sequence structures. The followings were the most popular structural types: *personal pronoun + verb phrase (+ complement-clause fragment)*, *(verb phrase +) that-clause fragment*, *FSs with wh-clause fragments* and *noun phrase with of-phrase fragment*. She produced 4 different types of these structures with a total of 8 token frequencies.

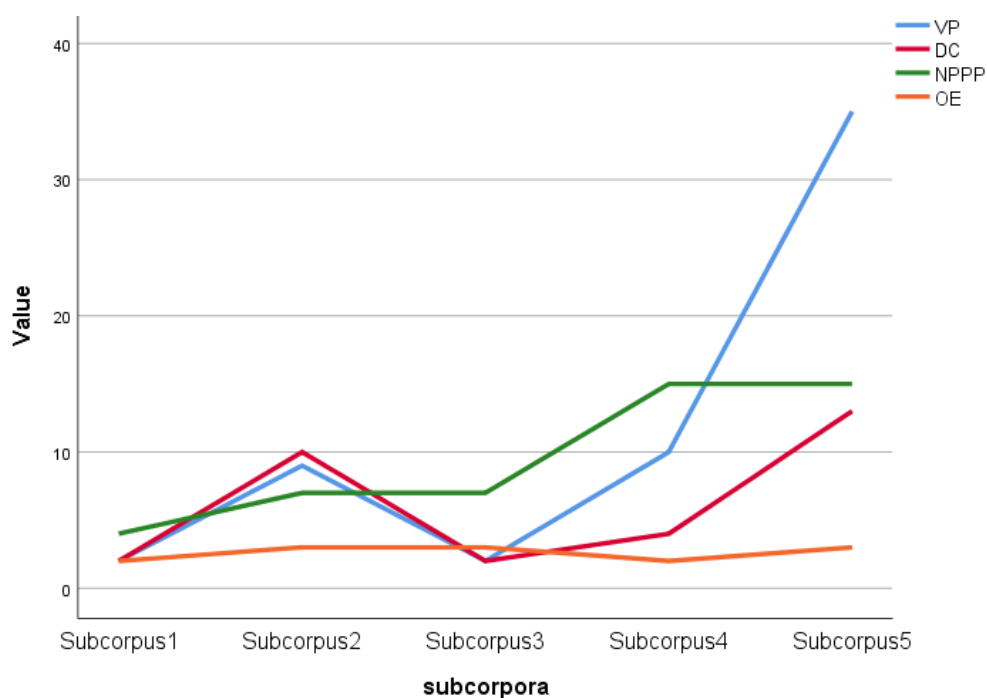
In sub-corpus 3, Elif produced 7 types of formulaic sequence structures. Similar to sub-corpus 2, the most utilized structural type was *noun phrase with of-phrase fragment*. She produced 4 different types of this structure with a total of 8 token frequencies, followed by 3 types of *other prepositional phrase fragment* and *other expressions* with a total of 7 and 8 token frequencies, respectively.

In sub-corpus 4, Elif produced 10 types of formulaic sequence structures. *Noun phrase with of-phrase fragment* was still the most frequent in terms of type (11) and token (27). It was followed by 5 types of *personal pronoun + verb phrase (+ complement-clause fragment)* with a total of 11 token frequencies.

In sub-corpus 5, Elif produced 14 types of formulaic sequence structures. The most frequent structural type was *personal pronoun + verb phrase (+ complement-clause fragment)*. She produced 14 different types of this structure with a total of 32 token frequencies, followed by 11 types of *verb phrase with active verb* with a total of 24 token frequencies.

The distribution of structural types of formulaic sequences produced by Elif across 5 sub-corpora is presented in Figure 11 below.

Figure 11: Distribution of Structural Types of Formulaic Sequences Produced by Elif across 5 Sub-corpora



4.3.2.2. Functions of Formulaic Sequences in Elif's each Sub-corpus

Tables 37 groups the functional types of the formulaic sequences produced by Elif across each sub-corpus.

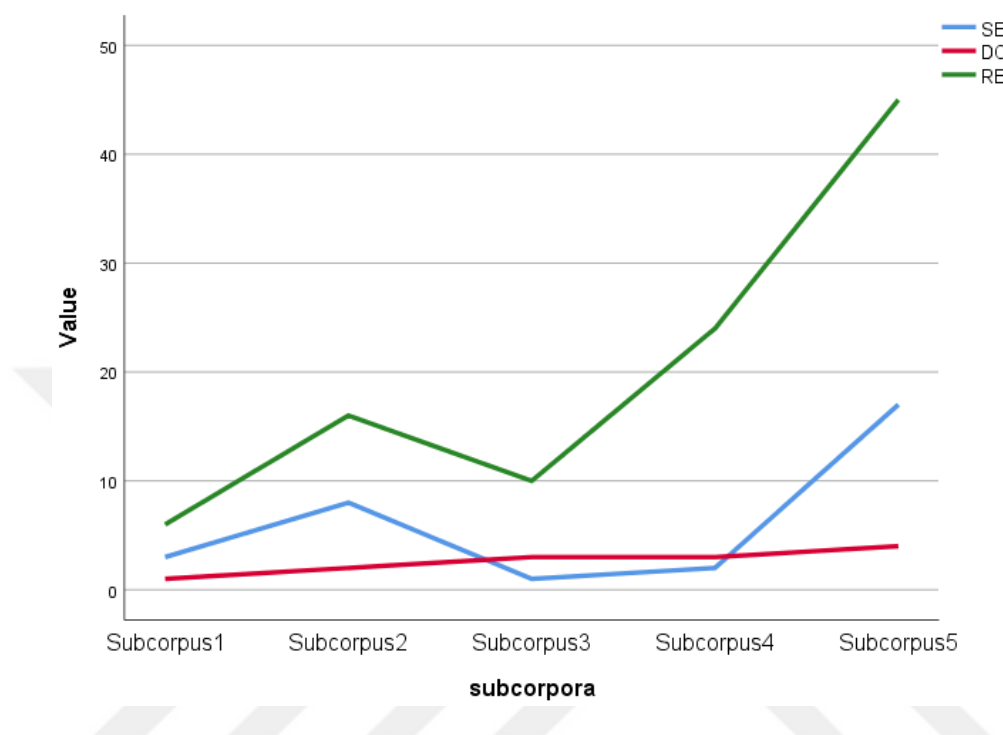
Table 37: Function of Formulaic Sequences in Elif's each Sub-corpus

Functional Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
1. Stance expressions	<i>can say that, think that it, according to the,</i>	<i>I firmly believe that, which I mentioned above, have you ever thought, the fact that, firmly believe that, you ever thought that, they cannot, cannot adapt the,</i>	<i>can be seen,</i>	<i>I do feel that, while I do feel,</i>	<i>I think that, I believe that, they think that, I do not think, do not think that, will not be, according to the, have to get permission. I do not agree, do not agree with, should be at, it will not be, you may not, they cannot, while I do, will be solved, he thinks that,</i>
2. Discourse organizers	<i>in this case,</i>	<i>due to the, due to the fact,</i>	<i>on the contrary, in final consideration, due to the,</i>	<i>regardless of the, on my research, due to the,</i>	<i>instead of exploiting, on the other hand, as a result of, as well as,</i>
3. Referential expressions	<i>they do not, for the students, most of the, it has a, he or she, a lot of,</i>	<i>who are cared by, who are too young, who is turkish, an who is, issues such as, at least three, the pace of, the rate of, the rate of, the reduction rate of, the difference between, day by day, slowing down the, that focused on, to increase population, they try to,</i>	<i>who live in, is that the, is difficult to, the busy pace of, the importance of, the value of, one of the, of nature on, inside than outside, more time inside than,</i>	<i>who is a, who is the, the most important, it is not a, it is not, is not a, such as the, of the person, of the person in, as the end of, the price of, one of the, status of the person, the standard of, in the article, the end of, end of the, the end of the, when people are, he mentioned that, when they are, failed to present, fails to mention, to claim that,</i>	<i>in the world, he claims that, the increase of, the inadequacy of, more nutritious than, is not important, that with the, who is a, he says that, that it is, that there is, the rate of, in terms of, gives an example, people who are, has the same, have a place in, as they grow early, he argues that, and it is, a solution to the, a solution to this, do not have, on this subject, one of the, percent of the, done was that, who is an, it is absolutely, it is a, it is an, so many people, spend more time, take care of, are a profit for, that are so important, less salary than, are not high, even if they, there are very serious, there is a, they are not so, way to solve, are not the, a lot of,</i>

For Elif, *referential expressions* were the most utilized functional type in each sub-corpus. In sub-corpus 1, while Elif produced 6 *referential expressions*, 3 *stance expressions* and 1 *discourse organizers*. In sub-corpus 2, *referential expressions* were the most popular with 16 different types while *stance expressions* were the second one with 8 types and *discourse organizers* were the last one with 2 types. Similarly, in sub-corpus 3, it consisted of 10 types of *referential expressions* and 1 type of *stance expressions* and 3 types of *discourse organizers*. *Referential expressions* were still the most frequent (24 types) in sub-corpus 4 and this functional type was followed by *stance expressions* (2 types). There were similar numbers in the category of *discourse organizers* (3 types). Sub-corpus 5 was consistent with the findings of each of the sub-corpus and it included 17 *stance expressions*, 4 *discourse organizers* and 45 *referential expressions*.

The distribution of functional types of formulaic sequences produced by Elif across 5 sub-corpora is displayed in Figure 12 below.

Figure 12: Distribution of Functional Types of Formulaic Sequences Produced by Elif across 5 Sub-corpora



4.3.2.3. Unique Formulaic Sequences in Elif's each Sub-corpus

In sub-corpus 1, Elif shared 70% of formulaic sequence types with learner corpora and 30% with LOCNESS on a word-for-word or partial basis. There were three unique formulaic sequences to her: *for the students, it has a* and *think that it*.

In sub-corpus 2, approximately 54% of formulaic sequence types were common with learner corpora while 12% of formulaic sequences were common with LOCNESS. The following five types of formulaic sequences were unique to her: *cannot adapt the, at least three, have you ever thought, who are too young, you ever thought that, an who is, slowing down the, they try to, that focused on, the difference between, to increase population*.

In sub-corpus 3, she shared 50% of formulaic sequence types with learner corpora and about 21% with LOCNESS. *Inside than outside, the busy pace of, more time inside than, is difficult to, the value of, of nature on* and *in final consideration* were unique to her.

In sub-corpus 4, she shared almost 62% of formulaic sequence types with learner corpora and about 31% with LOCNESS. The following eleven types of sequences were unique to her: *the price of,*

on my research, regardless of the, when people are, of the person, of the person in, failed to present, he mentioned that, status of the person, in the article and the standard of.

In sub-corpus 5, approximately 71% of formulaic sequence types were common with learner corpora, and about 18% of FSs were common with LOCNESS. *The increase of, the inadequacy of, that with the, gives an example, has the same, as they grow early, they produce has the, on this subject, instead of exploiting, done was that, should be at, spend more time, are a profit for, that are so important, you may not, are not high, there are very serious, way to solve and will be solved* were unique to her.

In line with the findings of Gizem, more than half of the formulaic sequences produced by Elif were shared on a word-for-word or partial basis with longitudinal learner corpora whereas the proportion of shared sequences were less with native corpus. On the other hand, the analysis of unique sequences showed that there was a regular increase in the type and frequency of unique sequences from sub-corpus 1 to sub-corpus 5.

4.3.3. Arda's Inventory of Formulaic Sequences

The top 20 frequently occurring formulaic sequences produced by Arda across 5 sub-corpora of the longitudinal learner corpus are presented in Table 38. Sub-corpus 1 consisted of 1,153 tokens and 15 formulaic sequence types. Sub-corpus 2 consisted of 1,166 tokens and 25 formulaic sequence types. Sub-corpus 3 consisted of 1,460 tokens and 23 formulaic sequence types. Sub-corpus 4 consisted of 1,840 tokens and 38 formulaic sequence types. Sub-corpus 5 consisted of 1,993 and 21 formulaic sequence types.

Table 38: Top 20 Frequent Formulaic Sequences Produced by Arda across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
the ones that	5	216,83	it is the	3	128,64	it is not	3	102,74	and it is	4	108,70	should be encouraged	4	100,35
because of the	3	130,10	it is a	3	128,64	won't be	3	102,74	since it is	3	81,52	the reason for	3	75,26
and the ones	3	130,10	would be the	3	128,64	if we are talking	2	68,49	at the same	3	81,52	it is not	3	75,26
the fact that	2	86,73	it is an	2	85,76	but it is	2	68,49	people to buy	3	81,52	as much as	2	50,18
do not like	2	86,73	but it is	2	85,76	because of the	2	68,49	in my opinion	3	81,52	for a long time	2	50,18
it helps us	2	86,73	be the better	2	85,76	different kind of	2	68,49	as soon as possible	2	54,35	because she is	2	50,18
for all of	2	86,73	for the world	2	85,76	for kids to	2	68,49	be able to	2	54,35	and there is	2	50,18
are the ones that	2	86,73	are three reasons why	2	85,76	and it is	2	68,49	and it is the	2	54,35	it is because	2	50,18
must be done	2	86,73	thing to do	2	85,76	is because there	2	68,49	also states that	2	54,35	we can see	2	50,18
most of the	2	86,73	to the life	2	85,76	is not important	2	68,49	not to be	2	54,35	less than men	2	50,18
the ones who	2	86,73	that it is	2	85,76	it is all	2	68,49	he is right about	2	54,35	one of the	2	50,18
there are also the	2	86,73	start to get	2	85,76	it is never	2	68,49	I do not	2	54,35	he is right about	2	50,18
there are some	2	86,73	the ones that	2	85,76	let go their	2	68,49	in the future	2	54,35	she did not	2	50,18
cannot be	2	86,73	in my opinion	2	85,76	since we are all	2	68,49	in this situation	2	54,35	in other words	2	50,18
a search for	2	86,73	grow up and	2	85,76	group of people	2	68,49	is not a	2	54,35	that we are	2	50,18
			the ones that are	2	85,76	are nothing but	2	68,49	is nothing but	2	54,35	is not because	2	50,18
			to be happy	2	85,76	the most important	2	68,49	cannot say	2	54,35	there are also	2	50,18
			in the world	2	85,76	the ones who	2	68,49	and there is	2	54,35	they are more	2	50,18
			we can prevent	2	85,76	I was a	2	68,49	more expensive than	2	54,35	they do not	2	50,18
			why it is	2	85,76	we are talking about	2	68,49	it is because	2	54,35	whether or not	2	50,18

4.3.3.1. Structures of Formulaic Sequences in Arda's each Sub-corpus

All the structural formulaic sequence types produced by Arda in her each sub-corpus are presented in Table 39.

Table 39: Structure of Formulaic Sequences in Arda's each Sub-corpus

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
Personal pronoun + verb phrase (+ complement-clause fragment)		<i>we can prevent,</i>	<i>since we are all, I was a, we are talking about,</i>	<i>he is right about, I do not, they want to, we are only, we cannot, we do not, when we are,</i>	<i>he is right about, she did not, they do not, because she is, we can see, they are more,</i>
Verb phrase with active verb	<i>do not like, are the ones that, must be done, cannot be,</i>	<i>would be the, start to get, grow up and,</i>	<i>won't be, let go their,</i>	<i>cannot say, do not think, sample also states,</i>	<i>should be encouraged,</i>
FSs with wh-clause fragments	<i>the ones who</i>		<i>the ones who,</i>		
Quantifier expressions				<i>a lot of,</i>	
Noun phrase with of-phrase fragment	<i>because of the, for all of,</i>		<i>because of the, different kind of, group of people, and all of,</i>	<i>of a product, of course we,</i>	<i>one of the,</i>
Noun phrase with other post-modifier fragment	<i>the fact that, the ones that,</i>	<i>the ones that,</i>		<i>the fact that, the ones that,</i>	
Other noun phrase expressions	<i>and the ones, a search for,</i>		<i>the most important,</i>		
Other prepositional phrase fragment	<i>most of the,</i>	<i>for the world, in my opinion, in the world,</i>	<i>for kids to,</i>	<i>at the same, in my opinion, in this situation,</i>	<i>the reason for, in other words,</i>
Anticipatory it + verb phrase/adjective phrase	<i>it helps us,</i>	<i>it is the, it is a, it is an, but it is, why it is, and it is,</i>	<i>it is not, but it is, and it is, it is all, it is never,</i>	<i>and it is, since it is, and it is the, it is because, it is the, it would be,</i>	<i>it is not, it is because,</i>
Copula be + noun phrase/adjective phrase		<i>be the better, are three reasons why, is a must,</i>	<i>is because there, is not important, are nothing but,</i>	<i>is not a, is nothing but,</i>	<i>is not because,</i>
(verb phrase +) that-clause fragment		<i>the ones that are, that it is,</i>		<i>also states that, said that the, that it was actually,</i>	<i>that we are,</i>
(verb/adjective +) to-clause fragment		<i>thing to do, to the life, to be happy,</i>	<i>to understand and,</i>	<i>people to buy, not to be, to be fooled by, be able to,</i>	
Adverbial clause fragment				<i>in the future,</i>	<i>for a long time,</i>
Pronoun/noun phrase + be (+ . . .)	<i>there are also the, there are some,</i>	<i>there are three reasons, there is no,</i>	<i>there will be more,</i>	<i>and there is, there are also,</i>	<i>there are also, and there is,</i>
Other expressions		<i>as soon as possible,</i>		<i>as soon as possible,</i>	<i>whether or not, as a servant</i>
If-clause fragments			<i>if we are talking,</i>		
Comparative expressions				<i>more expensive than,</i>	<i>less than men, as much as,</i>

In sub-corpus 1, Arda produced 8 types of formulaic sequence structures. The most utilized structural type was *verb phrase with active verb*. She produced 4 different types of this structure with a total of 8 token frequencies, followed by 2 types of *noun phrase with of-phrase fragment*, *noun phrase with other post-modifier fragment*, *other noun phrase expressions* and *pronoun/noun phrase + be (+ . . .)* with a total of 5, 7, 5 and 4 token frequencies, respectively.

In sub-corpus 2, Arda produced 10 types of formulaic sequence structures. The most frequent structural types were *anticipatory it + verb phrase/adjective phrase*. She produced 6 different types of these structures with a total of 14 token frequencies, followed by 3 types of *verb phrase with active verb*, *other prepositional phrase fragment* and *copula be + noun phrase/adjective phrase*.

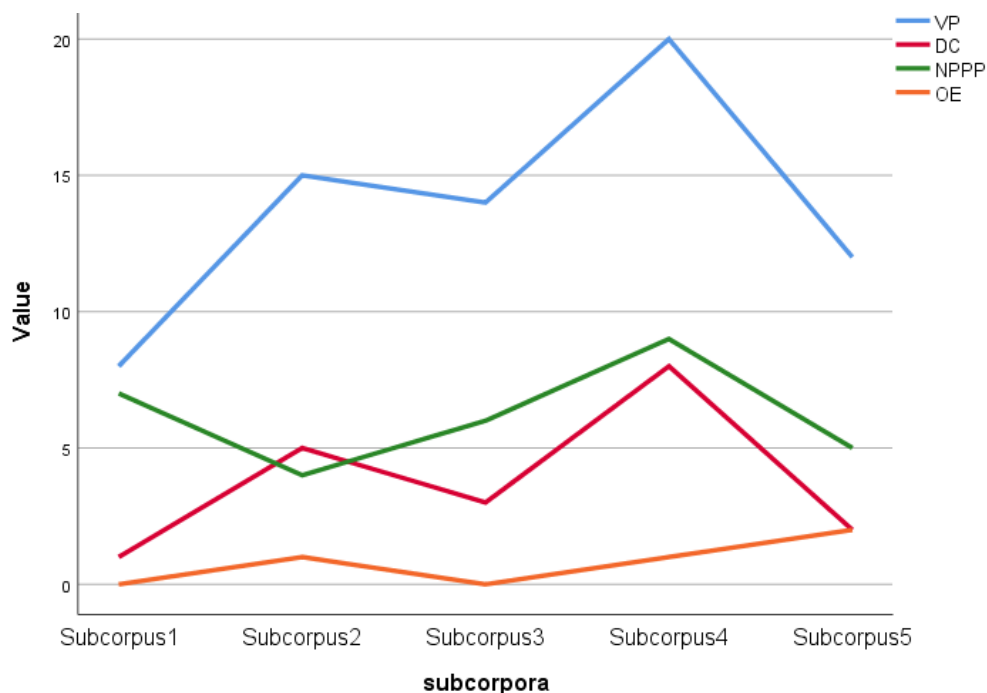
In sub-corpus 3, Arda produced 11 types of formulaic sequence structures. *Anticipatory it + verb phrase/adjective phrase* was still the most frequent in terms of type (5) and token (11). It was followed by 4 types of *noun phrase with of-phrase fragment* with a total of 8 token frequencies.

In sub-corpus 4, Arda produced 14 types of formulaic sequence structures. The most popular structural types were *personal pronoun + verb phrase (+ complement-clause fragment)*. She produced 7 different types of these structures with a total of 14 token frequencies, followed by 6 types of *anticipatory it + verb phrase/adjective phrase* with a total of 12 token frequencies.

In sub-corpus 5, Arda produced 11 types of formulaic sequence structures. *Personal pronoun + verb phrase (+ complement-clause fragment)* was still the most frequent in terms of type (6) and token (12). It was followed by 2 types of *other prepositional phrase fragment*, *anticipatory it + verb phrase/adjective phrase*, *pronoun/noun phrase + be (+ . . .)*, *other expressions* and *comparative expressions* with a total of 5 token frequency for the first category and 4 tokens for the others.

The distribution of structural types of formulaic sequences produced by Arda across 5 sub-corpora is demonstrated in Figure 13 below.

Figure 13: Distribution of Structural Types of Formulaic Sequences Produced by Arda across 5 Sub-corpora



4.3.3.2. Functions of Formulaic Sequences in Arda’s each Sub-corpus

Tables 40 shows the different functional types of the formulaic sequences produced by Arda in each sub-corpus.

Table 40: Function of Formulaic Sequences in Arda’s each Sub-corpus

Functional Types	<i>Sub-corpus 1</i>	<i>Sub-corpus 2</i>	<i>Sub-corpus 3</i>	<i>Sub-corpus 4</i>	<i>Sub-corpus 5</i>
	<i>3- and 4- FSs</i>	<i>3- and 4- FSs</i>	<i>3- and 4- FSs</i>	<i>3- and 4- FSs</i>	<i>3- and 4- FSs</i>
1. Stance expressions	<i>the fact that, do not like, must be done, cannot be,</i>	<i>we can prevent, be the better, to the life, would be the, in my opinion,</i>	<i>won't be</i>	<i>do not think, of course we, the fact that, they want to i do not, we do not, it would be, we cannot, cannot say, in my opinion,</i>	<i>she did not, they do not, should be encouraged, we can see,</i>
2. Discourse organizers					<i>the reason for, in other words,</i>

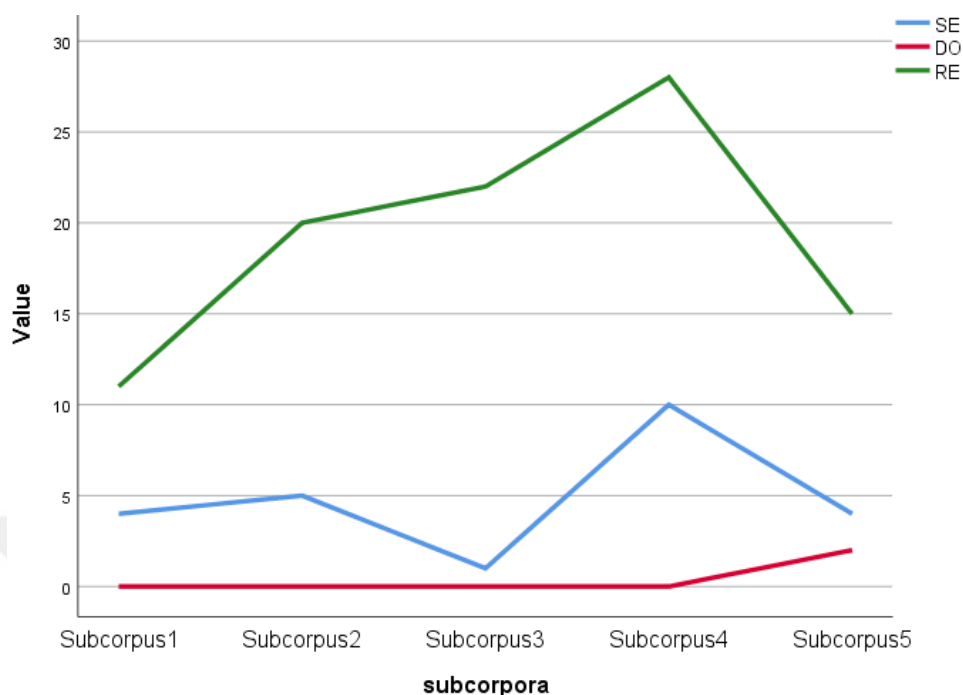
Table 40: (Continue)

Functional Types	<i>Sub-corpus 1</i> 3- and 4- FSs	<i>Sub-corpus 2</i> 3- and 4- FSs	<i>Sub-corpus 3</i> 3- and 4- FSs	<i>Sub-corpus 4</i> 3- and 4- FSs	<i>Sub-corpus 5</i> 3- and 4- FSs
3.Referential expressions	are the ones that, the ones who, because of the, for all of, the ones that, and the ones, most of the, there are also the, there are some, a search for, it helps us,	the ones that, it is the, it is a, it is an, but it is, why it is, and it is, are three reasons why, is a must, the ones that are, that it is, thing to do, there are three reasons, there is no, as soon as possible, for the world, in the world, start to get, grow up and, to be happy,	the ones who, because of the, the most important, it is not, but it is, and it is, it is all, it is never, is because there, is not important, are nothing but, there will be more, different kind of, group of people, and all of, since we are all, I was a, we are talking about, let go their, for kids to, to understand and, if we are talking,	the ones that, at the same, in this situation, and it is, since it is, and it is the, it is because, it is the, is not a, is nothing but, also states that, said that the, that it was actually, not to be, and there is, there are also, as soon as possible, a lot of, more expensive than, in the future, he is right about, we are only, when we are, sample also states, of a product, people to buy, to be fooled by,	it is not, it is because, is not because, there are also, less than men, for a long time, he is right about, because she is, as much as, and there is, one of the, that we are, they are more, whether or not, as a servant,

For Arda, *referential expressions* were the most recurrent functional type in each sub-corpus. In sub-corpus 1, he produced 11 types of *referential expressions* and 4 types of *stance expressions* while there was no type of *discourse organizers*. Similarly, in sub-corpus 2, the most popular type was *referential expressions* with 20 different types and there was no major difference in *stance expressions* types (5 types), and there was no type of *discourse organizers*. In sub-corpus 3, he produced 22 types of *referential expressions* and 1 type of *stance expressions* while there was no type of *discourse organizers*. In sub-corpus 4, he produced a higher number of *referential expressions* (28 types) and *stance expressions* (10 types) and then came *discourse organizers* with 1 type. In sub-corpus 5, *referential expressions* were the most common with 15 types and *stance expressions* were the second one with 4 types. Lastly, *discourse organizers* appeared with 2 types in this sub-corpus.

The distribution of functional types of formulaic sequences produced by Arda across 5 sub-corpora is submitted in Figure 14 below.

Figure 14: Distribution of Functional Types of Formulaic Sequences Produced by Arda across 5 Sub-corpora



4.3.3.3. Unique Formulaic Sequences in Arda's each Sub-corpus

In sub-corpus 1, Arda shared about 33% of formulaic sequence types with learner corpora and 20% with LOCNESS. There were ten unique formulaic sequences to him: *do not like, the ones that, and the ones, it helps us, for all of, are the ones that, the ones who, there are also the, must be done and a search for.*

In sub-corpus 2, he shared 40% of formulaic sequence types with learner corpora and 28% with LOCNESS. The following five types of formulaic sequences were unique to him: *for the world, to the life, start to get, would be the, be the better, are three reasons why, thing to do, the ones that, grow up and, the ones that are, we can prevent, there are three reasons, as soon as possible, to be happy and is a must.*

In sub-corpus 3, about 30% of formulaic sequence types were common with learner corpora, and about 13% of FSs were common with LOCNESS. *If we are talking, for kids to, won't be, different kind of, is because there, it is all, it is never, the ones who, I was a, we are talking about, to understand and, and all of, is not important, let go their, group of people and are nothing but* were unique to him.

In sub-corpus 4, he shared almost 53% of formulaic sequence types with learner corpora and about 18% with LOCNESS. The following eighteen types of sequences were unique to him: *people to buy, in this situation, is nothing but, of course we, since it is, as soon as possible, also states that, he is right*

about, cannot say, more expensive than, it is because, do not think, that it was actually, the ones that, we are only, when we are, to be fooled by and sample also states.

In sub-corpus 5, approximately 71% of formulaic sequence types were common with learner corpora while about 24% of FSs were common with LOCNESS. *Should be encouraged, she did not, that we are, they are more, whether or not and as a servant* were unique to him.

In line with the findings of Gizem, while more than half of the formulaic sequences produced by Arda were shared on the basis of word-for-word or partially with longitudinal learner corpora in sub-corpus 4 and 5, the proportion of common sequences were less than half in sub-corpus 1, 2 and 3. It was also clear that the proportion of shared sequences were at a quite low rate with native corpus. On the other hand, the analysis of unique sequences showed that whereas there was a continuous increase in the type and frequency of unique sequences from sub-corpus 1 to sub-corpus 4, the number of unique sequences steadily decreased in sub-corpus 5.

4.3.4. Lale's Inventory of Formulaic Sequences

The top 20 frequently occurring formulaic sequences produced by Lale across 5 sub-corpora of the longitudinal learner corpus are presented in Table 41. Sub-corpus 1 consisted of 1,004 tokens and 13 formulaic sequence types. Sub-corpus 2 consisted of 1,061 tokens and 13 formulaic sequence types. Sub-corpus 3 consisted of 1,656 tokens and 17 formulaic sequence types. Sub-corpus 4 consisted of 2,078 tokens and 55 formulaic sequence types. Sub-corpus 5 consisted of 1,924 and 36 formulaic sequence types.

Table 41: Top 20 Frequent Formulaic Sequences Produced by Lale across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
one of the	3	149,40	as a result	4	188,50	as long as	3	90,58	the end of	6	144,37	cannot be	4	103,95
given by the	3	149,40	more than one	3	141,38	in contact with	2	60,39	I do not	6	144,37	in the world	4	103,95
is not given	2	99,60	have to have more	2	94,25	in order to	2	60,39	can lead to	4	96,25	one of the	4	103,95
cannot achieve success	2	99,60	because they have	2	94,25	the strengthening of	2	60,39	are extremely useful	4	96,25	because of the	3	77,96
cannot distinguish	2	99,60	have to think about	2	94,25	in this way	2	60,39	of the human	4	96,25	the number of	3	77,96
in the house	2	99,60	to have more than	2	94,25	spend time in	2	60,39	an end to	4	96,25	the original ones	3	77,96
the subjects that	2	99,60	attention to the	2	94,25	to live together	2	60,39	can be the end	4	96,25	we try to	3	77,96
not understand the subject	2	99,60	on the contrary	2	94,25	such as +noun	2	60,39	the development of a	4	96,25	one of the biggest	2	51,98
of our lives	2	99,60	such as +noun	2	94,25	that there are	2	60,39	but the development of	4	96,25	in that point	2	51,98
prepared for the	2	99,60	very important for	2	94,25	live together in	2	60,39	of a full-scale	4	96,25	in this way	2	51,98
reinforce the subject	2	99,60	will be better	2	94,25	the opportunity to	2	60,39	be the end of	4	96,25	is the easiest	2	51,98
to be successful	2	99,60	the health of	2	94,25	to be able to	2	60,39	to the human	3	72,18	do not have any	2	51,98
and cannot	2	99,60	have more than one	2	94,25	we must ensure that	2	60,39	in contrast to	3	72,18	argued that they	2	51,98
						up in the	2	60,39	bring an end to	3	72,18	are the only alternative	2	51,98
						we must free	2	60,39	is actually in contrast	3	72,18	it can be	2	51,98
						we should not	2	60,39	contrast to an increase	3	72,18	because it is	2	51,98
						grow up in	2	60,39	contribute to people's	3	72,18	it is well known	2	51,98
									an increase of the	3	72,18	it may even	2	51,98
									it leads to the	3	72,18	longer than the	2	51,98
									do not agree with	3	72,18	according to the	2	51,98

4.3.4.1. Structures of Formulaic Sequences in Lale's each Sub-corpus

All the structural formulaic sequence types produced by Lale in her each sub-corpus are presented in Table 42.

Table 42: Structure of Formulaic Sequences in Lale's each Sub-corpus

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
Personal pronoun + verb phrase (+ complement-clause fragment)		<i>because they have,</i>	<i>we must ensure that, we must free, we should not,</i>	<i>I do not, we have seen that, people think like a, I do not agree, I do not support, I disagree with,</i>	<i>they should not, we try to, they produce more, we can easily,</i>
Verb phrase with active verb	<i>cannot achieve success, cannot distinguish, not understand the subjects, prepared for the, reinforce the subject, and cannot,</i>	<i>have to have more, have to think about, will be better, have more than one,</i>	<i>spend time in, live together in, grow up in,</i>	<i>can lead to, can be the end, bring an end to, do not agree with, does not bring an, cannot show, will also create a, look at your, may be right, not support this idea, have no effect on, will facilitate the,</i>	<i>cannot be, should not be, make a career, do not have any, try to keep,</i>
FSs with wh-clause fragments				<i>why people think like,</i>	
Noun phrase with of-phrase fragment	<i>one of the, of our lives,</i>	<i>the health of,</i>	<i>the strengthening of,</i>	<i>the end of, of the human, the development of a, but the development of, an increase of the, different areas of, of all people and, a number of, of people such as, a problem of, the ability of a, as a result of, of a full-scale,</i>	<i>one of the, because of the, the number of, one of the biggest, the problem of, the end of the, what kind of,</i>
Prepositional phrase with embedded of-phrase fragment	<i>in the house,</i>				
Other prepositional phrase fragment		<i>on the contrary, very important for,</i>	<i>in contact with, in order to, in this way, up in the,</i>	<i>in contrast to, in a similar way, in different areas of,</i>	<i>in the world, in that point, in this way, except for the, by the way,</i>
Anticipatory it + verb phrase/adjective phrase				<i>it leads to the, it is always, it is a,</i>	<i>it can be, because it is, it is well known, it may even,</i>
Passive verb + prepositional phrase fragment	<i>given by the,</i>				
Copula be + noun phrase/adjective phrase	<i>is not given,</i>			<i>are extremely useful, is actually in contrast, is the ability of a,</i>	<i>not be in, is the easiest, are the only alternative,</i>
(verb phrase +) that-clause fragment	<i>the subjects that,</i>		<i>that there are,</i>	<i>that knowledge is,</i>	<i>argued that they, that even though, that there is a, that we can,</i>

Table 42: (Continue)

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
(verb/adjective +) to-clause fragment	<i>to be successful,</i>	<i>to have more than, attention to the,</i>	<i>to live together, the opportunity to, to be able to,</i>	<i>an end to, to the human, contrast to increase, contribute to people's, to people's thoughts is, to an increase, to the development, an end to the, to contribute to, a similar way to, contrast to an increase,</i>	<i>way to produce,</i>
Other expressions		<i>as a result, such as +noun,</i>	<i>as long as, such as +noun,</i>	<i>and even in, according to his, as a result,</i>	<i>the original ones, according to the,</i>
Comparative expressions		<i>more than one,</i>			<i>longer than the,</i>

In sub-corpus 1, Lale produced 7 types of formulaic sequence structures. The most utilized structural type was *verb phrase with active verb*. She produced 6 different types of this structure with a total of 12 token frequencies, followed by 2 types of *noun phrase with of-phrase fragment* with a total of 5 token frequencies.

In sub-corpus 2, Lale produced 7 types of formulaic sequence structures. *Verb phrase with active verb* was still the most frequent in terms of type (4) and token (8). It was followed by 2 types of *other prepositional phrase fragment*, *(verb/adjective +) to-clause fragment* and *other expressions* with a total of 4, 4 and 6 token frequencies, respectively.

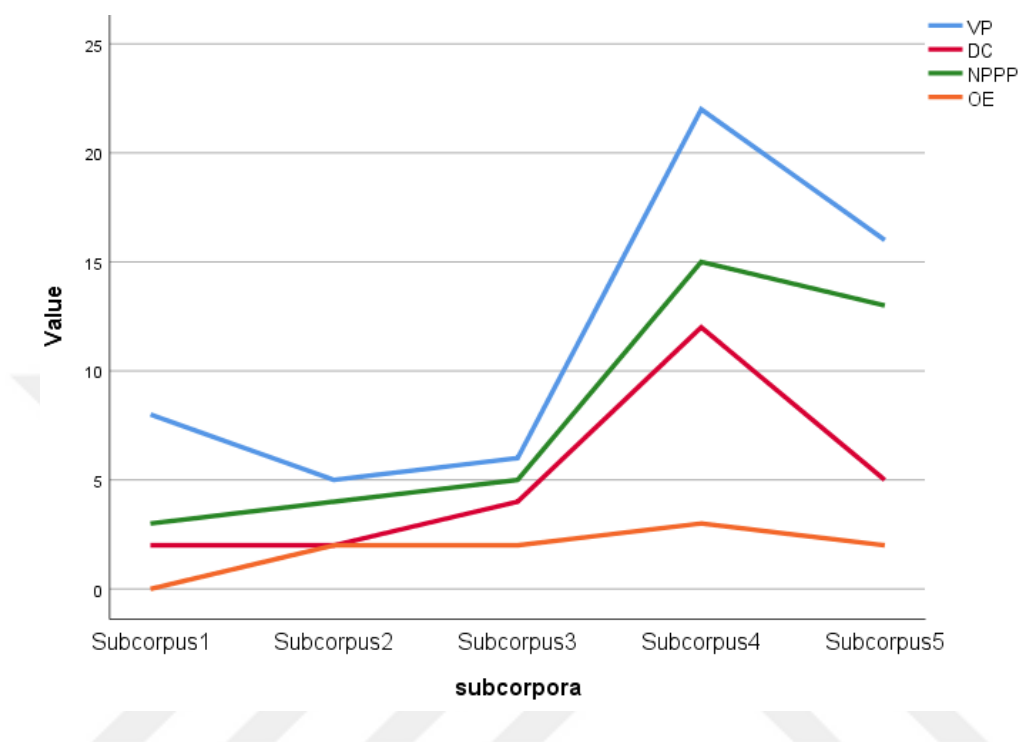
In sub-corpus 3, Lale produced 8 types of formulaic sequence structures. The most utilized structural type was *other prepositional phrase fragment*. She produced 4 different types of this structure with a total of 8 token frequencies, followed by 3 types of *verb phrase with active verb* with a total of 6 token frequencies.

In sub-corpus 4, Lale produced 10 types of formulaic sequence structures. Similar to sub-corpus 1 and 2, the most frequent structural type was *verb phrase with active verb*. She produced 13 different types of these structures with a total of 34 token frequencies, followed by 12 types of *noun phrase with of-phrase fragment* with a total of 37 token frequencies.

In sub-corpus 5, Lale produced 10 types of formulaic sequence structures. As other 3 sub-corpora, the most popular structural type was *verb phrase with active verb*. She produced 5 different types of this structure with a total of 12 token frequencies, followed by 4 types of *personal pronoun + verb phrase (+ complement-clause fragment)* with a total of 9 token frequencies.

The distribution of structural types of formulaic sequences produced by Lale across 5 sub-corpora is depicted in Figure 15 below.

Figure 15: Distribution of Structural Types of Formulaic Sequences Produced by Lale across 5 Sub-corpora



4.3.4.2. Functions of Formulaic Sequences in Lale's each Sub-corpus

Tables 43 submits the different functional types of the formulaic sequences in each sub-corpus.

Table 43: Function of Formulaic Sequences in Lale's each Sub-corpus

Functional Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
1.Stance expressions	<i>cannot achieve success, cannot distinguish, and cannot,</i>	<i>have to have more, have to think about, will be better,</i>	<i>we must ensure that, we must free, we should not, to be able to,</i>	<i>I do not agree, I do not support, I disagree with, can lead to, can be the end, do not agree with, cannot show, will also create a, not support this idea, have no effect on, will facilitate the, to people's thoughts is, according to his, may be right, I do not,</i>	<i>they should not, should not be, cannot be, it may even, we can easily, according to the, that we can,</i>
2.Discourse organizers		<i>on the contrary, as a result,</i>	<i>in contact with, in order to, as long as,</i>	<i>look at your, as a result of, in contrast to, as a result,</i>	<i>because of the, in that point, in this way, by the way,</i>

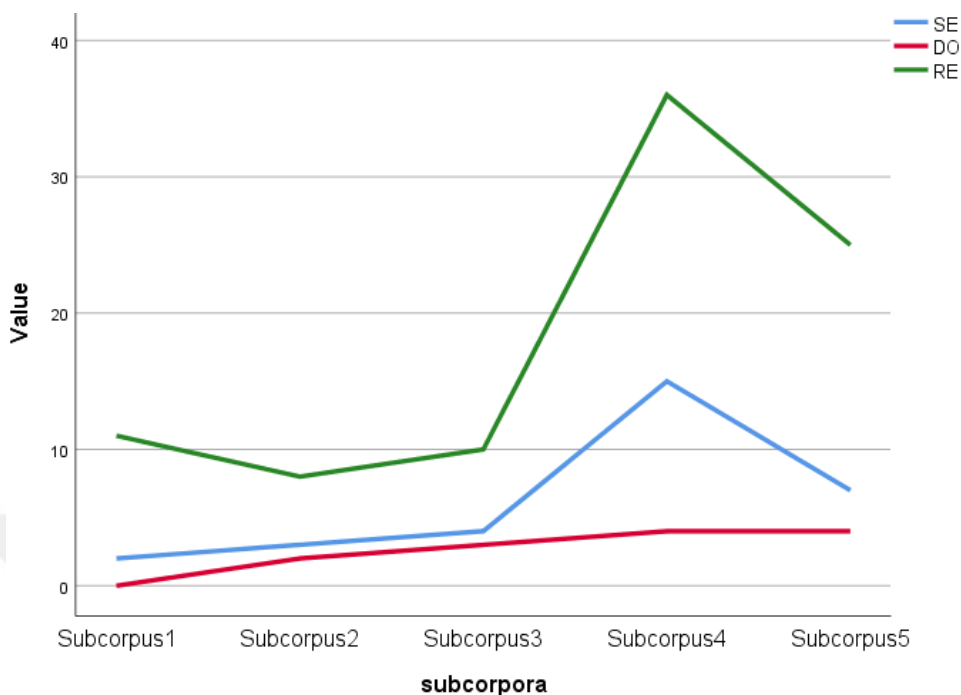
Table 43: (Continue)

Functional Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
3.Referential expressions	one of the, of our lives, the subjects that, in the house, not understand the subjects, prepared for the, reinforce the subject, given by the, is not given, to be successful,	very important for, such as +noun, the health of, because they have, have more than one, to have more than, attention to the, more than one,	in this way, up in the, that there are, such as +noun, the strengthening of, spend time in, live together in, grow up in, to live together, the opportunity to,	why people think like, of people such as, it leads to the, it is always, it is a, of a full-scale, contrast to an increase, a number of, the end of, of the human, the development of a, but the development of, an increase of the, different areas of, of all people and, a problem of, the ability of a, in a similar way, in different areas of, to the development, an end to the, we have seen that, people think like a, to an increase, to contribute to, a similar way to, and even in, are extremely useful, is actually in contrast, is the ability of a, that knowledge is, an end to, to the human, contrast to increase, contribute to people's, does not bring an, bring an end to,	except for the, that there is a, not be in, in the world, one of the, the number of, the original ones, we try to, one of the biggest, is the easiest, do not have any, argued that they, are the only alternative, it can be, because it is, it is well known, longer than the, the problem of, make a career, they produce more, that even though, the end of the, try to keep, way to produce, what kind of,

Similar to other samples, Lale produced *referential expressions* more frequently than other functional types in each sub-corpus. In sub-corpus 1, as stated before *referential expressions* were the most common with 11 types while *stance expressions* were the second one with 2 types and there was no type of *discourse organizers*. Sub-corpus 2 was similar to sub-corpus 1; it consisted of 8 types of *referential expressions* and 3 types of *stance expressions*. *Discourse organizers* appeared with 2 types. As the other 2 sub-corpora, *referential expressions* were still the most popular (10 types) and it was followed by *stance expressions* (4 types) and then came *discourse organizers* (3 types) in sub-corpus 3. In sub-corpus 4, there was a higher increase in the production of type frequency of formulaic sequence function across each sub-corpus: 36 types of *referential expressions*, 15 types of *stance expressions* and 4 types of *discourse organizers*. Similarly, *referential expressions* were the most recurrent with 25 types and *stance expressions* were the second popular one with 7 types. Sub-corpus 5 was similar to sub-corpus 4 with regard to *discourse organizers* because there were 4 types of it.

The distribution of functional types of formulaic sequences produced by Lale across 5 sub-corpora is depicted in Figure 16 below.

Figure 16: Distribution of Functional Types of Formulaic Sequences Produced by Lale across 5 Sub-corpora



4.3.4.3. Unique Formulaic Sequences in Lale's each Sub-corpus

In sub-corpus 1, Lale shared about 15% of formulaic sequence types with learner corpora and about 8% with LOCNESS. There were nine unique formulaic sequences to her: *given by the, cannot achieve success, cannot distinguish, the subjects that, not understand the subject, of our lives, prepared for the, is not given, reinforce the subject and to be successful.*

In sub-corpus 2, approximately 54% of formulaic sequence types were common with learner corpora, and about 15% of FSs were common with LOCNESS. *Have to have more, more than one, attention to the, have more than one, have to think about and the health of* were unique to her.

In sub-corpus 3, about 47% of formulaic sequence types were common with learner corpora while about 24% of FSs were common with LOCNESS. *In contact with, to live together, live together in, the opportunity to, grow up in, the strengthening of. we must ensure that, up in the and we must free,* were unique to Lale.

In sub-corpus 4, she shared almost 43% of formulaic sequence types with learner corpora and about 7% with LOCNESS. The following twenty-three types of sequences were unique to her: *can lead to, are extremely useful, of a full-scale, an increase of the, it leads to the, does not bring an, people think like a, to people's thoughts is, we have seen that, to an increase, why people think like, to contribute to, in a similar way, in different areas of, will also create a, and even in, is the ability of, may be right, of*

all people and, the ability of a, have no effect on, of the human, to the human, bring an end to, contrast to an increase, contribute to people's, to the development, cannot show, it is always, look at your, will facilitate the and I disagree with.

In sub-corpus 5, about 56% of formulaic sequence types were common with learner corpora, and about 28% of FSs were common with LOCNESS. *The original ones, we try to, in that point, is the easiest, are the only alternative, it is well known, it may even, longer than the, make a career, they produce more, except for the, that even though, try to keep, way to produce, we can easily and what kind of* were unique to her.

Lale's results nearly in each sub-corpus were consistent with the aforementioned other individual learners. The majority of the formulaic sequences were shared on a word-for-word or partial basis with longitudinal learner corpora while the rate of the shared sequence was less with native corpus. The analysis of unique sequences showed that there was a fluctuation in the type and frequency of unique sequences from sub-corpus 1 to sub-corpus 5.

4.3.5. Betül's Inventory of Formulaic Sequences

The top 20 frequently occurring formulaic sequences produced by Betül across 5 sub-corpora of the longitudinal learner corpus are presented in Table 44. Sub-corpus 1 consisted of 1,240 tokens and 8 formulaic sequence types. Sub-corpus 2 consisted of 1,320 tokens and 15 formulaic sequence types. Sub-corpus 3 consisted of 1,661 tokens and 17 formulaic sequence types. Sub-corpus 4 consisted of 1,999 tokens and 37 formulaic sequence types. Sub-corpus 5 consisted of 3,037 and 52 formulaic sequence types.

Table 44: Top 20 Frequent Formulaic Sequences Produced by Betül across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
in terms of	2	80,65	in terms of	4	151,52	in terms of	3	90,31	in terms of	5	125,06	in the world	17	279,88
it seems to me	2	80,65	has gone a	2	75,76	which is a	3	90,31	while a person	4	100,05	in terms of	7	115,25
seems to me that	2	80,65	in the matter of	2	75,76	for the sake of	2	60,20	people who live in	3	75,04	should not be	5	82,32
students who are	2	80,65	it appears that	2	75,76	as well as	2	60,20	a person is	3	75,04	I think that	5	82,32
success for students of	2	80,65	is connected with	2	75,76	it is an	2	60,20	not to mention	3	75,04	cannot be	5	82,32
who study by	2	80,65	it can be said	2	75,76	it is considered	2	60,20	it should be	3	75,04	the rate of	4	65,85
on technology in	2	80,65	can be said that	2	75,76	it is possible	2	60,20	firmly believe that	2	50,03	which has been	3	49,39
in the areas of	2	80,65	thanks to +noun	2	75,76	serve as a	2	60,20	have several issues	2	50,03	in same positions	3	49,39
			the data of	2	75,76	being in the	2	60,20	he claims that	2	50,03	does not cause	3	49,39
			one of the	2	75,76	so they may	2	60,20	he said that	2	50,03	cannot be ignored	3	49,39
			to grow their	2	75,76	to see that	2	60,20	according to the	2	50,03	the fact that	3	49,39
			where have the	2	75,76	to supply a	2	60,20	I mostly object	2	50,03	they want to	3	49,39
			point of view	2	75,76	when it is	2	60,20	I recognize him that	2	50,03	who work in	3	49,39
			cannot be	2	75,76	with each other	2	60,20	I think that	2	50,03	does not have	2	32,93
			according to the	2	75,76	in the same	2	60,20	if a person	2	50,03	even if the	2	32,93
						fact that the	2	60,20	in my opinion	2	50,03	I agree with him	2	32,93
						according to a	2	60,20	the cheapest thing	2	50,03	by the way	2	32,93
									is a tool to	2	50,03	in the same	2	32,93
									is clear that the	2	50,03	he ignores the fact	2	32,93
									I agree with	2	50,03	one of the	2	32,93

4.3.5.1. Structures of Formulaic Sequences in Betül's each Sub-corpus

All the structural formulaic sequence types produced by Betül in her each sub-corpus are presented in Table 45.

Table 45: Structure of Formulaic Sequences in Betül's each Sub-corpus

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
Personal pronoun + verb phrase (+ complement-clause fragment)			<i>so they may,</i>	<i>a person is, he claims that, he said that, I mostly object, I recognize him that, I think that, I agree with, and they will, I firmly believe that, a person should be, she should not, I have several,</i>	<i>I think that, he claimed that, they want to, I agree with him, he ignores the fact, when I consider,</i>
Verb phrase with active verb		<i>has gone a, cannot be,</i>	<i>serve as a,</i>	<i>have several issues, recognize him that there, should not be,</i>	<i>should not be, cannot be, does not cause, cannot be ignored, does not have, produced to increase the, do not want to, did not have, getting a high, has been performed, have an apprehension,</i>
FSs with wh-clause fragments	<i>students who are, who study by,</i>	<i>where have the,</i>	<i>which is a,</i>	<i>while a person, people who live in, which is used in, who is an, who is known,</i>	<i>which has been, who work in, who is a, whose name is,</i>
Noun phrase with of-phrase fragment	<i>success for students of, in the areas of,</i>	<i>in the matter of, the data of, one of the, point of view,</i>	<i>for the sake of,</i>	<i>one of the,</i>	<i>the rate of, one of the, the cost of, the number of, the only reason of, the reason of,</i>
Noun phrase with other post-modifier fragment					<i>the fact that,</i>
Prepositional phrase with embedded of-phrase fragment	<i>in terms of,</i>	<i>in terms of,</i>	<i>in terms of,</i>	<i>in terms of,</i>	<i>in terms of,</i>
Other prepositional phrase fragment	<i>on technology in,</i>		<i>being in the, with each other, in the same,</i>	<i>in my opinion, on the department of,</i>	<i>in the same, on the department of, on the journal of, in the world, in same positions, by the way, in addition to,</i>
Anticipatory it + verb phrase/adjective phrase	<i>it seems to me,</i>	<i>it appears that, it can be said,</i>	<i>it is an, it is considered, it is possible, when it is,</i>	<i>it should be,</i>	<i>it does not, it has right, it is quite, it is undeniable fact,</i>
Copula be + noun phrase/adjective phrase		<i>is connected with,</i>		<i>is a tool to,</i>	<i>is being produced to, is not the, be an obstacle,</i>
(verb phrase +) that-clause fragment	<i>seems to me that,</i>	<i>can be said that,</i>	<i>to see that, fact that the,</i>	<i>is clear that the, firmly believe that, it is clear that, that even if, undeniable fact that,</i>	<i>is undeniable fact that, change the reality that,</i>

Table 45: (Continue)

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
(verb/adjective +) to-clause fragment		<i>to grow their,</i>	<i>to supply a,</i>	<i>not to mention,</i>	<i>to be a, to increase the number,</i>
Pronoun/noun phrase + be (+ . . .)				<i>that there is a, there is a point,</i>	
Other expressions		<i>thanks to +noun, according to the,</i>	<i>as well as, according to a,</i>	<i>according to the,</i>	<i>as being a, according to the, as well as, rather than a,</i>
If-clause fragments				<i>if a person,</i>	<i>even if the,</i>
Comparative expressions				<i>the cheapest thing,</i>	

In sub-corpus 1, Betül produced 6 types of formulaic sequence structures. The most utilized structural types were *noun phrase with of-phrase fragment* and *with wh-clause fragments*. She produced 2 different types of this structure with a total of 4 token frequencies, followed by 1 type of *prepositional phrase with embedded of-phrase fragment*, *other prepositional phrase fragment*, *anticipatory it + verb phrase/adjective phrase* and *(verb phrase +) that-clause fragment* with a total of 2 token frequencies.

In sub-corpus 2, Betül produced 9 types of formulaic sequence structures. The most utilized structural type was *noun phrase with of-phrase fragment* was still the most frequent in terms of type (4) and token (8). It was followed by 2 types of *verb phrase with active verb*, *anticipatory it + verb phrase/adjective phrase* and *other expressions* with a total of 4, 4 and 6 token frequencies, respectively.

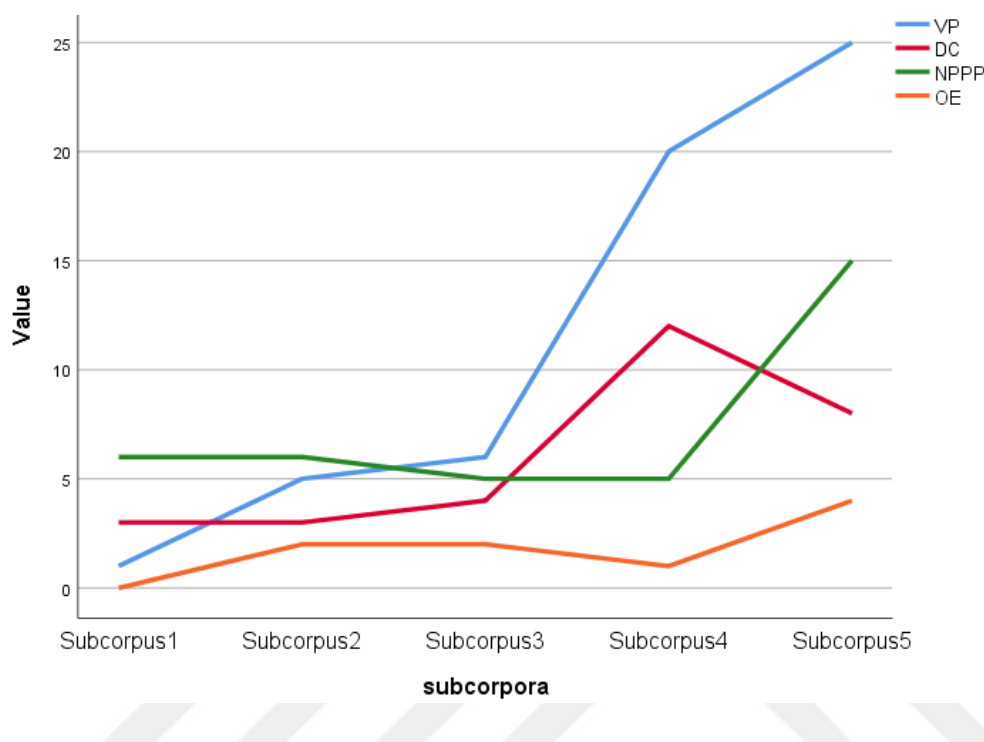
In sub-corpus 3, Betül produced 10 types of formulaic sequence structures. Different from the other 2 sub-corpora, the most frequent structural type was *anticipatory it + verb phrase/adjective phrase*. She produced 4 different types of this structure with a total of 8 token frequencies, followed by 3 types of *other prepositional phrase fragment* with a total of 6 token frequencies.

In sub-corpus 4, Betül produced 14 types of formulaic sequence structures. The most popular structural type was *personal pronoun + verb phrase (+ complement-clause fragment)*. She produced 12 different types of these structures with a total of 24 token frequencies, followed by 5 types of *(verb phrase +) that-clause fragment* with a total of 10 token frequencies.

In sub-corpus 5, Betül produced 13 types of formulaic sequence structures. The most utilized structural type was *verb phrase with active verb*. She produced 11 different types of this structure with a total of 27 token frequencies, followed by 7 types of *other prepositional phrase fragment* with a total of 30 token frequencies.

The distribution of structural types of formulaic sequences produced by Betül across 5 sub-corpora is shown in Figure 17 below.

Figure 17: Distribution of Structural Types of Formulaic Sequences Produced by Betül across 5 Sub-corpora



4.3.5.2. Functions of Formulaic Sequences in Betül's each Sub-corpus

Tables 46 presents the different functional types of the formulaic sequences in each sub-corpus.

Table 46: Function of Formulaic Sequences in Betül's each Sub-corpus

Functional Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
1. Stance expressions		<i>has gone a, can be said that, according to the, cannot be,</i>	<i>fact that the, according to a,</i>	<i>I mostly object, I think that, I agree with, and they will, I firmly believe that, a person should be, she should not, firmly believe that, according to the, should not be, in my opinion,</i>	<i>I think that, they want to, I agree with him, Should not be, cannot be, cannot be ignored, do not want to, the fact that, according to the, to be a,</i>
2. Discourse organizers			<i>for the sake of, as well as,</i>		<i>by the way, in addition to, rather than a, even if the, as well as,</i>

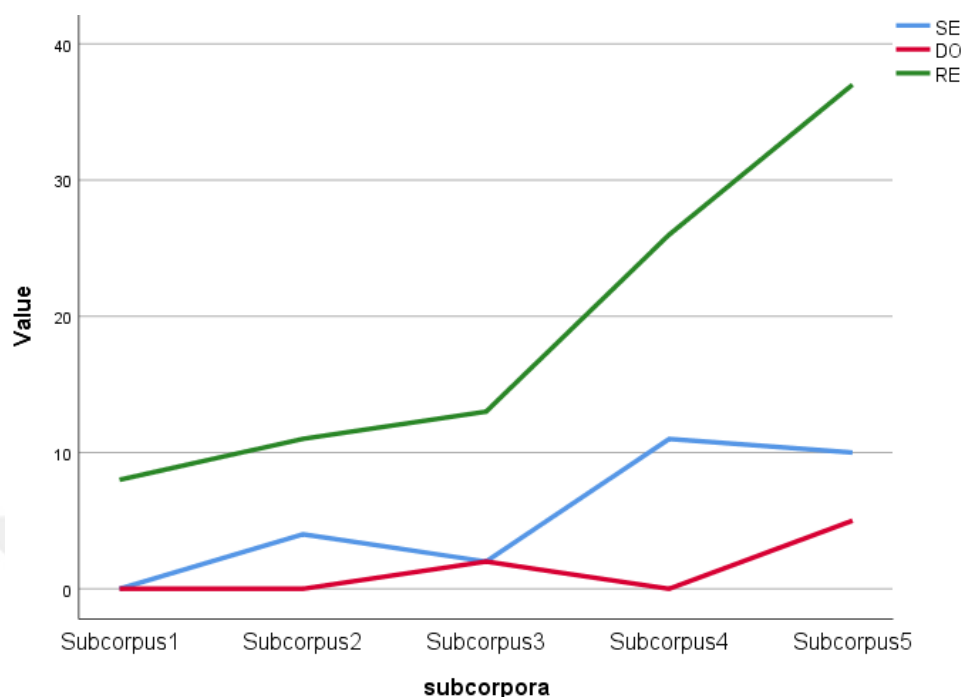
Table 46: (Continue)

Functional Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
3.Referential expressions	<p>students who are, who study by, it seems to me, in the areas of, in terms of, on technology in, success for students of, seems to me that,</p>	<p>where have the, it appears that, it can be said, thanks to +noun, in the matter of, the data of, one of the, point of view, in terms of, is connected with, to grow their,</p>	<p>which is a, with each other, in the same, it is an, it is considered, it is possible, when it is, in terms of, so they may, serve as a, being in the, to see that, to supply a,</p>	<p>people who live in, which is used in, who is an, who is known, one of the, it should be, is a tool to, it is clear that, that there is a, the cheapest thing, there is a point, in terms of, on the department of, a person is, he claims that, he said that, if a person, that even if, undeniable fact that, not to mention, I have several, have several issues, recognize him that there, while a person, is clear that the, I recognize him that,</p>	<p>he claimed that, he ignores the fact, when I consider, does not cause, does not have, produced to increase the, did not have, getting a high, has been performed, have an apprehension, which has been, who work in, who is a, whose name is, the rate of, one of the, the cost of, the number of, the only reason of, the reason of, in terms of, in the same, on the department of, on the journal of, in the world, in same positions, it does not, it has right, it is quite, it is undeniable fact, is being produced to, is not the, be an obstacle, is undeniable fact that, change the reality that, to increase the number, as being a,</p>

Referential expressions were the most recurrent in each sub-corpus. In sub-corpus 1, Betül produced 8 types of *referential expressions*, but there was no type in both *stance expressions* and *discourse organizers*. In sub-corpus 2, the amount of type in both *referential expressions* (11 types) and *stance expressions* (4 types) increased. In the same vein with sub-corpus 1, there was no type in *discourse organizers*. Similarly, *referential expressions* were the most common with 13 types in sub-corpus 3, and then came *stance expressions* (2 types). *Discourse organizers* appeared with 2 types in this sub-corpus. In sub-corpus 4, there was an increase in the production of type frequency of the formulaic sequence functions of the category *referential expressions* across each sub-corpus: 26 types of *referential expressions*. There were 11 types of *stance expressions*, and there was no type of *discourse organizers*. She used a smaller amount of *stance expressions* (10 types) whereas there were 37 types of *referential expressions* and 5 types of *discourse organizers* in sub-corpus 5.

The distribution of functional types of formulaic sequences produced by Betül across 5 sub-corpora is shown in Figure 18 below.

Figure 18: Distribution of Functional Types of Formulaic Sequences Produced by Betül across 5 Sub-corpora



4.3.5.3. Unique Formulaic Sequences in Betül's each Sub-corpus

In sub-corpus 1, Betül shared about 12% of formulaic sequence types with learner corpora whereas there was no shared pattern with LOCNESS on a word-for-word or partial basis. *It seems to me, success for students of, who study by, seems to me that, students who are, on technology in and in the areas of* were unique formulaic sequence to her.

In sub-corpus 2, she shared approximately 53% of formulaic sequence types with learner corpora and about 13% with LOCNESS. *Has gone a, in the matter of, it appears that, is connected with, the data of, to grow their and where have the* were unique to her.

In sub-corpus 3, about 59% of formulaic sequence types were common with learner corpora, and about 12% of FSs were common with LOCNESS. *For the sake of, serve as a, being in the, so they may, it is considered, to see that and to supply a* were unique to her.

In sub-corpus 4, she shared almost 51% of formulaic sequence types with learner corpora and about 14% with LOCNESS. The following ten types of sequences were unique to her: *I mostly object, I recognize him that, the cheapest thing, is a tool to, is clear that the, recognize him that there, there is a point, which is used in, while a person, a person is, not to mention, if a person, she should not, that even if, who is known, a person should be, it is clear that and on the department of.*

In sub-corpus 5, about 60% of formulaic sequence types were common with learner corpora while about 19% of FSs were common with LOCNESS. *In same positions, does not cause, cannot be ignored, who work in, he ignores the fact, rather than a, is being produced to, as being a, it has right, it is quite, change the reality that, when I consider, on the department of, on the journal of, the only reason of, the reason of, getting a high, be an obstacle, has been performed, whose name is and have an apprehension* were unique to her.

The results indicated that in each sub-corpus except for sub-corpus 1, more than half of the formulaic sequences were shared on a word-for-word or partial basis with longitudinal learner corpora. There were no common sequences in sub-corpus 1 whereas the proportion of shared sequences were less with native corpus than longitudinal learner corpora. The analysis of unique sequences showed that the type and frequency of unique sequences were similar numbers from sub-corpus 1 to sub-corpus 3 but increased slowly after then.

4.3.6. Ali's Inventory of Formulaic Sequences

The top 20 frequently occurring formulaic sequences produced by Ali across 5 sub-corpora of the longitudinal learner corpus are presented in Table 47. Sub-corpus 1 consisted of 1,367 tokens and 13 formulaic sequence types. Sub-corpus 2 consisted of 1,556 tokens and 30 formulaic sequence types. Sub-corpus 3 consisted of 2,103 tokens and 44 formulaic sequence types. Sub-corpus 4 consisted of 2,184 tokens and 58 formulaic sequence types. Sub-corpus 5 consisted of 2,666 and 42 formulaic sequence types.

Table 47: Top 20 Frequent Formulaic Sequences Produced by Ali across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
it is easy to	3	109,73	are the best	5	160,67	it should be	4	95,10	in the field of	4	91,58	in the world	9	168,79
are the most	3	109,73	is an undeniable fact	3	96,40	to create a	4	95,10	one of the	4	91,58	one of the	5	93,77
in order to	2	73,15	as they are	3	96,40	one of the	4	95,10	the most important	3	68,68	of the world	4	75,02
is easy to say	2	73,15	it should be	3	96,40	the most substantial	4	95,10	can be seen	3	68,68	in his book that	4	75,02
do my review	2	73,15	they can be	3	96,40	make people live in	3	71,33	it is not	3	68,68	in order to	4	75,02
get used to	2	73,15	according to this	3	96,40	is one of the	3	71,33	it is a	3	68,68	with the help of	3	56,26
claim that they	2	73,15	fond of to their	2	64,27	one of the most	3	71,33	it is the	3	68,68	can be a	2	37,51
how to use	2	73,15	be taken into consideration	2	64,27	in order to	3	71,33	would not be a	3	68,68	her study that	2	37,51
on the internet	2	73,15	an undeniable fact that	2	64,27	who live in	3	71,33	can eliminate that	2	45,79	cannot afford to	2	37,51
should be noted that	2	73,15	of this series	2	64,27	do not have	3	71,33	according to a	2	45,79	and they are	2	37,51
that they are	2	73,15	pay attention to	2	64,27	to make people	3	71,33	a person in	2	45,79	according to her	2	37,51
I strongly believe that	2	73,15	taken into consideration that	2	64,27	and it makes them	2	47,55	chance to save	2	45,79	cannot be a	2	37,51
as it is	2	73,15	have a chance to	2	64,27	have a problem of	2	47,55	do feel that	2	45,79	it is a	2	37,51
			have you ever	2	64,27	and make them	2	47,55	do not have	2	45,79	I strongly believe that	2	37,51
			I strongly believe that	2	64,27	according to this	2	47,55	fails to mention that	2	45,79	much more money	2	37,51
			according to a	2	64,27	how is it	2	47,55	at first appearance	2	45,79	I think that	2	37,51
			is necessary for	2	64,27	I strongly believe that	2	47,55	fill up one's	2	45,79	are equal in	2	37,51
			it is a	2	64,27	as a result of	2	47,55	be just a	2	45,79	a solution to	2	37,51
			it is an	2	64,27	a place where	2	47,55	an undeniable fact that	2	45,79	not only for	2	37,51
			can be a	2	64,27	a problem of	2	47,55	gives a chance to	2	45,79	on the same	2	37,51

4.3.6.1. Structures of Formulaic Sequences in Ali's each Sub-corpus

Tables 48 presents the different structural types of the formulaic sequences in each sub-corpus.

Table 48: Structure of Formulaic Sequences in Ali's each Sub-corpus

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
Personal pronoun + verb phrase (+ complement-clause fragment)	<i>I strongly believe that,</i>	<i>as they are, they can be, I strongly believe that,</i>	<i>I strongly believe that, they do not, people cannot, they do not have, they have an,</i>	<i>he fails to mention, I do feel that, I have several, the study shows, while I do feel,</i>	<i>and they are, I strongly believe that, I think that,</i>
Verb phrase with active verb	<i>do my review, get used to,</i>	<i>pay attention to, have a chance to, can be a, achieve their dreams, should be taken into, can be their,</i>	<i>make people live in, do not have, have a problem of, and make them, create a place, and can easily, plays a big role, should be taken into, spend time in, has a vital importance,</i>	<i>would not be a, do not have, fill up one's, gives a chance to, not have any, has failed to, have been made,</i>	<i>can be a, cannot afford to, cannot be a,</i>
FSs with wh-clause fragments		<i>have you ever,</i>	<i>who live in, how is it, a place where,</i>	<i>who is a, who is an,</i>	<i>people who live in,</i>
Quantifier expressions		<i>although many people,</i>			
Noun phrase with of-phrase fragment		<i>of this series, as a result of,</i>	<i>one of the, one of the most, as a result of, a problem of, of the most substantial,</i>	<i>one of the, the vast majority of, of the most, one of the most, point of view, these thoughts of,</i>	<i>because of their, one of the, of the world, parts of the, point of view, possible effects of, production of the, the cost of, these kind of,</i>
Other noun phrase expressions			<i>the most substantial, the most important,</i>	<i>the most important,</i>	
Prepositional phrase with embedded of-phrase fragment				<i>in the field of,</i>	
Other prepositional phrase fragment	<i>in order to, on the internet,</i>		<i>in order to, a big role in,</i>	<i>a person in, at first appearance, on the other hand, vital importance in, a study from the,</i>	<i>rise in the, in the same, in the whole, same conditions in their, on the same, a study from, in the world, in order to, with the help of, in a long term,</i>
Anticipatory it + verb phrase/adjective phrase	<i>it is easy to, it should be, as it is,</i>	<i>it should be, it is a, it is an, it is an undeniable, it is easy to,</i>	<i>it should be, and it makes them, it is possible,</i>	<i>it is not, it is a, it is the, it can be, it helps to, it is the most, it may have some, it should be taken, but it should,</i>	<i>it is a,</i>

Table 48: (Continue)

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
Passive verb + prepositional phrase fragment		<i>be taken into consideration,</i>		<i>can be seen,</i>	<i>based on my,</i>
Copula be + noun phrase/adjective phrase	<i>are the most, is easy to say,</i>	<i>are the best, is an undeniable fact, is necessary for, are the most,</i>	<i>is one of the, is an undeniable fact, is and how, is the biggest, is the most,</i>	<i>be just a, is not a, is the most important,</i>	<i>are equal in, be a solution, is not a,</i>
(verb phrase +) that-clause fragment	<i>claim that they, should be noted that, that they are</i>	<i>an undeniable fact that, taken into consideration that, it is essential that,</i>	<i>an undeniable fact that, should be noted that, take into consideration that, a place that,</i>	<i>can eliminate that, do feel that, fails to mention that, an undeniable fact that, kinds of stuff that, taken into consideration that, that is going to,</i>	<i>her study that, in his book that, that there are,</i>
(verb/adjective +) to-clause fragment	<i>how to use,</i>	<i>a chance to, fond of to their,</i>	<i>to create a, to make people, to learn the, to live in,</i>	<i>chance to save, going to be just, to keep a person, to manage money, a chance to,</i>	<i>a solution to,</i>
Pronoun/noun phrase + be (+ . . .)		<i>there are some,</i>			<i>there is no, there is a, there is no discrimination,</i>
Other expressions		<i>according to this, according to a,</i>	<i>according to this,</i>	<i>according to a,</i>	<i>according to her, not only for, the best way,</i>
If-clause fragments				<i>and if you can, but if one, if it is, if you can eliminate,</i>	
Comparative expressions			<i>than other as they,</i>	<i>the strongest claims,</i>	<i>much more money,</i>

In sub-corpus 1, Ali produced 7 types of formulaic sequence structures. The most utilized structural types were *anticipatory it + verb phrase/adjective phrase* and *(verb phrase +) that-clause fragment*. He produced 3 different types of these structures with a total of 7 token frequencies.

In sub-corpus 2, Ali produced 12 types of formulaic sequence structures. The most utilized structural type was *verb phrase with active verb*. She produced 6 different types of these structures with a total of 12 token frequencies, followed by 5 types of *anticipatory it + verb phrase/adjective phrase* with a total of 11 token frequencies.

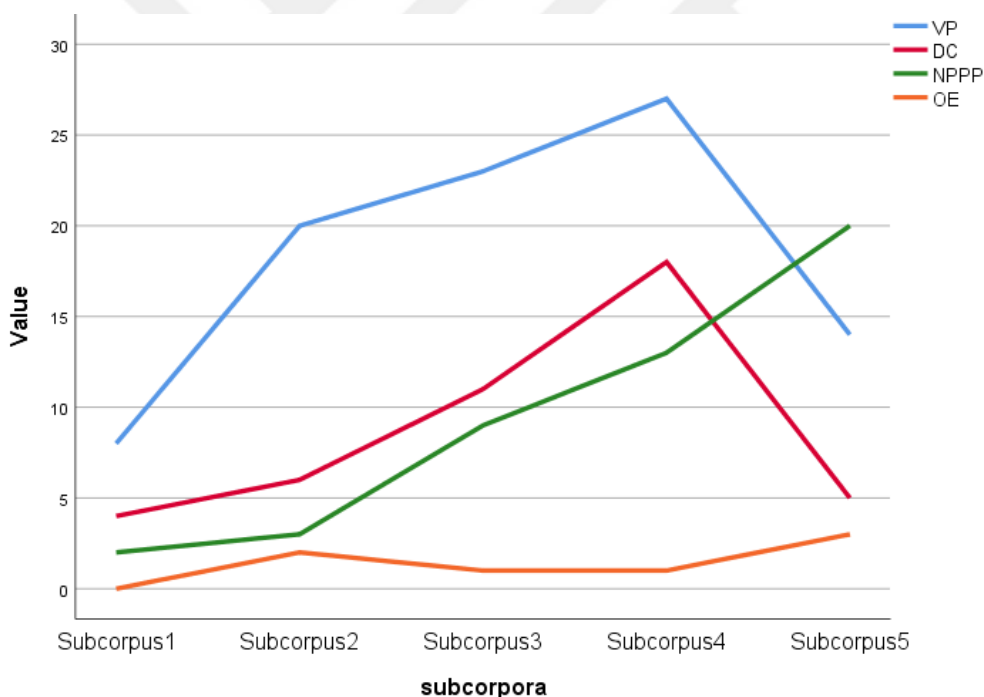
In sub-corpus 3, Ali produced 12 types of formulaic sequence structures. Similar to sub-corpus 2, the most frequent structural type was *verb phrase with active verb*. She produced 10 different types of this structure with a total of 22 token frequencies, followed by 5 types of *noun phrase with of-phrase fragment*, *copula be + noun phrase/adjective phrase* and *(verb phrase +) that-clause fragment*.

In sub-corpus 4, Ali produced 15 types of formulaic sequence structures. *Verb phrase with active verb* was still one of the most frequent in terms of type (7) and token (16) and *anticipatory it + verb phrase/adjective phrase* was another popular one with 9 different types and a total of 21 token frequencies. *Noun phrase with of-phrase fragment, other prepositional phrase fragment, (verb/adjective +) to-clause fragment* and *(verb phrase +) that-clause fragment* were the followings.

In sub-corpus 5, Ali produced 13 types of formulaic sequence structures. The most frequent structural type was *other prepositional phrase fragment*. She produced 10 different types of this structure with a total of 28 token frequencies, followed by 9 types of *noun phrase with of-phrase fragment* with a total of 23 token frequencies.

The distribution of structural types of formulaic sequences produced by Ali across 5 sub-corpora is presented in Figure 19 below.

Figure 19: Distribution of Structural Types of Formulaic Sequences Produced by Ali across 5 Sub-corpora



4.3.6.2. Functions of Formulaic Sequences in Ali's each Sub-corpus

All the functional types of formulaic sequences in each sub-corpus are presented in Table 49.

Table 49: Function of Formulaic Sequences in Ali's each Sub-corpus

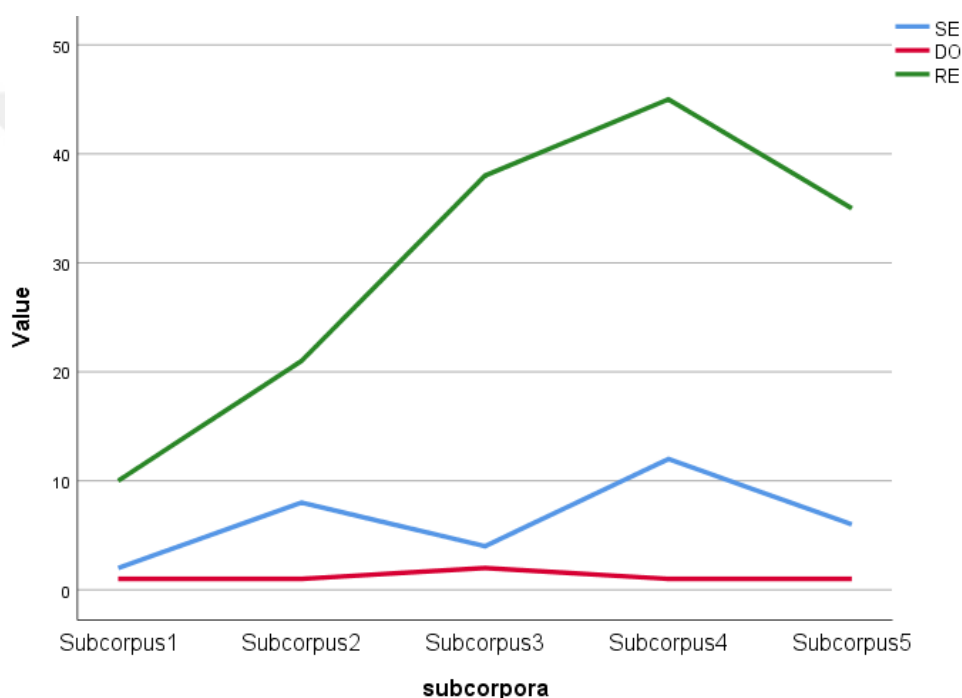
Functional Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
1. Stance expressions	<i>I strongly believe that, how to use,</i>	<i>they can be, I strongly believe that, can be a, achieve their dreams, should be taken into, can be their, according to this, according to a,</i>	<i>I strongly believe that, should be taken into, according to this, people cannot,</i>	<i>I do feel that, while I do feel, would not be a, these thoughts of, can be seen, can eliminate that, do feel that, that is going to, going to be just, if you can eliminate, according to a, and if you can,</i>	<i>I strongly believe that, I think that, can be a, cannot afford to, cannot be a, according to her,</i>
2. Discourse organizers	<i>in order to,</i>	<i>as a result of,</i>	<i>as a result of, in order to,</i>	<i>on the other hand,</i>	<i>in order to,</i>
3. Referential expressions	<i>it is easy to, it should be, as it is, are the most, should be noted that, that they are, do my review, get used to, on the internet, is easy to say, claim that they</i>	<i>it should be, it is a, it is an, it is an undeniable, it is easy to, are the best, is necessary for, are the most, there are some, although many people, a chance to, fond of to their, as they are, pay attention to, have a chance to, have you ever, of this series, be taken into consideration, is an undeniable fact, an undeniable fact that, taken into consideration that, it is essential that,</i>	<i>they do not, who live in, how is it, a place where, one of the, one of the most, of the most substantial, the most substantial, the most important, a big role in, it should be, and it makes them, it is possible, is one of the, is and how, is the biggest, is the most, a place that, should be noted that, take into consideration that, a problem of, they do not have, they have an, make people live in, do not have, have a problem of, and make them, create a place, and can easily, plays a big role, spend time in, has a vital importance, is an undeniable fact, an undeniable fact that, to create a, to make people, to learn the, to live in, than other as they,</i>	<i>who is a, who is an, one of the, of the most, one of the most, the most important, a person in, at first appearance, vital importance in, a study from the, it is not, it is a, it is the, it can be, it helps to, it is the most, it may have some, it should be taken, but it should, be just a, is not a, is the most important, but if one, if it is, the strongest claims, the vast majority of, point of view, in the field of, he fails to mention, I have several, the study shows, to keep a person, to manage money, a chance to, kinds of stuff that, chance to save, fails to mention that, an undeniable fact that, taken into consideration that, do not have, fill up one's, gives a chance to, not have any, has failed to, have been made,</i>	<i>and they are, people who live in, because of their, one of the, of the world, parts of the, point of view, possible effects of, production of the, the cost of, these kind of, rise in the, in the same, in the whole, same conditions in their, on the same, a study from, in the world, with the help of, in a long term, it is a, based on my, are equal in, be a solution, is not a, her study that, in his book that, that there are, a solution to, there is no, there is a, there is no discrimination, not only for, the best way, much more money,</i>

Referential expressions were the most frequent across each sub-corpus followed by *stance expressions* and *discourse organizers*. In sub-corpus 1, Ali produced 10 types of *referential expressions*, 2 types of *stance expressions* and 1 type of *discourse organizers*. In sub-corpus 2, there was an increase in the amount of *referential expressions* and *stance expressions*. He produced 21 types of *referential expressions*, 8 types of *stance expressions* and then came *discourse organizers* with 1 type. Similarly, in sub-corpus 3, there was an increase that *referential expressions* were the most recurrent with 38 types while *stance expressions* were the second one with 4 types. The last one was *discourse organizers* with

2 types of it. The steady increase in the amount of *stance expressions* and *referential expressions* occurred last time in sub-corpus 4. He produced 45 types of *referential expressions* and 12 types of *stance expressions*. Similar to sub-corpus 1 and 2, *discourse organizers* had 1 type. Lastly, in sub-corpus 5, *referential expressions* were the most popular with 35 types and it was followed by *stance expressions* (6 types). There was 1 type in *discourse organizers*.

The distribution of functional types of formulaic sequences produced by Ali across 5 sub-corpora is presented in Figure 20 below.

Figure 20: Distribution of Functional Types of Formulaic Sequences Produced by Ali across 5 Sub-corpora



4.3.6.3. Unique Formulaic Sequences in Ali's each Sub-corpus

In sub-corpus 1, Ali shared about 54% of formulaic sequence types with learner corpora and about 23% with LOCNESS on a word-for-word or partial basis. *It is easy to, is easy to say, do my review, get used to, on the internet* and *should be noted that* were unique formulaic sequence to her.

In sub-corpus 2, he shared approximately 77% of formulaic sequence types with learner corpora and about 3% with LOCNESS. *Fond of to their, of this series, can be their, it is easy to, although many people, achieve their dreams* and *it is essential that* were unique to him.

In sub-corpus 3, about 61% of formulaic sequence types were common with learner corpora, and about 11% of FSs were common with LOCNESS. *The most substantial, make people live in, to make*

people, and make them, how is it, a big role in, is and how, of the most substantial, is the biggest, create a place, and can easily, plays a big role, a place where, they have an, a place that, to learn the and has a vital importance were unique to him.

In sub-corpus 4, he shared almost 66% of formulaic sequence types with learner corpora and about 22% with LOCNESS. The following sixteen types of sequences were unique to him: *fill up one's, can eliminate that, chance to save, be just a, going to be just, it may have some, kinds of stuff that, have been made, the strongest claims, the study shows, to manage money, a person in, it helps to, not have any, but if one, these thoughts of, to keep a person, if you can eliminate, vital importance in and a study from the.*

In sub-corpus 5, about 74% of formulaic sequence types were common with learner corpora while about 26% of FSs were common with LOCNESS. *In his book that, her study that, much more money, not only for, on the same, in a long term, possible effects of, rise in the, production of the, same conditions in their and in the whole* were unique to him.

The results were consistent with aforementioned other particular learners that the majority of the formulaic sequences were shared on a word-for-word or partial basis with longitudinal learner corpora whereas the proportion of shared sequences were less with native corpus. The analysis of unique sequences revealed that there was a continuous increase in the type and frequency of unique sequences from sub-corpus 1 to sub-corpus 4, but in sub-corpus 5, it decreased in number.

4.3.7. Ash's Inventory of Formulaic Sequences

The top 20 frequently occurring formulaic sequences produced by Ash across 5 sub-corpora of the longitudinal learner corpus are presented in Table 50. Sub-corpus 1 consisted of 1,066 tokens and 14 formulaic sequence types. Sub-corpus 2 consisted of 1,272 tokens and 17 formulaic sequence types. Sub-corpus 3 consisted of 1,645 tokens and 11 formulaic sequence types. Sub-corpus 4 consisted of 3,962 tokens and 77 formulaic sequence types. Sub-corpus 5 consisted of 2,410 and 42 formulaic sequence types.

Table 50: Top 20 Frequent Formulaic Sequences Produced by Ashi across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
day by day	3	140,71	on the other hand	4	157,23	day by day	2	60,79	day by day	6	75,72	one of the	8	165,98
a result of this	3	140,71	day by day	2	78,62	even if they	2	60,79	in the future	6	75,72	in the world	5	103,73
the reason why	3	140,71	because of this	2	78,62	of nature for	2	60,79	I do not	6	75,72	this kind of	4	82,99
as a result of	3	140,71	one of the	2	78,62	on the contrary	2	60,79	people avoid to	6	75,72	in terms of	3	62,24
did not finish her	2	93,81	people cannot	2	78,62	on the other hand	2	60,79	is the most	6	75,72	the other biggest claims	3	62,24
had a daily	2	93,81	people have a	2	78,62	they do not	2	60,79	people do not	6	75,72	be a solution to	3	62,24
begin to use	2	93,81	level of their	2	78,62	who live in	2	60,79	in the same	6	75,72	I do not agree	2	41,49
and we begin to	2	93,81	reason is about	2	78,62	in ministry of	2	60,79	one of the	6	75,72	I have found that	2	41,49
we start to use	2	93,81	rising of personal	2	78,62	far away from	2	60,79	he said because of	4	50,48	can overcome all	2	41,49
is one of the	2	93,81	lots of children	2	78,62	because of this	2	60,79	he said that	4	50,48	cannot overcome	2	41,49
thanks to daily	2	93,81	most of people	2	78,62	as in the	2	60,79	a kind of	4	50,48	in some part of	2	41,49
of the most	2	93,81	stay with their	2	78,62				the reason why	4	50,48	claims of the	2	41,49
we can talk	2	93,81	think that grandparents	2	78,62				we are against	4	50,48	in the developing world	2	41,49
one of the most	2	93,81	who name is	2	78,62				in addition to	4	50,48	as a result	2	41,49
			what do you think	2	78,62				part of turkey	4	50,48	around the world	2	41,49
			cannot afford	2	78,62				cannot find a	4	50,48	the cause of	2	41,49
			both of them	2	78,62				whole of this	4	50,48	in the future	2	41,49
									are against each other	4	50,48	the most known	2	41,49
									people think that	4	50,48	is not a	2	41,49
									take care of	4	50,48	to the world's	2	41,49

4.3.7.1. Structures of Formulaic Sequences in Asl's each Sub-corpus

Tables 51 presents the different structural types of the formulaic sequences produced by Asl in each sub-corpus.

Table 51: Structure of Formulaic Sequences in Asl's each Sub-corpus

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
Personal pronoun + verb phrase (+ complement-clause fragment)	<i>and we begin to, we start to use, we can talk,</i>	<i>people cannot, people have a,</i>	<i>they do not,</i>	<i>I do not, people do not, he said because of, he said that, we are against, they do not, when I look, he said because, and he is aware, and I think he, and he adds, and he answered that, and he can understand, and he is, and he warned, and I found, and I look a, and I suggest you, and I think, and people became, again we are, and people make a, and they are, and they cannot, and they do not, and they will not, and he said that,</i>	<i>he does not, he supports these, I do not agree, and they can, she is the, I have found that,</i>
Verb phrase with active verb	<i>did not finish her, had a daily, begin to use,</i>	<i>stay with their, cannot afford,</i>		<i>cannot find a, take care of, start to give, will be the, added people to start, and make an agreement, agree with them,</i>	<i>cannot overcome, can overcome all, not want to, will be a, will be more, will emerge in, cannot be,</i>
FSs with wh-clause fragments		<i>what do you think, who name is,</i>	<i>who live in,</i>	<i>which is a,</i>	<i>who are the, who is a,</i>
Quantifier expressions		<i>lots of children,</i>		<i>most people still,</i>	
Noun phrase with of-phrase fragment	<i>a result of this, as a result of, one of the most,</i>	<i>because of this, one of the, level of their, rising of personal, most of people, both of them,</i>	<i>of nature for, because of this,</i>	<i>one of the, a kind of, part of turkey, whole of this, the end of, because of the, the number of, the rate of, a result of this, and the result of,</i>	<i>this kind of, one of the, in some part of, claims of the, the cause of, because instead of, of the report, of the world,</i>
Other noun phrase expressions	<i>the reason why,</i>			<i>the reason why, and the most,</i>	<i>the most known,</i>
Prepositional phrase with embedded of-phrase fragment				<i>in terms of,</i>	<i>in terms of,</i>

Table 51: (Continue)

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
Other prepositional phrase fragment	<i>thanks to daily, of the most,</i>	<i>on the other hand,</i>	<i>on the contrary, on the other hand, in ministry of, far away from, as in the,</i>	<i>in the same, in addition to, on the contrary, a research about which, against each other because, against each other he, a big deal for, at the beginning,</i>	<i>in the world, in the developing world, around the world, in the future, on the contrary,</i>
Anticipatory it + verb phrase/adjective phrase				<i>and it is obvious, and it was,</i>	<i>it does not,</i>
Copula be + noun phrase/adjective phrase	<i>is one of the,</i>	<i>reason is about,</i>		<i>is the most, are against each other, be the winner,</i>	<i>is the most, be a solution to, is not a, are completely safe,</i>
(verb phrase +) that-clause fragment		<i>think that grandparents,</i>		<i>it is obvious that, people think that, strongly believe that,</i>	<i>said that women, report concluded that, the report is that,</i>
(verb/adjective +) to-clause fragment				<i>people avoid to, and continue to, and start to,</i>	<i>to the world's, a solution to the,</i>
Adverbial clause fragment	<i>day by day,</i>	<i>day by day,</i>	<i>day by day,</i>	<i>day by day, in the future,</i>	
Other expressions				<i>because according to, according to his, according to my, according to interview, and for a while,</i>	<i>the other biggest claims, as a result,</i>
If-clause fragments			<i>even if they,</i>	<i>and if it is, and if people do,</i>	
Comparative expressions				<i>be stronger than,</i>	

In sub-corpus 1, Aslı produced 7 types of formulaic sequence structures. The followings were the most popular structural types: *personal pronoun + verb phrase (+ complement-clause fragment)*, *verb phrase with active verb* and *noun phrase with of-phrase fragment*. He produced 3 different types of these structures with a total of 6,6 and 8 token frequencies, respectively.

In sub-corpus 2, Aslı produced 9 types of formulaic sequence structures. *Noun phrase with of-phrase fragment* was still the most frequent in terms of type (6) and token (12). It was followed by 5 types of *anticipatory it + verb phrase/adjective phrase* with a total of 11 token frequencies.

In sub-corpus 3, Aslı produced 6 types of formulaic sequence structures. Different from the other 2 sub-corpora, the most utilized structural type was *other prepositional phrase fragment*. She produced

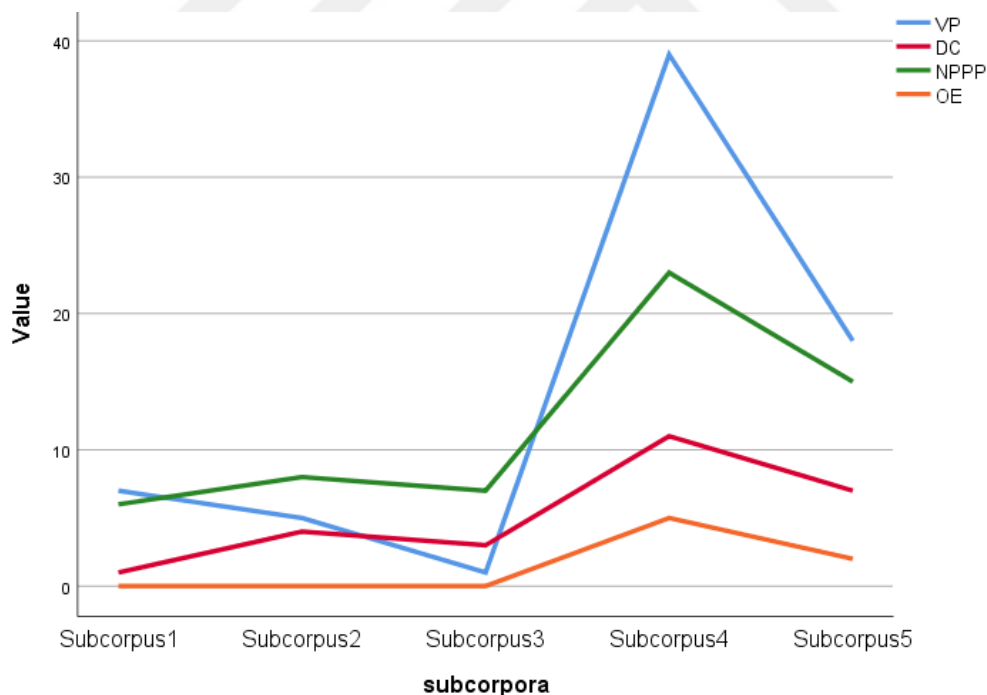
5 different types of this structure with a total of 10 token frequencies, followed by 2 types of *noun phrase with of-phrase fragment* with a total of 4 token frequencies.

In sub-corpus 4, Aslı produced 16 types of formulaic sequence structures. The most popular structural type was *personal pronoun + verb phrase (+ complement-clause fragment)* with 27 different types and a total of 74 token frequencies. *Noun phrase with of-phrase fragment* was the following one with 10 different types and a total of 38 token frequencies.

In sub-corpus 5, Aslı produced 12 types of formulaic sequence structures. *Noun phrase with of-phrase fragment* was still the most frequent in terms of type (8) and token (22). It was followed by 7 types of *verb phrase with active verb* with a total of 14 token frequencies, and *personal pronoun + verb phrase (+ complement-clause fragment)* had 6 types with 12 token frequencies, respectively.

The distribution of structural types of formulaic sequences produced by Aslı across 5 sub-corpora is displayed in Figure 21 below.

Figure 21: Distribution of Structural Types of Formulaic Sequences Produced by Aslı across 5 Sub-corpora



4.3.7.2. Functions of Formulaic Sequences in Aslı's each Sub-corpus

Tables 52 demonstrates the different functional types of the formulaic sequences produced by Aslı in each sub-corpus.

Table 52: Function of Formulaic Sequences in Ash's each Sub-corpus

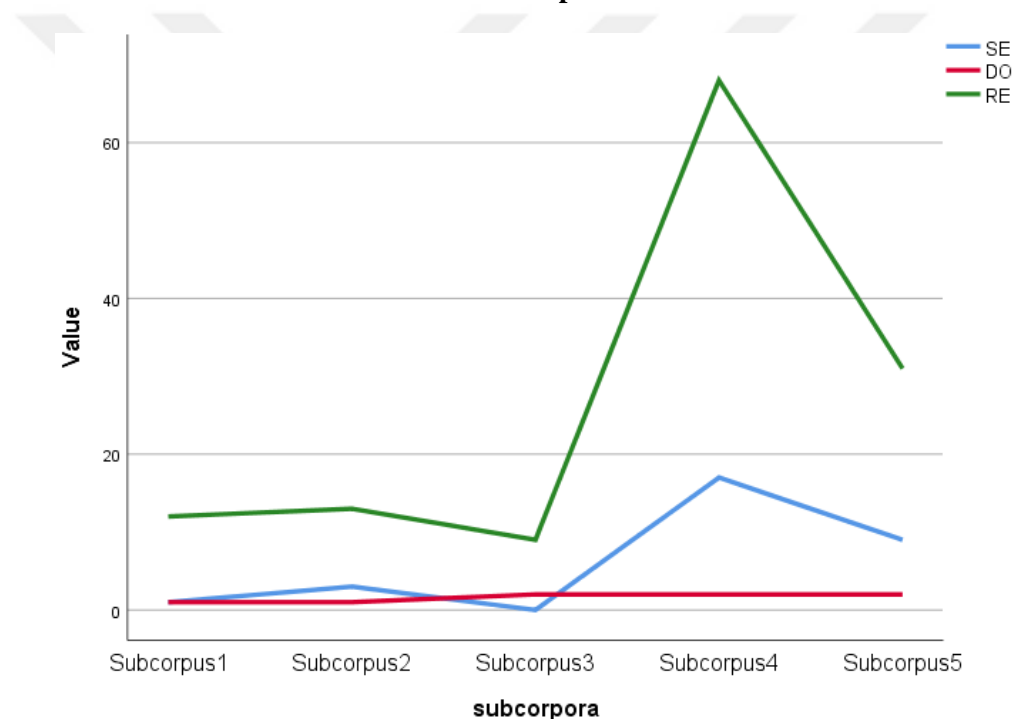
Functional Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSSs	3- and 4- FSSs	3- and 4- FSSs	3- and 4- FSSs	3- and 4- FSSs
1. Stance expressions	<i>we can talk,</i>	<i>think that grandparents, people cannot, cannot afford,</i>		<i>I do not, we are against, and they will not, and I think he, and he can understand, and I think, will be the, agree with them, strongly believe that, because according to, according to his, according to my, according to interview, and they do not, and they cannot, cannot find a, people think that,</i>	<i>I do not agree, and they can, cannot overcome, can overcome all, not want to, will be a, will be more, will emerge in, cannot be,</i>
2. Discourse organizers	<i>as a result of,</i>	<i>on the other hand,</i>	<i>on the contrary, on the other hand,</i>	<i>in addition to, on the contrary,</i>	<i>on the contrary, as a result,</i>
3. Referential expressions	<i>a result of this, one of the most, the reason why, thanks to daily, of the most, is one of the, day by day, and we begin to, we start to use, did not finish her, had a daily, begin to use,</i>	<i>what do you think, who name is, because of this, one of the, level of their, rising of personal, most of people, both of them, reason is about, lots of children, day by day, people have a, stay with their,</i>	<i>they do not, who live in, of nature for, because of this, in ministry of, far away from, as in the, day by day, even if they,</i>	<i>they do not, which is a, most people still, one of the, a kind of, whole of this, because of the, a result of this, and the result of, the reason why, and the most, a research about which, against each other because, against each other he, a big deal for, at the beginning, in the same, and it is obvious, and it was, is the most, are against each other, be the winner, it is obvious that, and for a while, and if it is, and if people do, and they are, when I look, the number of, the rate of, in terms of, part of turkey, day by day, in the future, the end of, be stronger than, people do not, he said because of, he said that, and he said that, he said because, and he is aware, and he adds, and he answered that, and he is, and he warned, and I found, and I look a, and I suggest you, and people became, again we are, and people make a, take care of, start to give, added people to start, and make an agreement, people avoid to, and continue to, and start to,</i>	<i>he does not, he supports these, she is the, I have found that, who are the, who is a, this kind of, one of the, in some part of, claims of the, the cause of, because instead of, of the report, of the world, the most known, in terms of, in the world, in the developing world, around the world, in the future, it does not, is the most, be a solution to, is not a, are completely safe, said that women, report concluded that, the report is that, to the world's, a solution to the, the other biggest claims,</i>

For Ash1, *referential expressions* were the most recurrent in each sub-corpus. In sub-corpus 1, 12 types of *referential expressions* were used whereas there was 1 type of both *stance expressions* and *discourse organizers*. In sub-corpus 2, she produced a similar amount of *referential expressions* with 13 types whereas the types of *stance expressions* (3 types) increased. In sub-corpus 3, there was a steady

decrease of both *referential expressions* and *stance expressions* that *referential expressions* were used with 9 types and there was no type of *stance expressions* while 2 types of *discourse organizers* were found. In contrast to the decrease of types in sub-corpus 3, there was a steady increase in types of both *referential expressions* and *stance expressions* in sub-corpus 4: 68 types of *referential expressions* and 17 types of *stance expressions*. *Discourse organizers* were still the least frequent (2 types). In sub-corpus 5, the number of types of *discourse organizers* was similar to sub-corpus 4 while *referential expressions* were still the most frequent (31 types) and then came *stance expressions* (9 types).

The distribution of functional types of formulaic sequences produced by Aslı across 5 sub-corpora is displayed in Figure 22 below.

Figure 22: Distribution of Functional Types of Formulaic Sequences Produced by Aslı across 5 Sub-corpora



4.3.7.3. Unique Formulaic Sequences in Aslı's each Sub-corpus

In sub-corpus 1, Aslı shared about 57% of formulaic sequence types with learner corpora and about 29% with LOCNESS on a word-for-word or partial basis. *Did not finish her, had a daily, begin to use, we can talk, and we begin to* and *we start to use* were unique formulaic sequence to her.

In sub-corpus 2, she shared approximately 53% of formulaic sequence types with learner corpora and about 12% with LOCNESS. *People have a, level of their, reason is about, rising of personal, stay with their, cannot afford, who name is* and *what do you think* were unique to her.

In sub-corpus 3, about 72% of formulaic sequence types were common with learner corpora, and about 18% of FSs were common with LOCNESS. *In ministry of, far away from* and *as in the* were unique to her.

In sub-corpus 4, she shared almost 69% of formulaic sequence types with learner corpora and about 9% with LOCNESS. The following twenty types of sequences were unique to her: *People avoid to, we are against, whole of this, start to give, be stronger than, most people still, and he is aware, a research about which, and continue to, and for a while, and he answered that, and he warned, and I found, and I suggest you, be the winner, and if people do, and make an agreement, and people became, a kind of, and I look a, again we are, and start to, and people make a and a big deal for.*

In sub-corpus 5, almost 79% of formulaic sequence types with learner corpora while about 17% of FSs were common with LOCNESS. *The other biggest claims, cannot overcome, claims of the, report concluded that, of the report, the report is that, will emerge in, he supports these* and *are completely safe* were unique to her.

For Ashi, the results showed that in each sub-corpus, the majority of the the formulaic sequences were shared on a word-for-word or partial basis with longitudinal learner corpora whereas the proportion of shared sequences were less with native corpus. The analysis of unique sequences indicated that there was a fluctuation in the type and frequency of unique sequences from sub-corpus 1 to sub-corpus 5, and the peak point was in sub-corpus 4.

4.3.8. Zeynep's Inventory of Formulaic Sequences

The top 20 frequently occurring formulaic sequences produced by Zeynep across 5 sub-corpora of the longitudinal learner corpus are presented in Table 53. Sub-corpus 1 consisted of 1,066 tokens and 16 formulaic sequence types. Sub-corpus 2 consisted of 1,249 tokens and 12 formulaic sequence types. Sub-corpus 3 consisted of 1,453 tokens and 19 formulaic sequence types. Sub-corpus 4 consisted of 1,698 tokens, and 30 formulaic sequence types. Sub-corpus 5 consisted of 2,201 and 38 formulaic sequence types.

Table 53: Top 20 Frequent Formulaic Sequences Produced by Zeynep across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
to do their	6	281,43	they cannot	3	120,10	the rules of	14	481,76	the quality of	7	206,12	in the world	5	113,58
she cannot	4	187,62	they can learn	3	120,10	with each other	3	103,23	in addition to	4	117,79	one of the	4	90,87
his or her	3	140,71	as they want	2	80,06	another fact that	2	68,82	people think that	3	88,34	there is a	3	68,15
more than ever	2	93,81	due to the	2	80,06	are also improve	2	68,82	is not the	2	58,89	the genetics of	3	68,15
fed up this	2	93,81	of age and	2	80,06	and it makes	2	68,82	cannot find anything	2	58,89	because of the	3	68,15
he or she	2	93,81	effects on the	2	80,06	of people living in	2	68,82	content of the	2	58,89	answer to the	3	68,15
because of that	2	93,81	and some other	2	80,06	are the rules	2	68,82	anything to support this	2	58,89	the structure of	3	68,15
to sum up	2	93,81	the benefits of	2	80,06	rules to ensure	2	68,82	many people think	2	58,89	about this issue	3	68,15
not want it	2	93,81	things like that	2	80,06	in the street	2	68,82	determine the quality of	2	58,89	to the world's	3	68,15
they push their	2	93,81	to sum up	2	80,06	and with the	2	68,82	agree with this	2	58,89	are equal in	2	45,43
they are bad in	2	93,81	too bad for	2	80,06	show that the	2	68,82	at first appearance	2	58,89	are harmful to	2	45,43
time to do	2	93,81	and it is	2	80,06	some rules to	2	68,82	about this issue	2	58,89	people think that	2	45,43
to begin with	2	93,81				the ability to	2	68,82	a person cannot	2	58,89	are thought to be	2	45,43
too much and	2	93,81				the source of	2	68,82	think that the	2	58,89	as we know	2	45,43
to do her	2	93,81				to sum up	2	68,82	is sufficient or not	2	58,89	at first appearance	2	45,43
according to +smb	2	93,81				researches also show that	2	68,82	with the development of	2	58,89	is looking hard enough	2	45,43
						with the rules	2	68,82	may seem as a	2	58,89	according to +smb	2	45,43
						with the society	2	68,82	be an alive person	2	58,89	a person cannot	2	45,43
						a rule of	2	68,82	find anything to support	2	58,89	of the biggest problems	2	45,43
									because of the	2	58,89	of the products	2	45,43

4.3.8.1. Structures of Formulaic Sequences in Zeynep's each Sub-corpus

Tables 54 presents the different structural types of the formulaic sequences produced by Zeynep in each sub-corpus.

Table 54: Structure of Formulaic Sequences in Zeynep's each Sub-corpus

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
Personal pronoun + verb phrase (+ complement-clause fragment)	<i>she cannot, they push their, they are bad in,</i>	<i>they cannot, they can learn, as they want,</i>		<i>a person cannot, we can provide,</i>	<i>a person cannot, many people believe, person cannot find, as we know, a person can,</i>
Verb phrase with active verb	<i>fed up this, not want it,</i>			<i>cannot find anything, determine the quality of, agree with this, may seem as a, find anything to support, should be an, support this idea, can help us,</i>	<i>should not work, issue may seem as, seem as a senseful, may actually be, cannot be, do not take,</i>
Quantifier expressions				<i>many people think,</i>	<i>too much work,</i>
Noun phrase with of-phrase fragment	<i>because of that,</i>	<i>of age and, the benefits of,</i>	<i>the rules of, of people living in, the source of, a rule of,</i>	<i>the quality of, content of the, with the development of, because of the,</i>	<i>the beauty of, the prestige of, one of the, the genetics of, because of the, the structure of, of the biggest problems, of the products, the yield of the,</i>
Other prepositional phrase fragment		<i>effects on the,</i>	<i>with each other, and with the, in the street, with the rules, with the society,</i>	<i>in addition to, at first appearance, about this issue, for me and, in my opinion,</i>	<i>at first appearance, about this issue, in the world, on their own,</i>
Anticipatory it + verb phrase/adjective phrase		<i>and it is,</i>	<i>and it makes,</i>		
Copula be + noun phrase/adjective phrase			<i>are also improve, are the rules,</i>	<i>is not the, be an alive person, is sufficient or not,</i>	<i>are thought to be, is looking hard enough, are equal in, are harmful to, is one of the,</i>
(verb phrase +) that-clause fragment		<i>things like that,</i>	<i>another fact that, show that the, researches also show that,</i>	<i>people think that, think that the, people believe that,</i>	<i>that they will, people think that,</i>
(verb/adjective +) to-clause fragment	<i>to do their, time to do, to begin with, to do her,</i>		<i>rules to ensure, some rules to, the ability to,</i>	<i>anything to support this,</i>	<i>answer to the, to the world's, solution to this,</i>

Table 54: (Continue)

Structural Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
Pronoun/noun phrase + be (+ . . .)				<i>there are many,</i>	<i>there is a,</i>
Other expressions	<i>his or her, he or she, to sum up, too much and, according to +smb</i>	<i>due to the, and some other, to sum up, too bad for,</i>	<i>to sum up,</i>	<i>according to their,</i>	<i>according to+smb,</i>
If-clause fragments				<i>but if someone is,</i>	<i>but if someone is,</i>
Comparative expressions	<i>more than ever,</i>				

In sub-corpus 1, Zeynep produced 6 types of formulaic sequence structures. The most utilized structural type was *other expressions*. She produced 5 different types of these structures with a total of 11 token frequencies.

In sub-corpus 2, Zeynep produced 6 types of formulaic sequence structures. *Other expressions* structural type was still the most frequent in terms of type (4) and token (8). It was followed by 3 types of *personal pronoun + verb phrase (+ complement-clause fragment)* with a total of 8 token frequencies.

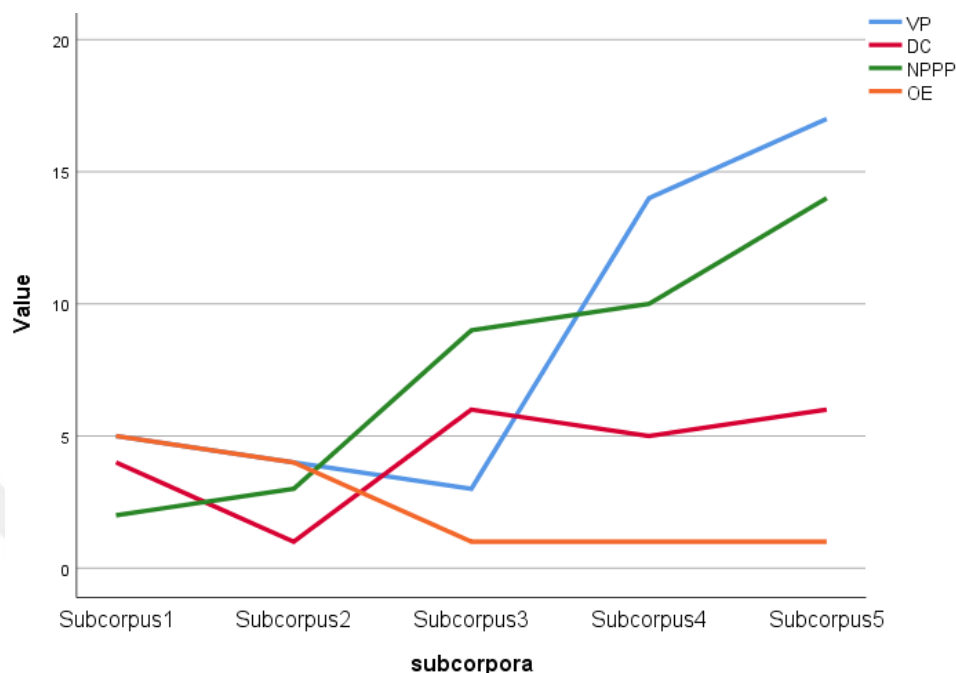
In sub-corpus 3, Zeynep produced 7 types of formulaic sequence structures. Different from the other 2 sub-corpora, the most utilized structural type was *other prepositional phrase fragment*. She produced 5 different types of this structure with a total of 11 token frequencies, followed by 4 types of *noun phrase with of-phrase fragment* with a total of 20 token frequencies.

In sub-corpus 4, Zeynep produced 11 types of formulaic sequence structures. The most popular structural type was *verb phrase with active verb* with 8 different types and a total of 16 token frequencies. *Other prepositional phrase fragment* was the following one with 5 different types and a total of 10 token frequencies.

In sub-corpus 5, Zeynep produced 11 types of formulaic sequence structures. The most popular structural type was *noun phrase with of-phrase fragment* with 9 different types and a total of 23 token frequencies. It was followed by 6 types of *verb phrase with active verb* with a total of 12 token frequencies.

The distribution of structural types of formulaic sequences produced by Zeynep across 5 sub-corpora is illustrated in Figure 23 below.

Figure 23: Distribution of Structural Types of Formulaic Sequences Produced by Zeynep across 5 Sub-corpora



4.3.8.2. Functions of Formulaic Sequences in Zeynep's each Sub-corpus

Tables 55 shows the different functional types of the formulaic sequence in each sub-corpus.

Table 55: Function of Formulaic Sequences in Zeynep's each Sub-corpus

Functional Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
1. Stance expressions	<i>according to +smb, not want it, she cannot,</i>	<i>as they want, they cannot, they can learn,</i>	<i>another fact that,</i>	<i>agree with this, may seem as a, should be an, can help us, many people think, people think that, think that the, people believe that, according to their, a person cannot, we can provide, cannot find anything, in my opinion,</i>	<i>a person cannot, person cannot find, a person can, should not work, issue may seem as, may actually be, cannot be, according to+smb,</i>
2. Discourse organizers	<i>to sum up,</i>	<i>due to the, to sum up,</i>	<i>to sum up,</i>	<i>in addition to,</i>	

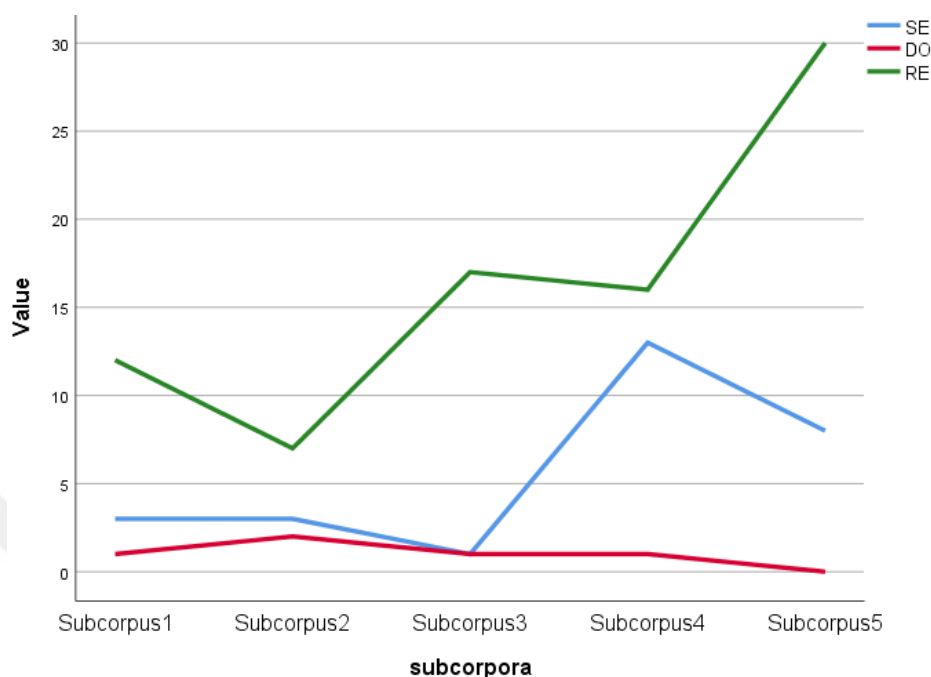
Table 55: (Continue)

Functional Types	Sub-corpus 1	Sub-corpus 2	Sub-corpus 3	Sub-corpus 4	Sub-corpus 5
	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs	3- and 4- FSs
3.Referential expressions	<i>because of that, to do their, time to do, to begin with, to do her, his or her, he or she, more than ever, too much and, they push their, they are bad in, fed up this,</i>	<i>of age and, effects on the, and it is, things like that, too bad for, and some other, the benefits of,</i>	<i>with each other, and with the, and it makes, are also improve, are the rules, rules to ensure, some rules to, the ability to, the rules of, the source of, a rule of, with the rules, with the society, of people living in, in the street, show that the, researches also show that,</i>	<i>because of the, content of the, at first appearance, about this issue, for me and, is not the, be an alive person, is sufficient or not, but if someone is, there are many, the quality of, with the development of, determine the quality of, anything to support this, find anything to support, support this idea,</i>	<i>many people believe, as we know, seem as a senseful, do not take, too much work, the beauty of, the prestige of, one of the, the genetics of, because of the, the structure of, of the biggest problems, of the products, the yield of the, at first appearance, about this issue, in the world, on their own, are thought to be, is looking hard enough, are equal in, are harmful to, is one of the, that they will, people think that, answer to the, to the world's, solution to this, there is a, but if someone is,</i>

Referential expressions were the most recurrent functional type in each sub-corpus. In sub-corpus 1, Zeynep produced 12 types of *referential expressions* and 3 types of *stance expressions* and then came 1 type of *discourse organizers*. In sub-corpus 2, the amount of *referential expressions* decreased to 7 types while *stance expressions* remained the same with 3 types and there was a similar amount of *discourse organizers* (2 types). In sub-corpus 3, *referential expressions* were still the most utilized with 17 types and the types of *stance expressions* and *discourse organizers* were similar to sub-corpus 2 with 1 type. In sub-corpus 4, there was a steady increase in the types of *stance expressions* (16 types) whereas *referential expressions* (16 types) and *discourse organizers* (1 type) were similar to sub-corpus 3. In sub-corpus 5, she produced a higher amount of *referential expressions* (30 types) while there were 8 types of *stance expressions*, there were no *discourse organizers*.

The distribution of functional types of formulaic sequences produced by Zeynep across 5 sub-corpora is illustrated in Figure 24 below.

Figure 24: Distribution of Functional Types of Formulaic Sequences Produced by Zeynep across 5 Sub-corpora



4.3.8.3. Unique Formulaic Sequences in Zeynep's each Sub-corpus

In sub-corpus 1, Zeynep shared about 56% of formulaic sequence types with learner corpora whereas there was no shared pattern with LOCNESS on a word-for-word or partial basis. *Fed up this, they push their, she cannot, more than ever, too much and, they are bad in and time to do* were unique formulaic sequence to her.

In sub-corpus 2, 50% of formulaic sequence types were common with learner corpora, and about 8% of FSs were common with LOCNESS. *They can learn, of age and, effects on the, and some other, things like that and too bad for* were unique to her.

In sub-corpus 3, she shared about 37% of formulaic sequence types with learner corpora whereas there was no shared pattern with LOCNESS. *Are also improve, of people living in, rules to ensure, in the street, and with the, some rules to, the source of, are the rules, the ability to, a rule of, with the rules and with the society* were unique to her.

In sub-corpus 4, 80% of formulaic sequence types were common with learner corpora while 10% of FSs were common with LOCNESS. The following six types of sequences were unique to her: *content of the, is sufficient or not, be an alive person, for me and, can help us and we can provide.*

In sub-corpus 5, she shared about 68% of formulaic sequence types with learner corpora and about 16% shared patterns with LOCNESS. *The genetics of, the structure of, are harmful to, as we know, on their own, are thought to be, the prestige of, may actually be, the beauty of, the yield of the, do not take and too much work* were unique to her.

For Zeynep, the results revealed that in each sub-corpus except for sub-corpus 3, the majority of the formulaic sequences were shared on a word-for-word or partial basis with longitudinal learner corpora. There were no shared formulaic sequences in sub-corpus 1 and 3 while the proportion of shared sequences in sub-corpus 2, 4 and 5 were relatively small percentages with native corpus. The analysis of unique sequences indicated that there was a fluctuation in the type and frequency of unique sequences from sub-corpus 1 to sub-corpus 5. The amount of type and frequency of unique sequences were similar in sub-corpus 1,2 and 4 while the number of them was high in sub-corpus 3 and 5.

4.4. Data Analysis of the Retrospective Protocols

Qualitative data were collected through the use of retrospective protocols from selected sample students. Obtained qualitative data of retrospective protocols were analysed through encoded categories. The formulaic sequence samples were highlighted in the essays of samples and these students were asked protocol questions related to their uses, preferences, awareness and decisions towards the use of these sequences. These categories and codes are submitted as below in Table 56.

Table 56: Protocol Questions and their Coded Categories

	Protocol Questions	Categories/Codes
1	What are the things that you pay the most attention while writing?	Primary considerations in writing
2	What difficulties do you have most while writing?	Difficulties in writing
3	How important do you think “word choice” is while you are writing and how do you choose them?	Importance and selection of word choice
4	Have you ever heard of “formulaic sequences”? Or collocations?	Familiarity with FSs
5	Do you pay attention to use formulaic sequences while writing?	Attention to FSs in writing
6	What do you think of your essays in terms of formulaic sequences usage patterns?	FSs patterns in writing
7	How did you think of using the formulaic sequences samples underlined in your essays?	Why used FSs?
8	In these essays, were there any FSs you used often since you knew well these sequences?	Frequently used FSs
9	Have you noticed any development with regards to the use of formulaic sequences in your essays across two semesters? When? How?	Development of FSs usage
10	In English lessons in previous years, did you ever encounter with the formulaic sequences?	Previous exposure to FSs
11	How much do you think these encounters helped you increase your awareness towards them?	Awareness

The above protocol questions were asked the participants and their responses to these questions were analysed descriptively after then the relevant themes were classified. Classified themes are displayed in Table 57 below.

Table 57: The Encoded Analysis of the Protocol Questions

Protocol Questions	Codes	Participants
Primary considerations in writing?	<ol style="list-style-type: none"> 1. Organization 2. Define the Topic of Essay or Idea 3. Coherence-Cohesion 4. Grammar 5. Teacher Feedback 6. Lexical Patterns 7. Content of The Topic 8. Sentence Patterns 9. Dictionary Check for FSs 10. Academic Word/Phrase List 11. Simple and Clear Sentences 12. Tense Agreement 13. Essay Topic 14. Grammar Check (Grammarly) 15. Usage of Different Range of Lexical Patterns 16. Advanced Grammar 17. More Complex and Compound Sentences 18. Creative ideas 19. Using Accurate Grammar 	P1, P2, P3, P4, P5, P1, P1, P1, P2, P1, P3, P4, P1, P2, P3, P5, P3, P4, P4, P4, P4, P5, P5, P5, P6, P6, P6, P6,
Difficulties in writing?	<ol style="list-style-type: none"> 1. Difficulty in Identifying an Idea 2. Choosing the Right Word 3. Difficulty in Creating Unity 4. Difficulty to Creating Body 5. Difficulty in Using Conjunctions 6. Difficulty in Using Words 7. Difficulty in Paraphrasing 8. Difficulty in Recognizing Prefabricated Patterns 9. Difficulty in Grammar 10. Difficulty in Tense Agreement 11. Difficulty in Sentence Constructions 12. Difficulty in Writing Unfamiliar Topic 13. Difficulty in Using Unfamiliar Words 14. Difficulty in Choosing the Appropriate Words 15. Difficulty in Using Academic Collocation 16. Difficulty in Creating New Idea 	P1, P1, P1, P2 P2 P2 P3 P3 P3 P3 P4 P4 P4 P4 P4 P5 P5 P6
Importance and selection of word choice?	<ol style="list-style-type: none"> 1. Important 2. Very Important 3. Choosing the Right Terminology 4. Searching FSs in Google N-Gram 5. Using a Dictionary (checking for words) 6. Google Search 7. Checking from Sample Essays 8. Checking Dictionary for Synonym 9. Checking Articles for Synonym 	P1, P4, P2, P3, P5, P6, P1, P2, P2, P2, P2, P3, P3,

Table 57: (Continue)

Protocol Questions	Codes	Participants
Importance and selection of word choice?	<ol style="list-style-type: none"> 10. Checking Dictionary for FSs 11. Choosing from Academic Word List 12. Checking Context of the Word 13. Choosing Formal Words 14. Using Coca 15. Reading and Selecting Word 16. Paraphrasing 17. Synonym Check 18. Collocation Check 19. Teacher Feedback 20. Choosing Less Familiar Words 	<p>P4, P4, P4, P5, P5, P5, P5, P5, P5, P5, P6,</p>
Familiarity with FSs?	<ol style="list-style-type: none"> 1. Familiarity with FSs and Collocation 2. No Familiarity with the FSs 3. Familiarity with Collocation 	<p>P1, P2, P5, P3, P4, P6, P3, P4, P6,</p>
Attention to FSs in writing?	<ol style="list-style-type: none"> 1. Pay Attention to Use of FS 2. Usage of Well-Known FSs 3. Pay attention to Accurate Usage of FSs 	<p>P1, P2, P3, P4, P5, P6, P2, P6,</p>
FSs patterns in writing?	<ol style="list-style-type: none"> 1. Lack of Usage 2. Frequent Usage of Well-Known FSs 3. Necessity to include More Fs into Essays 4. Lack of Awareness 5. Frequent Use of Fs 6. Lack of Variety/Limited Range of it 7. Adequate Usage of FSs 	<p>P1, P2, P3, P5, P1, P1, P2, P3, P4, P5, P6, P3, P4, P4, P6,</p>
Why used FSs?	<ol style="list-style-type: none"> 1. Background Knowledge 2. Context Help 3. Need of Appropriate Use of FS 4. Teacher Feedback 5. Self-Awareness 6. More Academic 7. Individual Search for FSs 8. COCA 	<p>P1, P3, P4, P6, P1, P2, P3, P5, P3, P3, P5, P3, P5, P5,</p>
Frequently used FSs?	<ol style="list-style-type: none"> 1. More in Passive Constructions 2. More in If Clause Sentences 3. More in Longer Word Patterns 4. Usage of Frequently Encountered FSs 5. Use Through Teacher Feedback 6. Usage of Limited Patterns 7. Use Popular FS Pattern 8. Use from More Academic Fs 9. Usage of Well-Known FSs 10. Academic Word/Phrase Lists 11. More in Passive Construction 	<p>P1, P1, P1, P2, P3, P5, P3, P3, P4, P6, P3, P4, P4, P5,</p>

Table 57: (Continue)

Protocol Questions	Codes	Participants
Development of FSs usage?	<ol style="list-style-type: none"> 1. Raise Awareness on FS 2. Not Fully Conscious of FS 3. More FS Through Teacher Feedback 4. Using More Complex FS 5. Using More Academic FS 6. Increase in the Variety/Range of FS 7. Frequent and Different Range of FSs 8. Writing Quickly with FS 9. Less Dictionary Check 10. Longer Sentences 11. Better Grammar 12. Across the Course Schedule 	<p>P1, P1, P2, P3, P4, P3, P3, P6, P4, P6, P5, P5, P5, P6, P6, P6,</p>
Previous exposure to FSs	<ol style="list-style-type: none"> 1. Not Really 2. Very Limited 3. Very Limited in High School 4. More Frequent After Teacher Feedback 5. More Frequent in Prep Class 6. Very Limited Through Language Course 7. Through Reading Articles 8. Through English Courses in High School 9. Individual Attention on FS 10. Teacher' Emphasis on FSs 	<p>P1, P3, P1, P2, P3, P4, P3, P4, P5, P6, P6, P6, P6,</p>
Awareness	<ol style="list-style-type: none"> 1. More Awareness to FS 2. More Conscious Usage 3. Increase Through Reading 4. More awareness Through Teacher Feedback 5. More Conscious Attitude 6. Increase Through Writing Courses 7. Increase Through Prep Class Writing Courses 8. Through Conversation Outside the Class 	<p>P1, P2, P3, P4, P5, P6, P1, P3, P3, P4, P3, P4, P5, P6</p>

In this protocol, for general purpose questions as given in the first three questions (primary considerations, difficulties and importance and selection), most of the participants mainly focused their attention on organization. While teacher feedback was emphasized by three participants, it was followed by grammar and sentence patterns with two participants. The other themes that were asserted by only one participant were the followings: define the topic of essay or idea, coherence-cohesion, lexical patterns, content of the topic, dictionary check for FSs, academic word/phrase list, simple and clear sentences, tense agreement, essay topic, grammar check (Grammarly), usage of different range of lexical patterns, advanced grammar, more complex and compound sentences, creative ideas, using accurate grammar.

The second question asked the participants about the difficulties they experienced while writing. Although there were no shared themes by participants, several of the difficulties which

participants experienced could be gathered under broader themes. It is important to note that in here, one of the broader themes might be word-level difficulties such as choosing the right word (P1), using words (P2), using unfamiliar words (P4) and choosing the appropriate words (P5). Similarly, the matter of word patterns might be the second one, for instance, they had difficulty in recognizing prefabricated patterns (P3) and using academic collocation (P5). Participant 1 stated that she had difficulty in identifying an idea, choosing the right word and creating unity. Participant 2 had difficulty in creating body, using conjunctions and using words. The fifth participant's difficulty in writing was related to issues of using words such as choosing the appropriate words and using academic collocation. She expressed her thoughts on difficulties while writing as the following;

“Choosing the right word was challenging, maybe that was why I used too many relative and that clauses for the explanation. Academic words and word combinations, that is, I noticed thanks to teacher feedback that I could not use the appropriate word combinations” (Participant 5).

The participants were asked their opinion on the importance and selection of word choice. While four of them stated that word selection was very important for them, two of the participants reported as an important. Some of the prominent ways of choosing words were using a dictionary on the purpose of checking for words (P2), synonyms (P3) and formulaic sequences (P4). Besides of dictionary check, searching formulaic sequences in Google N-grams (P2), using Google Search (P2), checking from sample essays (P2), checking articles for synonym (P3), choosing from academic word list (P4), using COCA (P5) were among them.

The fourth question asked whether the participants were familiar with formulaic sequences and collocations. Whereas participants 1, 2 and 5 reported that they were familiar with formulaic sequences and collocation, participants 3, 4 and 6 had no familiarity with FSs. There were no participants that did not familiar with collocation and most of them stated that they knew collocations because they encountered with these collocations from the courses they have taken and from their teachers that used them in the classes. The following quotation from the protocol of one of the participants illustrates this notion:

“We covered them step by step in our course, the course instructor explained them to us using slide shares. We used these structures as well” (Participant 5).

The fifth question tried to find an answer to whether they paid attention to FSs in writing or not. All of the participants reported that they paid attention to the use of formulaic sequences in writing. Participants 6 also stated that she paid attention to accurate usage of formulaic sequences since she thought that using these combinations made her use of language (English) even better. On the other hand, participant 2 remarked the following statement;

“In fact, I could construct sentences which had background, so I was not looking for different word combinations” (Participant 2).

For question 6 (FSs patterns in writing), all of the participants responded that there was a necessity to include more FSs into their essays. Another frequent response was lack of usage. Participant 1 mentioned about lack of usage with the following sentence;

“I mean, considering what was highlighted in my essays, it seems to me that I did not use these combinations enough” (Participant 1).

She added that she used frequently some FSs since she knew them well. Participant 4 talked about an important issue that though she used frequently formulaic sequences in her essays, there were mostly the same patterns in each paragraph of each text, so there was no pattern variety.

The essays were shown to formulaic sequence samples from their argumentative essays and asked to account for their choice. Participants 1, 3, 4 and 6 responded that they used these sequences thanks to their background knowledge while participants 3 and 5 used these sequences through teacher feedback and individual search for FSs. Participant 3 remarked that she had probably heard most of 3- to 4- word sequences from the course instructor, so these longer combinations were seemed to be more academic to her. In a similar vein, participants 5 indicated the same theme for this question. The following codes also were emphasized: context help, need of appropriate use of FSs and self-awareness.

For question 8, whether there was any FS they used often since they knew well was asked. Participants 3, 4 and 6 replied that they used frequently popular FSs patterns in their essays. Participant 2 remarked that the following combinations were the first sequences that spring to her mind: *as a result of, in a way, such as a, according to the and in terms of*.

“In fact, I can say that these are popular and well-known that they were learned and embedded subconsciously. These sequences are repeated over and over again” (Participant 2).

The codes of usage of well-known FSs and sequences from academic word/phrase lists were marked by participant 4. The statements of participant 4 correspond with the claim of Granger (1998) that the safe or well-known FSs were used by L2 learners. While participant 3 stated that she used limited patterns, participant 2 used frequently encountered FSs.

Question 9 asked the participants to find out whether they noticed any development with regards to the use of formulaic sequences in your essays across the term. Participants 2, 3 and 4 declared that they used more FSs through teacher feedback. Participant 1 responded that thanks to

teacher feedback, she improved the use of FSs across each essay. One of the codes asserted by participants 4 and 6 was the increase of the range and variety of FSs. Participant 4 stated that;

“As I encountered or used FSs in the courses, maybe towards the end of the course, I gained courage and took a risk, so I used different range of FSs” (Participant 4).

Participants 3 and 6 pointed out that they used more academic FSs. Participant 5 reported the following codes: frequent and different range of FSs, writing quickly with FSs and less dictionary check.

The 10th question asked participants whether they have ever encountered with formulaic sequences. Three of the participants (P2, P3 and P4) reported that they had very limited exposure to FSs in their high school years. Participant 6 stated that she encountered formulaic sequences through reading articles, through English courses in high school, teacher’ emphasis and individual attention on FS. Participant 4 emphasized the increase of exposure in prep class that;

“Due to the very limited exposure in high school, we encountered more frequently with some academic patterns and particular sequences in prep class because we attended writing courses and encountered sequences in that lesson. I noticed such word patterns mostly in the prep class” (Participant 4).

Participant 3 responded that she provided somehow awareness in prep class and the exposure and use of FSs increased after teacher feedback, so she used more frequently.

For question 11, although they had very limited previous exposure to FSs, they reported that their awareness towards formulaic sequences increased. Participant 1 said that this very limited exposure provided more conscious usage for her. Two of the participants (P3 and P4) claimed that they had more awareness through teacher feedback. Whereas the awareness of participant 3 raised through reading and this brought about more conscious attitude, writing courses created awareness of participants 4 and 5. Except for these, participant 6 focused on different codes, adding that conversation outside the classroom provided awareness to her.

4.5. Discussion

The current thesis is a true longitudinal design, and it was aimed to explore the use and development of three- and four-word formulaic sequences in longitudinal learner corpora composed of tertiary level EFL learners’ essays across two semesters. Using a frequency-driven approach, the most frequent formulaic sequences were extracted in a period of two consecutive semesters in 2018-2019. A total of ten argumentative essays for each participant were analysed in terms of formulaic

sequence content, frequency and type. The analysis included two groups of EFL learners with different levels of language proficiency. The researcher used a frequency-based approach in the analysis of the data and the raw frequencies obtained from the five sub-corpora of each group were normalized and the resulting data was subjected to Pearson correlation analysis. The analysis also included structural and functional categorisation of the FSs in the form of tables and graphics for each group. In doing so, the researcher aimed to investigate collective trends in the use of FSs. Finally, the researcher employed an individual analysis of the eight participants who were selected through purposive sampling, for four participants from each group. The retrospective protocol was made with the 6 participants in an attempt to better understand how they learn FSs and how FSs change over time. In the processes of analysing retrospective protocols, encoded categories for the participants' retrospective accounts were analysed. As mentioned in the methodology part, this thesis study aimed to investigate the FSs usage and development of EFL learners in a longitudinal design which is as a window that opens to do non-native writers' frequent lexical patterns in writing compared to those of the native writers. The sub-corpora that were used in the study included academic and argumentative essays ranging between 248 to 1198 words for two consecutive semesters. The written productions of EFL learners in two groups were compiled according to a strict design criterion and contained un-timed expository and academic essays in which FSs were used frequently.

All the analysis and the results yielded interesting results regarding the FSs developmental levels of EFL learners, and they are discussed in this chapter.

Regarding the first research question, the frequency analysis of three- to four-word formulaic sequences concluded that the number and the range of FSs seemed to reveal an increasing pattern in number and type. Conklin and Schmitt (2008) stated that “formulaic sequences are frequent and relatively salient (because they are typically linked to specific meanings or functions)...”(2008: 79). The rationale behind the frequent usage of FSs such as *in order to*, *one of the*, *a lot of*, *they do not*, *it is not*, *day by day*, *there is a*, *there is no* and *on the other hand* might be due to the fact that those units are highly frequent and, thus, salient (i.e., noticeable and prominent) in the written input of L2 learners (Conrad and Biber, 2005). In this sense, the findings of this thesis were consistent with the study of Biber et al. (1999), who stated that *in order to*, *one of the*, *part of the*, *the number of*, *the presence of*, *the use of*, *the fact that*, *there is a* and *there is no* were the most common three-word sequences in academic prose while *in the case of* and *on the other hand* were the most common four-word lexical units in academic prose.

The frequency analysis which was done in the study is due to the fact that high frequency FSs can be learned and processed more easily than less frequent FSs according to O'Donnell et al. (2013). Such features as frequency, familiarity, conventionality, prototypicality/stereotypicality (Giora, 2003) are the significant factors for EFL learners to learn. According to Giora (2003), “the more

frequent, familiar, conventional, or prototypical/stereotypical the information in the mind of the individual or in a certain linguistic community, the more salient it is in that mind or among the community members” (2003: 15-16). This is also supported by O’Donnell et al. (2013) who stated that “humans learn more easily and process more fluently high frequency forms and “regular” which are exemplified by many types and which have few competitors” (2013: 89). In the literature of phraseology, the significance of frequencies of exposure to formulaic sequences was supported by various researchers such as Ellis, (2002), Wood, (2002), Wray, (2000), Ellis et al. (2008), Webb et al. (2013), Tekmen and Daloglu, (2006) and Üstünbaş (2014). The findings of this thesis clearly revealed that FSs were used saliently in the academic and expository essays in the two groups. Many FSs especially in Group 1 sub-corpora were used in frequent FSs such as *one of the, a lot of, there is a and day by day*. Formulaic patterns which are similar to the academic ones such as *for this reason, as long as, it is an undeniable and there will be* were not frequent in two groups since they demand extensive reinforcement not easily retrieved by the learners (Elturki, 2015). Academic formulaic sequences are difficult for L2 learners to attain since “they are simply less prominent in the speech stream” (Boyd and Goldberg, 2009: 419). This finding corroborates the ideas of Stubbs (2007), who suggested that recurrent lexical sequences in a corpus are “good evidence of what is typical and routine in language use” (2007: 130).

The examination of the most frequent FSs across two semesters gave the researcher various FSs differing in length and type. Three-word FSs were used more commonly in the argumentative essays of L2 learners. The main 4-word formulaic sequences on the other hand that were frequently used in the longitudinal learner corpora were *on the other hand, is one of the, I strongly believe that, I firmly believe that* and *one of the most*. These findings are concurrent with the finding of Conrad and Biber (2005) who found that 3-word FSs are more frequent in the corpora as evidence in NES written corpora of academic prose. They remarked that three-word formulaic sequences occur over 60,000 times per million words while four-word sequences over 5,000 times per million words in academic prose. This finding is also similar to the findings of Juknevičienė’s (2009), Pavesi’s (2013), Huang’s (2014) and Ulfa and Muthalib (2020). In a study of continuous recurrent sequences in speech despite being outside the scope of the study, Altenberg (1998) reported that “the word-combinations are on the whole not very long: they range from 3 to 5 words, with a mean length of 3.15 words” (1998: 103) in the London-Lund Corpus.

The frequent use of 3- and 4-word FSs may be given to the fact that they can be found in naturally occurring spoken and written language, and thus their exposure must have been easier for EFL learners. This claim is supported by the findings of Conrad and Biber (2005) who investigated the use of FSs in conversation and academic prose. They found that there were almost 4,000 different three- and four-word formulaic sequences in conversation, and about 3,000 different three- and four-word sequences in academic prose. As can be seen from the findings, the three- and four-word FSs occur frequently in both registers. They also stated that the most frequent sequences in academic

prose such as *in order to*, *one of the* and *part of the* occur between 200 and 400 times per million words.

Regarding the types of the FSs overtime showed frequent but limited FSs usage patterns. In other words, FSs were almost similar types of FSs or repeated across two semesters of both groups. This may be given to the learners' limited stock of L2 FSs (Granger, 1998; Wang, 2016; Elturki, 2015; Kuosmanen, 2020) and which is also supported by Ellis (2012) who stated that learners tend to use were common FSs and familiar constructions. According to Granger (1998), learners feel confident when they use safe or familiar sequences because these sequences help them to compensate for their limited repertoires of formulaic sequences.

Our hypothesis put forward at the beginning of the study was “if EFL learners are more frequently exposed to different sets of formulaic sequence types from their previous language instruction onward, this will be a positive contributing factor to their language development”. The results discussed so far in the study partly confirm this hypothesis in that EFL learners seem to show a reliance on frequent FSs. With this in mind, however, the researcher noticed that the type of FSs did not seem to increase to a great extent. Therefore, our findings suggest that more instruction and increasing teacher feedback may play a role in the number of FSs rather than their types.

Regarding the second research question, the structural and functional analysis concluded that verb phrase fragments were the dominant structural type, followed by noun phrase and prepositional phrase fragments. The category of dependent clause fragments and other expressions were the other two structural types with the lower proportions in longitudinal learner corpora. The structure of a great number of the frequent FSs in the category of VP-fragments across two semesters of observation contained the personal pronoun such as *we cannot*, *I cannot*, *we do not*, *they do not* and *I believe that* (Biber et al., 1999) while the majority of recurrent FSs in the category of NPPP-fragments occurred in other prepositional phrase fragments. These findings are in agreement with Fattani's (2018) findings that VP-based structures are the most frequently occurring in the textbooks and the written AFL sub-list, and these structures account for the highest proportion of formulaic sequences. In another study conducted by Alamri (2017) on the use of FSs in introduction, methods, results, discussion and conclusion sections of English language research articles published in Saudi Arabian and international journals in the field of applied linguistics, the usage proportions of noun phrase fragments, prepositional phrase and verb phrase fragments were similar. The findings of the current study are not consistent with the results of the study conducted by Cooper (2016) who investigated four-word FSs in the IELTS writing tests, student essays and published writing within the field of psychology. Cooper (2016) found that in these corpora, the dominant structural types were prepositional phrase and noun phrase fragments, followed by verb phrase fragments.

The FSs in terms of functional category indicated that writers used high amount of referential expressions followed by stance expressions and discourse organizers. In this longitudinal corpus study, it was observed similar tendencies in FSs functions by EFL learners. The referential functions such as *one of the, it is not, it is a, is one of the, there is a* and *there is no* were among the most frequent ones across two semesters. One of the possible reasons for this may be that the learners were frequently exposed to these FSs functions in their previous instructions. This claim may be supported by usage-based theories of language learning since the frequency is crucial for acquisition (Ellis, 2002).

FSs such as *one of the, a lot of* and *in terms of* are some of the most frequent examples from the referential category and which are used by learners while writing. Stance expressions, on the other hand, become slightly less frequent in all sub-corpora. Such stance expressions such as *the fact that, I do not want, do not agree with* and *should not be* were the most frequent ones. FSs of discourse organizers were the least employed ones by the EFL learners and their frequencies and types remain relatively lower during two semesters. The overall findings are that EFL learners predominantly used referential categories of FSs, making their essays more and more impersonal. From the early stages of L2 development, EFL learners rely on referential FSs and this finding is consistent with the findings of Vidaković and Barker (2009) who stated that “learning conventionalised word strings starts emerging after the lowest proficiency level” (2009: 144). With this in mind however, the FSs employed so far were the most common and invariant ones.

The findings of this research have revealed that the *referential expressions* are the most employed ones, corresponding to Vidaković and Barker’s (2009) findings which showed that *referential expressions* were dominant in the written learner corpus at all levels. These results matched to those observed in a great number of earlier studies. For instance, the study of Tomankova (2016) showed that referential expressions present the most frequently occurring functional type. A study conducted by Breeze (2013), which investigates FSs employed in four legal corpora: academic law, case law, legislation, and documents indicated the same results. Biber and Barbieri (2007) studied on different registers and found strikingly different results in terms of functional types that institutional writing comprised of approximately 70% of *referential expressions* whereas written course management involved over 70% of stance expressions. They also claimed that referential expressions were dominant in academic writing (e.g., academic prose and textbooks). This finding was also in agreement with Kashiha and Heng (2014) findings which showed *referential expressions* were the most common functional type in the two disciplines, namely politics and chemistry lectures, and the investigators suggested that a great number of formulaic sequences in academic lectures were used to describe entities, characteristics, place, time and topic references. Lastly, the study of Fattani (2018) may be a good example of functional usage of formulaic sequences in different registers that in the instructors’ materials and the written AFL sub-list, the proportion of *referential expressions* were high compared to *stance expressions* and *discourse organizers*. In contrast to these two

registers, the distribution of functional categories seemed to have different findings and *stance expressions* were the most common in the textbooks. All in all, this study produced results which corroborate the findings of a great deal of the previous work in the field (e.g., Appel, 2011).

The stance expressions were the second, and the discourse organizers were the least employed across two semesters, which corresponds with the findings of Biber and Barbieri (2007) indicating the usage of over 10% of *stance expressions* and *discourse organizers*. The findings of the current study also were consistent with those of Kashiha and Heng (2014) that *stance expressions* were the second common functional type including 26% in politics and 28% in the chemistry of FSs.

Regarding the third research question, the Pearson correlation test showed that the frequent three- to four-word FSs in the two learner corpora seemed to moderately correlate, in Group 1, with the native learner corpora in the four sub-corpora, however, it was not the case in Group 2. This group revealed unique patterns when compared to Group 1 and native corpus (LOCNESS). One possible reason for this may be that Group 2 learners' language proficiencies were comparatively higher and these learners must have developed and preferred to use their own specific formulaic sequences in their essays. The study of Adel and Erman (2012) seems to confirm with this finding in that the shared formulaic sequences in their corpora with highly advanced learners revealed more unique patterns (60) with non-native learners when compared to those (55) with native learners.

The FSs similarities and the differences in Group 1 and Group 2 sub-corpora and native corpus seemed to be consistent with past learner corpus research for various reasons. The results of Group 1 and Group 2 confirm a general pattern, which non-native speakers had more restricted repertoire of formulaic sequences than native speakers, found by several investigators such as Ma, 2009; Adel and Erman, 2012; Karabacak and Qin, 2013; Salazar, 2014; Kashiha and Chan, 2015; Taşkaya, 2019).

Regarding the fourth research question, the study also aimed to investigate individual participants' inventories of formulaic sequences look like in terms of unique patterns used by EFL learners. The results revealed partly similar observations to the ones discussed above.

1. FSs increased in frequency and type across two semesters.
2. Three-word FSs were more frequent than four-word FSs.
3. Common structural main categories used by the five participants were: verb phrase fragments and noun phrase and prepositional phrase fragments.
4. Referential expressions were the most frequent which is followed by stance expressions and discourse organizers.

The analysis of FSs usages resulted in novel findings that may contribute to the body of literature on learner corpus research. The analysis of individual FSs usages resulted in the emergence of several unique FSs. In other words, certain types of FSs were repeatedly used by the EFL learners, but the number and type of FSs were almost absent in native corpus (LOCNESS). Some of the examples of unique FSs are given: *in higher positions, the formation of, may be beneficial to, by changing the, in some parts of, is a serious, are at the, they change their, transferred to the and an important role to*. These unique FSs were used for various reasons. This finding corroborates the ideas of Elturki (2015), who suggested that learners used unique sequences frequently to link and contrast their past experiences, personalities, and lives to theirs in the L2 world. That is to say, these sequences were used to refer to learners' uncovered aspects concerning their identity, interests, attitudes, and concerns (Elturki, 2015).

Regarding the fifth research question, the overall usage patterns of FSs indicated that the participants made conscious decisions to use several formulaic sequences in the context of writing. When the participants were asked about the things that they paid the most attention while writing, out of 6 protocol participants, 5 participants noted about organization while teacher feedback was the second commonly uttered ones, followed by grammar and sentence patterns. The second protocol question asked about the difficulties while writing. Each of the participant reported various difficulties while writing, such as writing unfamiliar topic, using unfamiliar words, choosing the appropriate words, using academic collocation and creating unity, creating body, using conjunctions, using words, paraphrasing and recognizing prefabricated patterns. For the third protocol question about the importance and selection of word choice, all of the participants reported that word selection was important for them, adding that they used several sources in order to select words, such as, searching FSs in google n-gram, using a dictionary (checking for words), using google search, checking from sample essays, checking dictionary for synonym and FSs, checking articles for synonym, choosing from academic word list, checking context of the word, using COCA , receiving feedback from teacher and choosing less familiar words. The fact that the participants used dictionaries, COCA corpus and background knowledge appropriately may indicate a conscious effort of doing so. When the fourth protocol question was asked about the familiarity with FSs and collocation, three out of six participants reported that they were familiar with both FSs and collocation. Other three participants stated that they were not familiar with FSs, but they knew collocation. Another protocol question (Protocol 5) asked about the attention of FSs. Four participants declared that they paid attention to the use of FSs while other two participants remarked that they used well-known FSs in their essays. When the sixth protocol question was asked about their thoughts on their ten essays in terms of formulaic sequences usage patterns, all the participants responded that there was a necessity to include more FS into their essays since they used limited number of FSs, and these FSs were similar. The seventh protocol question asked to the participants how they thought of using the FSs were underlined in their essays. Four participants reported that they used these sequences because they had these sequences in their background knowledge. It is

important to note that teacher feedback was one of the other reasons behind the participants' use of FSs. When the eighth protocol question was asked about whether there were any FSs they used often since they knew well, three of them responded that they used popular formulaic sequences, and the usage of FSs through teacher feedback was also one of the common themes for this question. One of the participants stated that he/she used frequently encountered FSs. It is important to note that the frequency of exposure is important to the use of these sequences by EFL learners. The ninth question was asked the participants whether they noticed any development with regards to the use of FSs in their essays across two semesters. Three of the participants stated that they used more FS through teacher feedback. In the same vein, three of them responded that they noticed an increase in terms of the variety of FSs and used frequent and different ranges of FSs over time. Participants did not point out an exact time for development, adding that their FSs usage developed across the course schedule. For the tenth protocol question about the previous exposure, four of them responded that they had very limited exposure in their previous educational background, but more frequent in the prep class. According to some of the participants, teacher emphasis or teacher feedback during writing courses increased their exposure of FSs, and they used these FSs more often. The last protocol question was asked about whether their awareness of FSs increased via previous encounters. All of the participants responded that previous exposures increased their awareness towards these constructions, and these exposures provide more conscious attitude and usage. They added that receiving teacher feedback, reading, writing course, conversation outside the classroom were the other reasons behind the increase of their awareness.

CONCLUSION AND RECOMMENDATIONS

The current thesis aimed to investigate the usage of formulaic sequences in longitudinal learner corpora composed of EFL learners' essays across two semesters of observation. To this aim, the analysis phase included two-fold: group and individual analysis. The first phase of the study examined collective usage of formulaic sequences over five sub-corpora of each group while the second phase investigated unique formulaic sequences used by eight individual learners over five sub-corpora. In accordance with frequency analysis, 3- to 4-word formulaic sequences were extracted from five sub-corpora of Group 1 and Group 2, and then they were classified into the Biber's et al. structural and functional taxonomy (1999; 2004). Next, the usage of three- and four-word formulaic sequences extracted from longitudinal learner corpora were compared to frequent sequences found in the native corpus (LOCNESS). In individual analysis, the formulaic sequences were extracted from particular learners' five sub-corpora, classified into structural and functional categories, and also the unique sequences from each sub-corpus were defined. Lastly, the retrospective protocol was made with the 6 participants in an attempt to better understand how they learn formulaic sequences and whether they were aware of these sequences.

Overall, the frequency analyses, correlation statistical test, and the structural and functional analyses and the protocol analyses of the FSs showed that the number and the range of FSs seemed to reveal an increasing pattern in terms of frequency, as the learners were given instruction and teacher feedback for each week during two semesters. In other words, as they get more teacher feedback their general writing quality seemed to increase in terms of FSs development, which is also validated through correlation analysis. This claim is also supported by the participants' responses to the protocol questions since they emphasized that their exposure and awareness of FSs increased through teacher feedback, so they used more FS through teacher feedback.

The findings from longitudinal design seem to provide some understanding to EFL teachers regarding the significance of formulaic sequences. As mentioned earlier, the results indicated that learners used FSs frequently but limited in type and length. EFL teachers may help to increase learners' FSs repertoire by implementing the lists of common FSs in their courses. In other words, EFL teachers need to encourage their learners and draw their attention to the use of FSs in writing. If the EFL learners are exposed to as many FSs as possible over time, the learners' FSs capacity may be enhanced. Several researchers such as Hyland (2008) and Chen and Baker (2010) emphasize also the significance of integrating FSs in writing curriculums. Overall, it can be claimed that as EFL

learners are exposed to FSs, it can be much attainable for them to use sequences in a language environment.

It is hypothesized that there may be a relation between the use of FSs and overall writing proficiency since the use of FSs can impact the quality of learners' writing. This hypothesis is supported in the literature by several researchers such as Liou and Chen (2018), Conrad and Biber, (2005) and Boers et al., (2006). Liou and Chen (2018) asserted that if L2 learners intend to be successful in writing, a high degree of proficiency in respect to the usage of formulaic sequences is required.

Limitations of the Study

In the current study, the limitations as they occurred were submitted and in fact there are still several issues that need to be revisited or explored. First of all, the study is limited to a corpus of EFL students registered at only one state university in Turkey, and so, other advanced level tertiary level EFL students were not included in this study. For this reason, the results of this study cannot be generalized to all advanced level university students in Turkey. Secondly, this study was strictly limited to investigating formulaic sequence aspects of the essays of tertiary level Turkish EFL students, which means that no other aspects (e.g., pragmatics, discourse markers, syntax) were targeted in this study. Thirdly, all the data collected with untimed essays were limited to a certain number and type, so they cannot be generalized to other essay design criteria such as timed essays. Another limitation was that operational restrictions of this study did not allow time for comparison between learner corpora and reference spoken corpora which would provide specific insight into spoken data with regards to how they are similar or different. By virtue of the fact that there was no tool to determine the structural and functional categories of obtained formulaic sequences in the process of qualitative analysis, the categorisations were carried on manually by the researcher, hence, there may have been some possible inconsistencies with the arranged items. On the other hand, time should also be noted as another limitation of this study since the data in the current study were gathered within two semesters of an academic year. If the duration of the data gathering is extended over another year, a more in-depth analysis on developmental stages could have been conducted on the data.

Implications

The results of this study have demonstrated that FSs usage patterns increase as the language proficiency of learners and teacher feedback increase in consecutive weeks. However, it also seems that there is a need for language teachers to create an air of repeated exposure in the classroom for the sequences. This must be done on a systematic basis by attaching importance to the FSs from different structural and functional categories. Similarly, it seems that there is a need for immediate

pedagogical focus on the use of FSs by the language teachers during their instruction. This pedagogical focus includes processes of “noticing”, “retrieving”, and “generating” (Hatami, 2015). The course instructors should prepare an appropriate setting to offer repeated exposure of formulaic sequences for the learners. For instance, the carefully-prepared list of these sequences may be used during the instruction. Following this, these sequences can be implemented into the teaching activities through speaking, listening, reading and writing. The findings may also be especially useful to EFL teachers for gaining insight into the development of formulaic sequences in learners’ written language. From a theoretical basis, EFL learners may recognize and increase their formulaic sequence knowledge by attaching particular focus onto the language sequences that are frequent and salient. Through a corpus-based analysis language teachers and learners may be informed about the word usage patterns that they go through during L2 development.

Pedagogical and Contextual Implications: The current study indicated that EFL learners produced invariant formulaic sequences across two semesters, and they responded that their level of awareness and the range and frequency of FSs have increased through teacher feedback. This shows that the use of FSs can be highlighted by the course instructors through teacher feedback. The use of more FSs should be integrated in the curriculums, by highlighting the FSs content of the coursebooks, and the other materials used in the courses. The integration of FSs into the course materials can hold an efficient means of increasing awareness, and they may be able to increase the usage of formulaic sequences in terms of the range and frequency over time. Particularly, there is a need for the frequently occurring FSs seen in native speakers’ written outputs to be integrated into both the receptive and productive language materials for Turkish EFL learners to be able to produce them in a more appropriate manner, and thus the frequency of use or the diversity of formulaic sequences can be increased in learners’ outputs. The pedagogically useful list of formulaic sequences for academic writing should be created by the researchers.

Theoretical Implications: The development of formulaic sequences may be affected by the selection of these sequences, that is, the salient and idiosyncratic sequences can be learned and used more easily compared to infrequent occurrences. EFL learners can be sensitive to frequency information of sequences. In this way, it is possible that the production of these sequences becomes automatized. This notion was supported by the researcher’s findings. The results of the study on the usage of FSs in each sub-corpus were also in line with the assumptions of the theory of holistic storage since the learners produced recurrently these sequences as a single lexical item such as *one of the*, *in the world* and *a lot of*. This validated that the learners acquired and used these sequences as a whole. In addition, the recurrent use of FSs in the learners’ essays can be supported by the idiom principle and lexical priming theories because of the fact that the use of formulaic sequences was affected by the idiomatic nature of these sequences. Overall, the results of the present study implied that the use and development of formulaic sequences by EFL learners can be supported by the theory of holistic storage, the idiom principle and lexical priming.

Practical Implications: It may be implied that the appropriate usage of formulaic sequences helps learners develop their essay writing. On the other hand, when the learners' awareness of FSs increased, EFL learners can be better in both written and conversational settings by using these sequences. That is to say, they can easily understand the statements uttered by the other users of English, and the use of FSs allows them to express themselves more clearly and understandably. It may improve the fluency of their utterances. The use of formulaic sequences can ensure the processing advantage on account of their holistic natures compared to retrieval of word-by-word. As the learners stated in retrospective protocols, using popular sequences makes them feel more secure because they know them well and use these FSs as a safe-belt.

Further Suggestions

The usage and development of formulaic sequences by non-native learners in conversational settings by compiling spoken corpora can also be investigated to uncover which types of FSs are more preferred in this discourse. In addition, the duration of the data gathering process can be extended over two semesters, so the intervals can be increased. By following the same procedures, new groups of learners on the basis of proficiency levels could be included in the study. Lastly, it can be suggested that larger-scale longitudinal studies are required to see the use and development of FSs with other Turkish learners in different regions. It would be beneficial to compare the structures and functions of the formulaic sequences found in the learner corpora to the sequences found in the native corpus to see whether EFL learners used the same sequences in the same way or not.

REFERENCES

- Adel, Annelie and Erman, Britt (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), 84-91.
- Adolphs, Svenja and Durow, Valerie (2004), "Social-Cultural Integration and The Development of Formulaic Sequences", Norbert Schmitt (Ed.), **Formulaic Sequences Acquisition, Processing and Use**, In (107-126.), John Benjamins, Amsterdam.
- Aijmer, Karin and Altenberg, Bengt (2004), "Introduction" (pp.1-7), In Aijmer, Karin and Altenberg, Bengt (Eds.), (2004). *Advances in Corpus Linguistics: Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23)*, Göteborg 22-26 May 2002 (No. 49). Rodopi.
- Alamri, Basim M. (2017), **Connecting Genre-Based and Corpus-Driven Approaches in Research Articles: A Comparative Study of Moves and Lexical Bundles in Saudi and International Journals**, Unpublished Doctoral Dissertation, The University of New Mexico.
- Allan, Rachel (2009), "Can a Graded Reader Corpus Provide 'Authentic' Input?", **ELT Journal Volume**, 63(1), 23-32, doi:10.1093/elt/ccn011.
- Allen, David (2009), "Lexical Bundles in Learner Writing: An Analysis of Formulaic Language in the ALESS Learner Corpus", **Komaba Journal of English Education**, 1, 105-127.
- Altenberg, Bengt (1998), "On the Phraseology of Spoken English: The Evidence of Recurrent Word Combinations", Anthony Paul Cowie (Ed.), **Phraseology: Theory, analysis, and applications**, In (101-122), Oxford University Press, Oxford.
- Altenberg, Bengt and Granger, Sylviane (2001), "The Grammatical and Lexical Patterning of MAKE in Native and Non-native Student Writing", **Applied Linguistics**, Oxford University Press, 22(2), 173-195.
- Alquraishi, Mohammed A. (2014), **Lexical Bundles in an Advanced INTOCSU Writing Class and Engineering Texts: A Functional Analysis**, Unpublished Master's Thesis, Colorado State University.
- Anthony, Laurence (2004), AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit (pp. 7-13). *Proceedings of IWLeL 2004: An Interactive Workshop on Language E-learning 2004*.

- _____ (2013), “A Critical Look at Software Tools in Corpus Linguistics”, **Linguistic Research**, 30(2), 141-161.
- Appel, Randy (2011), **Lexical Bundles in University EAP Exam Writing Samples: CAEL Test Essays**, Unpublished Master’s Thesis, Carleton University.
- Bal, Betül (2010), **Analysis of Four-word Lexical Bundles in Published Research Articles Written by Turkish Scholars**, Unpublished Master’s Thesis, Georgia State University.
- Bal-Gezegin, Betül (2019), “Lexical Bundles in Published Research Articles: A Corpus-based Study”, **Journal of Language and Linguistic Studies**, 15(2), 520-534.
- Bannard, Colin and Lieven, Elena (2012), “Formulaic Language in L1 Acquisition”, **Annual Review of Applied Linguistics**, 32, 3-16. doi: 10.1017/S0267190512000062
- Benigno, Veronica and Kraif, Olivier (2016), “Core Vocabulary and Core Collocations: Combining Corpus Analysis and Native Speaker Judgement to Inform Selection of Collocations in Learner Dictionaries”, Adriana Orlandi and Laura Giacomini (Eds.), **Defining Collocation for Lexicographic Purposes. From Linguistic Theory to Lexicographic Practice**, In (236-270), PeterLang, Bern, Switzerland.
- Bernardini, Silvia (2004). “Corpora in The Classroom: An Overview and Some Reflections on Future Developments”, John McH. Sinclair (Ed.), **How to Use Corpora in Language Teaching**, In (15-36), John Benjamins Publishing Company, Amsterdam.
- Bestgen, Yves and Granger, Sylviane (2014), “Quantifying the Development of Phraseological Competence in L2 English Writing: An Automated Approach”, **Journal of Second Language Writing**, 26, 28-41.
- Bestgen, Yves (2017), “Beyond Single-Word Measures: L2 Writing Assessment, Lexical Richness and Formulaic Competence”, **System**, 69, 65-78. <http://dx.doi.org/10.1016/j.system.2017.08.004>
- Biber, Douglas (1993), “Representativeness in Corpus Design”, **Literary and Linguistic Computing**, 8(4), 243-257.
- _____ (2006), **University Language, A Corpus Based Study of Spoken and Written Registers**, John Benjamin Publishing Company, Amsterdam/Philadelphia.
- Biber, Douglas et al. (1998), **Corpus Linguistics: Investigating Language Structure and Use**, Cambridge University Press, Cambridge.
- _____ (1999), **The Longman Grammar of Spoken and Written English**, Longman, London.
- _____ (2004), “If You Look at...: Lexical Bundles in University Teaching and Textbooks”, **Applied linguistics**, 25(3), 371-405.

- Biber, Douglas and Barbieri, Federica (2007), "Lexical Bundles in University Spoken and Written Registers", **English for specific purposes**, 26(3), 263-286.
- Biber, Douglas and Gray, Bethany (2013), "Nominalizing the Verb Phrase in Academic Science Writing", Bas Aarts, Joanne Close, Geoffrey Leech and Sean Wallis (Eds.), **The Verb Phrase in English**, In (99-132), Cambridge University Press.
- Bishop, Hugh (2004), "Noticing formulaic sequences - A problem of measuring the subjective", **LSO Working Papers in Linguistics**, 4, 15-19.
- Boers, Frank et al. (2006), "Formulaic Sequences and Perceived Oral Proficiency: Putting a Lexical Approach to the Test", **Language Teaching Research**, 10(3), 245-261. doi:10.1191/1362168806lr195oa.
- Bonelli, Elena T. (2010), "Theoretical Overview of the Evolution of Corpus Linguistics", Anne O'Keeffe and Michael McCarthy (Eds.), **The Routledge Handbook of Corpus Linguistics**, In (14-28), Routledge, London-New York.
- Boyd, Jeremy K. and Goldberg, Adele E. (2009), "Input Effects Within a Constructionist Framework", **The Modern Language Journal**, 93(3), 418-429.
- Breeze, Ruth (2013), "Lexical Bundles across Four Legal Genres", **International Journal of Corpus Linguistics**, 18(2), 229-253.
- Brooks, Gavin (2013), "Assessment and Academic Writing: A Look at the Use of Rubrics in the Second Language Writing Classroom", **Kwansei Gakuin University Humanities Review**, 17, 227-240.
- Buerki, Andreas (2016), "Formulaic Sequences: A Drop in the Ocean of Constructions or Something More Significant?", **European Journal of English Studies**, 20(1), 15-34. doi: 10.1080/13825577.2015.1136158
- Bychkovska, Tetyana and Lee, Joseph 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.
- Chen, Yu-Hua and Baker, Paul (2010), "Lexical Bundles in L1 and L2 Academic Writing", **Language Learning and Technology**, 14(2), 30-49.
- _____ (2016), "Investigating Criterial Discourse Features across Second Language Development: Lexical Bundles in Rated Learner Essays, CEFR B1, B2 and C1", **Applied Linguistics**, 37(6), 849-880. <https://doi.org/10.1093/appln/amu065>
- Cheng, Winnie et al. (2008), "Uncovering the Extent of The Phraseological Tendency: Towards a Systematic Analysis of Concgrams", **Applied Linguistics**, 30(2), 236-252. doi:10.1093/applin/amn039.

- Chenu, Florence and Jisa, Harriet (2009), "Reviewing Some Similarities and Differences in L1 And L2 Lexical Development", **Acquisition Et Interaction En Langue Étrangère**, 1, 17-38. doi: <https://doi.org/10.4000/aile.4506>
- Colson, Jean-Pierre (2008), "Cross-linguistic Phraseological Studies: An Overview", Fanny Meunier and Sylviane Granger (Eds.), **Phraseology in Foreign Language Learning and Teaching**, In (191-206), John Benjamins Publishing Company, Amsterdam / Philadelphia.
- Conklin, Kathy and Schmitt, Norbert (2008), "Formulaic Sequences: Are They Processed More Quickly Than Non-Formulaic Language by Native and Nonnative Speakers?", **Applied Linguistics**, 29(1), 1-18.
- Conrad, Susan M. (1996), "Investigating Academic Texts with Corpus-Based Techniques: An Example from Biology", **Linguistics and Education**, 8(3), 299-326.
- Conrad, Susan and Biber, Douglas (2005), "The Frequency and Use of Lexical Bundles in Conversation and Academic Prose", **Lexicographica**, 20, 56-71.
- Conrad, Susan (2000), "Will Corpus Linguistics Revolutionize Grammar Teaching in the 21st Century?", **Tesol Quarterly**, 34(3), 548-560.
- _____ (2010), "What Can a Corpus Tell Us about Grammar", Anne O'Keeffe and Michael McCarthy (Eds.), **The Routledge Handbook of Corpus Linguistics**, In (227-240), Routledge, London-New York.
- Cooper, Patricia A. (2016), **Academic Vocabulary and Lexical Bundles in The Writing of Undergraduate Psychology Students**, Unpublished Doctoral Dissertation, University of South Africa.
- Cortes, Viviana (2002), "Lexical Bundles in Freshman Composition", Randi Reppen, Susan M. Fitzmaurice and Douglas Biber (Eds.), **Using Corpora to Explore Linguistic Variation (Vol. 9)**, In (131-145). John Benjamins Publishing Company, Amsterdam.
- _____ (2004), "Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology", **English for Specific Purposes**, 23, 397-423. doi:10.1016/j.esp.2003.12.001.
- _____ (2006), "Teaching Lexical Bundles in the Disciplines: An Example from a Writing Intensive History Class", **Linguistics and Education**, 17(4), 391-406.
- _____ (2013), "The Purpose of This Study Is to: Connecting Lexical Bundles and Moves in Research Article Introductions", **Journal of English for Academic Purposes**, 12, 33-43.
- Cowie, Anthony P. (1981), "The Treatment of Collocations and Idioms in Learners' Dictionaries", **Applied Linguistics**, 2(3), 223-235.

- _____ (1988), "Stable and Creative Aspects of Vocabulary Use", Ronald Carter and Michael McCarthy (Eds.), **Vocabulary and Language Teaching**, In (126-137), Routledge, New York.
- Cowie, A. P. (1994). "Phraseology", Ronald E. Asher (Ed.), **The Encyclopedia of Language and Linguistics**, In (3168–3171), Oxford University Press, Oxford.
- Cowie, Anthony P. (Ed.), (1998), **Phraseology: Theory, Analysis, and Applications**, Oxford University Press, New York.
- Cowie, Anthony P. (1998), "Introduction", Anthony Paul Cowie (Ed.), **Phraseology: Theory, Analysis, and Applications**, In (1-20), Oxford University Press, Oxford.
- _____ (1992), "Multiword Lexical Units and Communicative Language Teaching", Pierre J. L. Arnaud and Henri Bejoint (Eds.), **Vocabulary and Applied Linguistics**, In (1-12), Macmillan, Basingstoke.
- Cowie, Anthony P. (2001), **Phraseology: Theory, Analysis, and Applications**, Oxford University Press, Oxford.
- Coxhead, Averil (2000), "A New Academic Word List" **TESOL Quarterly**, 34(2), 213-238.
- Coxhead, Averil and Byrd, Pat (2007), "Preparing Writing Teachers to Teach the Vocabulary and Grammar of Academic Prose", **Journal of Second Language Writing**, 16(3), 129-147.
- Davies, Alan (2003), **The Native Speaker: Myth and Reality**, Multilingual Matters LTD, Clevedon.
- Davies, Michael J. (2014), "An Overview of Formulaic Language and its Possible Role in L2 Fluency Development", **Ritsumeikan Language and Culture Studies**, 25(3), 221-231.
- De Cock, Sylvie (1998), "A Recurrent Word Combination Approach to the Study of Formulae in the Speech of Native and Non-native Speakers of English", **International Journal of Corpus Linguistics**, 3(1), 59-80.
- De Cook, Sylvie (2000), "Repetitive Phrasal Chunkiness and Advanced EFL Speech and Writing.", Christian Mair and Marianne Hundt (Eds.), **Corpus Linguistic and Linguistic Theory**, In (51-68), Rodopi, Amsterdam.
- Dontcheva-Navratilova, Olga (2012), "Lexical Bundles in Academic Texts by Non-native Speakers", **Brno Studies in English**, 38(2), 37-58.
- Drouhamane, Alhassane (2016), **Lexical Bundles in Master's-Level Finance Research Articles**, Unpublished Master's Thesis, Colorado State University.
- Duan, Shiping, and Shi, Zhiliang (2021), "A longitudinal study of formulaic sequence use in second language writing: Complex dynamic systems perspective", **Language Teaching Research**, 1-34. doi:10.1177/13621688211002942.

- Ellis, Nick C. (1996), "Sequencing in SLA", **Studies in Second Language Acquisition**, 18(1), 91-126.
- Ellis, Nick C. and Sinclair, Susan G. (1996), "Working Memory in the Acquisition of Vocabulary and Syntax: Putting Language in Good Order", **The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology**, 49(1), 234-250. <http://dx.doi.org/10.1080/713755604>.
- Ellis, Nick C. (1997), "Vocabulary Acquisition: Word Structure, Collocation, Grammar, And Meaning", Norbert Schmitt and Michael McCarthy (Eds.), **Vocabulary: Description, Acquisition and Pedagogy**, In (122-139), Cambridge University Press, Cambridge.
- _____ (2002), "Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition", **Studies in Second Language Acquisition**, 24(2), 143-188. doi: <https://doi.org/10.1017/S0272263102002024>.
- _____ (2008), "Phraseology: The Periphery and the Heart of Language", Fanny Meunier and Sylviane Granger (Eds.), **Phraseology in Foreign Language Learning and Teaching**, In (1-13). John Benjamins Publishing Company, Amsterdam / Philadelphia.
- _____ (2012) "Formulaic Language and Second Language Acquisition: Zipf and the Phrasal Teddy Bear", **Annual Review of Applied Linguistics**, 32, 17-44. doi: 10.1017/S0267190512000025
- _____ (2013). **Construction Grammar and Second Language Acquisition**, Oxford University Press, Oxford.
- _____ (2013), "Frequency-based Accounts of Second Language Acquisition", Gass, S. M., and Mackey, A. (Eds.), **The Routledge Handbook of Second Language Acquisition**, In (211-228). Routledge.
- Ellis, Nick C. et al. (2008), "Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL", **Tesol Quarterly**, 42(3), 375-396.
- Ellis, Nick C., and Ogden, Dave C. (2017), "Thinking about Multiword Constructions: Usage-based Approaches to Acquisition and Processing", **Topics in Cognitive Science**, 9(3), 604-620.
- Elturki, Eman M. (2015), **The Development of Formulaic Sequences: A Longitudinal Learner Corpus Investigation**, Unpublished Doctoral Dissertation, Washington State University.
- Evert, Stefan (2004), **The Statistics of Word Cooccurrences: Word Pairs and Collocations**, Unpublished Doctoral Dissertation, University of Stuttgart - Institut Für Maschinelle Sprachverarbeitung.
- Erman, Britt and Warren, Beatrice (2000), "The Idiom Principle and The Open Choice Principle", **Text**, 20(1), 29-62.

- Fattani, Reem Abdulkareem I. (2018), **Profiling Lexical Bundles in an EAP Pre-Sessional Course: A Corpus-Based Study on Textbooks and Instructors' Materials**, Unpublished Doctoral Dissertation, University of Sheffield.
- Flowerdew, John (1996), "Concordancing in Language Learning", Martha C. Pennington (Ed.), **The power of CALL**, In (97-113), Athelstan, Houston, TX.
- Foster, Pauline (2001), "Rules and Routines: A Consideration of Their Role in The Task-Based Language Production of Native and Non-native Speakers", Martin Bygate, Peter Skehan and Merrill Swain (Eds.), **Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing**, 2nd Edition in (75-93), Routledge, London.
- Garner, James et al., (2019), "N-gram Measures and L2 Writing Proficiency", **System**, 80, 176-187. doi: 10.1016/j.system.2018.12.001
- Geluso, Joe, and Yamaguchi, Atsumi (2014), "Discovering Formulaic Language Through Data-Driven Learning: Student Attitudes and Efficacy", **ReCALL**, 26(2), 225-242. doi:10.1017/S0958344014000044.
- Gibbs, Raymond, et al. (1997), "Metaphor in Idiom Comprehension", **Journal of Memory and Language**, 37(2), 141-154.
- Gilquin, Gaëtanelle and Paquot, Magali (2008), "Too Chatty. Learner Academic Writing and Register Variation", **English Text Construction**, 3(1), 41-61.
- Giora, Rachel (2003). **On Our Mind: Salience, Context, and Figurative Language**, Oxford University Press.
- Granger, Sylviane (1998), "Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae", Anthony Paul Cowie (Ed.), **Phraseology: Theory, Analysis and Applications**, In (145-60), England, Oxford University Press, Oxford.
- _____ (2002), "A Bird's-Eye View of Learner Corpus Research", Sylviane Granger, Joseph Hung, and Stephanie Petch-Tyson (Eds.), **Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching**, In (3-33), John Benjamins Publishing, Amsterdam.
- _____ (2003), "The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research", **TESOL Quarterly**, 37(3), 538-546.
- Granger, Sylviane and Tyson, Stephanie (1996), "Connector Usage in the English Essay Writing of Native and Non-native EFL Speakers of English", **World Englishes**, 15(1), 17-27.

- Granger, Sylviane, and Paquot, Magali (2008), “Disentangling the Phraseological Web”, Sylviane Granger and Fanny Meunier (Eds.), **Phraseology: An Interdisciplinary Perspective**, In (27-49), John Benjamins Publishing Company, Amsterdam / Philadelphia.
- Granger, Sylviane, and Meunier, Fanny (2008), “Phraseology in Language Learning and Teaching: Where to From Here?”, Sylviane Granger and Fanny Meunier (Eds.), **Phraseology in Foreign Language Learning and Teaching**, In (247-252), John Benjamins Publishing Company: Amsterdam / Philadelphia.
-
- _____ (2008), **Phraseology: An Interdisciplinary Perspective**, John Benjamins, Amsterdam.
- Greaves, Chris., and Warren, Martin (2010), “What Can a Corpus Tell Us About Multi-word Units”, Anne O’Keeffe and Michael McCarthy (Eds.), **The Routledge Handbook of Corpus Linguistics**, In (212-226), Routledge, London-New York.
- Gries, Stefan (2008), “Phraseology and Linguistic Theory: A Brief Survey”, In Sylviane Granger and Fanny Meunier (Eds.), **Phraseology: An Interdisciplinary Perspective**, In (3–25), John Benjamins Publishing Company, Amsterdam / Philadelphia.
- Grigaliunienu, Jone and Juknevičienė, Rita (2011), “Formulaic Language, Learner Speech and The Spoken Corpus of Learner English LINDSEI-LITH”, *Žmogus ir žodis*, 12-18.
- Güngör, Fatih (2016), **Crosslinguistic Analysis of Lexical Bundles in L1 English, L2 English, and L1 Turkish Research**, Unpublished Doctoral Dissertation, Gazi University - Graduate School of Educational Sciences.
- Güngör, Fatih and Uysal, Hacer H. (2016), “A Comparative Analysis of Lexical Bundles Used by Native and Non-native Scholars”, **English Language Teaching**, 9(6), 176-188. doi: 10.5539/elt.v9n6p176
- Halliday, Michael A. K. (2004), “Lexicology”, Halliday, Michael A. K. et al., **Lexicology and Corpus Linguistics: An Introduction**, In (1-22), Continuum, London, New York.
- Han, Sumi (2015), **Processing Formulaic Sequences by Native and Nonnative Speakers of English: Evidence from Reading Aloud**, Unpublished Doctoral Dissertation, Northern Arizona University.
- Hatami, Sarvenaz (2015), “Teaching Formulaic Sequences in the ESL Classroom”, **TESOL Journal**, 6(1), 112-129. doi: 10.1002/tesj.143.
- Hill, Jimmie (2000), “Revising Priorities: From Grammatical Failure to Collocational Success”, Michael Lewis (Ed.), **Teaching Collocation: Further Developments in the Lexical Approach**, In (47-69), Language Teaching Publications, Hove.

- Hidalgo, Encarnación; Quereda, Luis; Santana, Juan (2007) "Corpora in the Foreign Language Classroom:" Selected papers from the Sixth International Conference on Teaching and Language Corpora (TaLC 6). University of Granada, Spain, 4-7 July, 2004". BRILL.
- Housen, Alex et al., (2012), "Complexity, Accuracy and Fluency; Definitions, Measurement and Research", Alex Housen, Folkert Kuiken, and Ineke Vedder (Eds.), **Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA**, In (1-20), John Benjamins Publishing.
- Howarth, Peter (1998), "Phraseology and Second Language Proficiency", **Applied linguistics**, 19(1), 24-44.
- _____ (1998). "The Phraseology of Learners' Academic Writing", Anthony Paul Cowie (Ed.), **Phraseology: Theory, Analysis, and Application**, In (161-186), Oxford University Press, Oxford.
- Huang, Kaisheng (2014), **A Corpus Study of Chinese EFL Majors' Phraseological Performance**, Unpublished Doctoral Dissertation, University of Hong Kong, Pokfulam, Hong Kong SAR.
- Huang, Kaisheng (2015), "More Does Not Mean Better: Frequency and Accuracy Analysis of Lexical Bundles in Chinese EFL Learners' Essay Writing", **System**, 53, 13-23. <http://dx.doi.org/10.1016/j.system.2015.06.011>
- Hunston, Susan (2002), **Corpora in Applied Linguistics**, Cambridge University Press, Cambridge.
- _____ (2011). **Corpus Approaches to Evaluation: Phraseology and Evaluative Language**, Routledge, New York.
- Hyland, Ken (2004), **Disciplinary Discourses: Social Interactions in Academic Writing**, University of Michigan Press.
- Hyland, Ken (2008), "As can be seen: Lexical Bundles and Disciplinary Variation", **English for Specific Purposes**, 27(1), 4-21.
- _____ (2012), "Bundles in Academic Discourse", **Annual Review of Applied Linguistics**, 32, 150–169. doi: 10.1017/S0267190512000037
- Ishikawa, Yuka (2013), "Recurrent Word Clusters Used by Asian Learners: A Statistical Study of Differences", **Learner Corpus Studies in Asia and the World**, 1, 67-76.
- Iwatsuki, Kenichi, and Aizawa, Akiko (2018). Using Formulaic Expressions in Writing Assistance Systems. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 2678-2689). Santa Fe, New Mexico, USA, August 20-26, 2018.
- Jespersen, J. O. (1924), "Formulas and Free Expressions", **The Philosophy of Grammar**, 18-24.

- Jones, Martha, and Haywood, Sandra (2004), "Facilitating the Acquisition of Formulaic Sequences: An Exploratory Study in An EAP Context", Norbert Schmitt (Ed.), **Formulaic Sequences Acquisition, Processing and Use**, In (269-300), Benjamins, Amsterdam.
- Juknevičienė, Rita (2009), "Lexical Bundles in Learner Language: Lithuanian Learners vs. Native Speakers", **KALBOTYRA**, 61(3), 61-72.
- Karabacak, Erkan and Qin, Jingjing (2013), "Comparison of Lexical Bundles Used by Turkish, Chinese, and American University Students", **Procedia-Social and Behavioral Sciences**, 70, 622-628. doi:10.1016/j.sbspro.2013.01.101
- Kashiha, Hadi and Heng, Chan S. (2014), "Discourse Functions of Formulaic Sequences in Academic Speech across Two Disciplines", **GEMA Online® Journal of Language Studies**, 14(2), 15-27. <http://dx.doi.org/10.17576/GEMA-2014-1402-02>
- Kennedy, Graeme (2008), "Phraseology and Language Pedagogy: Semantic Preference Associated with English Verbs", Fanny Meunier and Sylviane Granger (Ed.s), **Phraseology in Foreign Language Learning and Teaching**, In (21-41), John Benjamins Publishing Company, Amsterdam / Philadelphia.
- Khusnita, Dafi and Rukmini, Dwi (2016), "The EFL Learners' Perceptions and Realizations of Formulaic Sequences in Casual Conversation", **English Education Journal**, 6(2), 68-78.
- Kilgarriff, Adam et al. (2004), "The Sketch Engine", **Proceedings of EURALEX, European Association for Lexicography**, July, 105-116.
- Kilgarriff, Adam (2005), "Language is Never, Ever, Ever, Random", **Corpus Linguistics and Linguistic Theory**, 1(2), 263-276.
- Kilgarriff, Adam et al. (2014), "The Sketch Engine: Ten Years on", **Lexicography**, 1(1), 7-36.
- Kashiha, Hadi and Chan, Swee Heng (2014), "Discourse Functions of Formulaic Sequences in Academic Speech across Two Disciplines", **GEMA Online® Journal of Language Studies**, 14(2), 15-27.
-
- (2015), "A Little Bit About: Differences in Native and Non-native Speakers' Use of Formulaic Language", **Australian Journal of Linguistics**, 35(4), 297-310. doi: 10.1080/07268602.2015.1067132
- Kuosmanen, Anna (2020), **The Relationship Between Knowledge of Formulaic Sequences and Willingness to Communicate in Finnish EFL Speakers: A Correlational Study**, Unpublished Master's Thesis, University of Jyväskylä.
- Kyle, Kristopher and Crossley, Scott A. (2015), "Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application", **TESOL Quarterly**, 49(4), 757-786.

- Lenko-Szymanska, Agnieszka (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>.
- Lewis, Michael (2000), "Language in the Lexical Approach", Michael Lewis (Ed.), **Teaching Collocation: Further Developments in the Lexical Approach**, In (126-154), Language Teaching Publications, Hove.
- _____ (2000), "Learning in the Lexical Approach", Michael Lewis (Ed.), **Teaching Collocation: Further Developments in the Lexical Approach**, In (155-185), Language Teaching Publications, Hove.
- _____ (2000), "Introduction", Michael Lewis (Ed.), **Teaching Collocation: Further Developments in the Lexical Approach**, In (8-9), Language Teaching Publications, Hove.
- _____ (2002), **The Lexical Approach**, Thomson, Boston.
- Li, Jie and Schmitt, Norbert (2009), "The Acquisition of Lexical Phrases in Academic Writing: A Longitudinal Case Study", **Journal of second Language Writing**, 18(2), 85-102.
- Liou, Hsien C. and Chen, Weng F. (2018), "Effects of Explicit Instruction on Learning Academic Formulaic Sequences for EFL College Learners' Writing", **Taiwan Journal of TESOL**, 15(1), 61-100.
- Lozano, Cristóbal and Mendikoetxea, Amaya (2007), "Learner corpora and the acquisition of word order: A study of the production of Verb-Subject structures in L2 English", in M. Davies, P. Rayson, S. Hunston and P. Danielsson (eds.), *Proceedings of the Corpus Linguistics Conference 2007*, University of Birmingham.
- Ma, GH (2009), "Lexical Bundles in L2 Timed Writing of English Majors", **Foreign Language Teaching and Research**, (41) 54-60.
- Martinez, Ron (2011), **The Development of a Corpus-Informed List of Formulaic Sequences for Language Pedagogy**, Unpublished Doctoral Dissertation, University of Nottingham.
- Martinez, Ron and Schmitt, Norbert (2012), "A Phrasal Expressions List", **Applied Linguistics**, 33(3), 299-320. doi:10.1093/applin/ams010
- McCarthy, Michael (1991), **Discourse Analysis for Language Teachers**, Cambridge University Press, Cambridge.
- Mel'cuk, Igor (1998), "Collocations and Lexical Functions", Anthony Paul Cowie (Ed.), **Phraseology: Theory, Analysis, and Applications**, In (23-53), Clarendon Press, Oxford.
- Meyer, Charles F. (2002). **English Corpus Linguistics: An Introduction**, Cambridge University Press, Cambridge.

- Millar, Neil (2011), "The Processing of Malformed Formulaic Language", **Applied Linguistics**, 32(2), 129-148.
- Müjdeci, Şafak (2014), **The Effects of Focused Instruction on the Receptive and Productive Knowledge of Formulaic Sequences**, Unpublished Master's Thesis, Gazi University - Institute of Educational Sciences.
- Nation, I. S. P. and Waring, R. (1997), "Vocabulary Size, Text Coverage and Word Lists", Norbert Schmitt and Michael McCarthy (Eds.), **Vocabulary: Description, Acquisition and Pedagogy**, In (6-19), Cambridge University Press, Cambridge.
- Nation, I. S. P. (2001), **Learning Vocabulary in Another Language**, Cambridge University Press, Cambridge.
- Nattinger, James R. (1980), "A lexical phrase grammar for ESL", **TESOL Quarterly**, 14(3), 337-344.
- Nattinger, James R. and DeCarrico, Jeanette S. (1992), **Lexical Phrases and Language Teaching**, Oxford University Press, Oxford.
- Nesselhauf, Nadja (2003), "The Use of Collocations by Advanced Learners of English and Some Implications for Teaching", **Applied Linguistics**, 24(2) 223-242. doi:10.1093/applin/24.2.223.
- _____ (2004), "Learner Corpora and Their Potential for Language Teaching", John McH. Sinclair (Ed.), **How to Use Corpora in Language Teaching**, In (125-152), John Benjamins Publishing Company, Amsterdam.
- _____ (2004), "What are collocations?" D. J. Allerton, Nadja Nesselhauf, and Paul Skandera (Eds.), **Phraseological Units: Basic Concepts and Their Application**, In (1- 21), Schwabe, Basel.
- _____ (2005), **Collocations in a Learner Corpus**, John Benjamins, Netherlands.
- Oakey, David (2002). "Formulaic Language in English Academic Writing: A Corpus-Based Study of The Formal and Functional Variation of a Lexical Phrase in Different Academic Disciplines", Randi Reppen, Susan Fitzmaurice and Douglas Biber (Eds.), **Using Corpora to Explore Linguistic Variation: Studies in Corpus Linguistics**, In (111-130), John Benjamins, Amsterdam.
- Oakey, David (2020), "Phrases in EAP Academic Writing Pedagogy: Illuminating Halliday's Influence on Research and Practice", **Journal of English for Academic Purposes**, 44, 1-16.
- Ohlrogge, Aaron (2009), "Formulaic Expressions in Intermediate EFL Writing Assessment", Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali and Kathleen M. Wheatley (Eds.), **Formulaic Language: Volume 2, Acquisition, Loss, Psychological Reality, and Functional Explanations**, In (375-386), John Benjamins Publishing Company, Amsterdam / Philadelphia.

- Ortega, Lourdes, and Iberri-Shea, Gina (2005), “Longitudinal Research in Second Language Acquisition: Recent Trends and Future Directions”, **Annual Review of Applied Linguistics**, 25, 26-45.
- Osborne, Jason (2002), “Notes on the Use of Data Transformations”, **Practical Assessment, Research, and Evaluation**, 8(6), 1-7. doi: <https://doi.org/10.7275/4vng-5608>
- O'Donnell, Matthew et al. (2013). “The Development of Formulaic Sequences in First and Second Language Writing: Investigating Effects of Frequency, Association, and Native Norm”, **International Journal of Corpus Linguistics**, 18(1), 83-108. <https://doi.org/10.1075/ijcl.18.1.07odo>
- O'Keeffe, Anne et al., (2007), **From Corpus to Classroom: Language Use and Language Teaching**, Cambridge University Press.
- Özbay, Ali Şükrü (2015), **Corpus Analysis of the Support Verb Construction Development and Use by EFL Learners**, Unpublished Doctoral Dissertation, Karadeniz Technical University - Institute of Social Sciences.
- Özbay, Ali Şükrü and Kayaoğlu, Mustafa Naci (2016), “Computerized Corpus Based Investigation of the Use of Multi Word Combinations and the Developmental Stages by Tertiary Level EFL Learners”, **EPiC Series in Language and Linguistics**, 1, 341-350.
- Öztürk, Yusuf and Köse, Gül D. (2016), “Turkish and Native English Academic Writers’ Use of Lexical Bundles”, **Journal of Language and Linguistic Studies**, 12(1), 149-165.
- Paikeday, Thomas M. (1985), **The Native Speaker is Dead!**, Paikeday, Toronto and New York.
- Pang, Ping (2009), “A Study on the Use of 4-Word Bundles in Argumentative Essays by Chinese English Majors: A Comparative Study Based on WECLL and LOCNESS”, **CELEA Journal**, 32(3), 25-45.
- Pastor, Corpas and Mitkov, Ruslan (2019), Preface, Computational and corpus-based phraseology. Third International Conference, Europhras 2019, Malaga, Spain, September 25-27, 2019 Proceedings, Springer International Publishing.
- Paquot, Magali and Granger, Sylviane (2012), “Formulaic Language in Learner Corpora”, **Annual Review of Applied Linguistics**, 32, 130-149. doi: 10.1017/S0267190512000098.
- Pavesi, Caterina (2013), **Recurrent Sequences in a Learner Corpus of Computer-Mediated Communication**, Unpublished Doctoral Dissertation, Università Cattolica del Sacro Cuore.
- Pawley, Andrew and Syder, Frances (1983), “Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency”, Jack C. Richards and R.W. Schmidt (Eds.), **Language and Communication**, In (191–225), Routledge.

- Peters, Elke and Pauwels, Paul (2015), "Learning Academic Formulaic Sequences", **Journal Of English for Academic Purposes**, 20, 28-39. <http://dx.doi.org/10.1016/j.jeap.2015.04.002>
- Philip, Gill (2008), "Reassessing the Canon: "Fixed" Phrases in General Reference Corpora", Sylviane Granger and Fanny Meunier (Eds.), **Phraseology: An Interdisciplinary Perspective**, In (95-109), John Benjamins Publishing Company, Amsterdam / Philadelphia.
- Piirainen, Elisabeth (2008), "Figurative Phraseology and Culture", Sylviane Granger and Fanny Meunier (Eds.), **Phraseology: An Interdisciplinary Perspective**, In (207-228), John Benjamins Publishing Company, Amsterdam / Philadelphia.
- Qureshi, Abrar Hussain and Akhter, Shamim (2019), "Core English Vocabulary for EFL Learners: A Corpus-Based Perspective", **Pakistan Journal of Social Sciences (PJSS)**, 39(2), 505-512.
- Ranjbar, Nosrat et al. (2012), "The Effect of Lexical Bundles on Iranian EFL Learners Linguistic Production Fluency", **International Education Studies**, 5(4), 243-251. doi:10.5539/ies.v5n4p243.
- Razak, Harniza Binti Abd (2015), **An analysis of ESL lexical bundles in learners' academic writing**, Unpublished Master's Thesis, International Islamic University Malaysia.
- Reppen, Randi (2010), "Building a Corpus: What Are The Key Considerations?", Anne O'Keeffe and Michael McCarthy (Eds.), **The Routledge Handbook of Corpus Linguistics**, In (31-37), Routledge, London-New York.
- Richards, Jack C. and Rogers, Theodore. S. (2001), **Approaches and Methods in Language Teaching**, 2nd Edition, Cambridge University Press, Cambridge.
- Römer, Ute (2010), "Establishing the Phraseological Profile of a Text Type: The Construction of Meaning in Academic Book Reviews", **English Text Construction**, 3(1), 95-119. <https://doi.org/10.1075/etc.3.1.06rom>
- Salazar, Danica J. L. (2011), **Lexical Bundles in Scientific English: A Corpus-Based Study of Native and Non-native Writing**, Unpublished Doctoral Dissertation, Universitat de Barcelona.
- Schauer, Gila A., and Adolphs, Svenja (2006), "Expressions of Gratitude in Corpus and DCT Data: Vocabulary, Formulaic Sequences, and Pedagogy", **System**, 34(1), 119-134. doi:10.1016/j.system.2005.09.003.
- Schmitt, Norbert (2000), **Vocabulary in Language Teaching**, Cambridge University Press, Cambridge.
- Schmitt, Norbert (2004), **Formulaic Sequences: Acquisition, Processing and Use**, John Benjamins, Amsterdam.

- _____ (2005). Processing advantages of formulaic sequences vs non-formulaic sequences in text passages. In C. Cosme, C. Gouverneur, F. Meunier and M. Paquot (Eds.), *Proceedings of the Phraseology 2005 conference* (381–382). Universite Catholique de Louvain: Louvain-la-Neuve, 13–15 October 2005.
- _____ (2013), “Formulaic Language and Collocation”, Carol A. Chapelle (Ed.), **The Encyclopedia of Applied Linguistics**, In (2190–2200), Blackwell Publishing Ltd.
- Schmitt, Norbert and Carter, Ronald (2004), “Formulaic Sequences in Action: An Introduction.”, Norbert Schmitt (Ed.), **Formulaic Sequences: Acquisition, Processing and Use**, In (1-22), John Benjamins, Amsterdam.
- Schmitt, Norbert and Underwood, Geoffery (2004), “Exploring The Processing of Formulaic Sequences through a Self-Paced Reading Task”, Norbert Schmitt (Ed.), **Formulaic Sequences: Acquisition, Processing and Use**, In (173-190), John Benjamins, Amsterdam.
- Schmitt, Norbert, et al., (2004), “Knowledge and Acquisition of Formulaic Sequences: A Longitudinal Study”, Norbert Schmitt (Ed.), **Formulaic Sequences: Acquisition, Processing and Use**, In (55-86), John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Shrout, Patrick and Fleiss, Joseph (1979), “Intraclass Correlations: Uses in Assessing Rater Reliability”, **Psychological Bulletin**, 86(2), 420-428. DOI: 10.1037/0033-2909.86.2.420
- Simpson-Vlach, Rita and Ellis, Nick C. (2010), “An Academic Formulas List: New Methods in Phraseology Research”, **Applied Linguistics**, 31(4), 487-512. doi:10.1093/applin/amp058.
- Sinclair, John (1991), **Corpus, Concordance, Collocation**, Oxford University Press, Oxford.
- _____ (2004), **Trust the Text: Language, Corpus and Discourse**, Routledge, London.
- _____ (2008), “The Phrase, The Whole Phrase, and Nothing But The Phrase”, Sylviane Granger and Fanny Meunier (Eds.), **Phraseology: An Interdisciplinary Perspective**, In (407-410), John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Staples, Shelley et al. (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. <http://dx.doi.org/10.1016/j.jeap.2013.05.002>.
- Staples, Shelley and Reppen, Randi (2016), “Understanding First-Year L2 Writing: A Lexico-Grammatical Analysis across L1s, Genres, and Language Ratings”, **Journal of Second Language Writing**, 32, 17-35.
- Stubbs, Michael (2007), “On Texts, Corpora and Models of Language”, Wolfgang Teubert and Michaela Mahlberg (Eds.), **Text, Discourse and Corpora: Theory and Analysis**, In (127-161), Continuum, London.

- Svartvik, J. (1992). Corpus linguistics comes of age. In J. Svartvik (ed.) *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*. Berlin and New York: Mouton de Gruyter, pp. 7-16.
- Stengers, Helene et al. (2011), "Formulaic Sequences and L2 Oral Proficiency: Does the Type of Target Language Influence the Association?", **International Review of Applied Linguistics (IRAL)** 49, 321-343.
- Siyanova, Anna and Schmitt, Norbert (2008), "L2 learner production and processing of collocation: A multi-study perspective", **The Canadian Modern Language Review**, 64(3), 429-458.
- Siyanova-Chanturia, Anna and Spina, Stefania (2020), "Multi-word Expressions in Second Language Writing: A Large-Scale Longitudinal Learner Corpus Study", **Language Learning**, 70(2), 420-463.
- Şanal, Fahrettin (2007), **A Learner Corpus Based Study on Second Language Lexicology of Turkish Students of English**, Unpublished Doctoral Dissertation, Çukurova University - The Institute of Social Sciences.
- Taşkaya, Hakan (2019), **Phraseological Awareness of Tertiary Level Turkish EFL Learners as Evidenced in Native and Non-native Corpora**, Unpublished Master's Thesis, Karadeniz Technical University - The Institute of Social Sciences.
- Tekmen, Anne F. and Daloğlu, Ayşegül (2006), "An Investigation of Incidental Vocabulary Acquisition in Relation to Learner Proficiency Level and Word Frequency", **Foreign Language Annals**, 39(2), 220-243.
- Tognini-Bonelli, Elena (2001), **Corpus Linguistics at Work**, John Benjamins, Amsterdam/Philadelphia.
- Tomankova, Veronika (2016), "Lexical Bundles in Legal Texts Corpora—Selection, Classification and Pedagogical Implications", **Discourse and Interaction**, 9(2), 75-94. doi: 10.5817/DI2016-2-75
- Tribble, Chris and Jones, Glyn (1990), **Concordances in The Classroom: A Resource Book for Teachers**, Longman, Essex.
- Ulfa, Nadia and Muthalib, Kismullah A. (2020), "Lexical Bundles in Students' Essay Writing", **English Education Journal (EEJ)**, 11(3), 367-379.
- Underwood, Geoffrey (2004), "The Eyes Have It: An Eye-Movement Study into The Processing of Formulaic Sequences", Norbert Schmitt (Ed.), **Formulaic Sequences**, In (155–72), John Benjamins, Amsterdam.
- Üstünbaş, Ümran (2014), **The Use Formulaic Language by English as a Foreign Language (EFL) Learners in Oral Proficiency Exams**, Unpublished Master' Thesis, Bilkent University.

- Vidaković, Ivana and Barker, Fiona (2009), “Lexical Development across Second Language Proficiency Levels: A Corpus-Informed Study”, In R. M. Millar & M. Durham (Eds.), *Language, Learning & Context: Proceedings of the 42nd Annual Meeting of the British Association for Applied Linguistics* (pp. 143-146). London: Scitsiugnill Press.
- Wang, Ying (2016), **The Idiom Principle and L1 Influence. A Contrastive Learner-corpus Study of Delexical Verb + Noun Collocations**, John Benjamins Publishing Company.
- Webb, Stuart et al. (2013), “Incidental learning of collocation”, **Language Learning**, 63, 91-120.
- Weinert, R. (1995), “The Role of Formulaic Language in Second Language Acquisition: A review”, **Applied Linguistics**, 16(2), 180-205.
- Wray, Alison and Perkins, Michael (2000), “The Functions of Formulaic Language: An Integrated Model”, **Language and Communication**, 20(1), 1-28.
- Wray, Alison and Fitzpatrick, Tess (2008), “Why Can’t You Just Leave It Alone? Deviations from Memorized Language as a Gauge of Nativelike Competence”, Fanny Meunier and Sylviane Granger (Eds.), **Phraseology in Foreign Language Learning and Teaching**, In (123-147), John Benjamins, Amsterdam.
- _____ (2010), “Pushing Learners to the Extreme: the Artificial Use of Prefabricated Material in Conversation”, **Innovation in Language Learning and Teaching**, 4(1), 37-51.
- Wang, Ping (2009), “The Inter-Rater Reliability in Scoring Composition”, **English Language Teaching**, 2(3), 39-43.
- Wang, Ying (2018), “As Hill Seems to Suggest: Variability in Formulaic Sequences with Interpersonal Functions in L1 Novice and Expert Academic Writing”, **Journal of English for Academic Purposes**, 33, 12-23. <https://doi.org/10.1016/j.jeap.2018.01.003>.
- Warren, Martin (2006), **Features of Naturalness in Conversation**, John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Wei, Liu and Ying, Huo (2011), “On the role of formulaic sequences in second language acquisition”, **US-China Foreign Language**, 9(11), 708-713.
- Wood, David (2002), “Formulaic Language in Acquisition and Production: Implications for Teaching”, **TESL Canada Journal**, 20(1), 1-15.
- _____ (2010), **Formulaic Language and Second Language Speech Fluency: Background, Evidence and Classroom Applications**, Bloomsbury Publishing.
- _____ (2015), **Fundamentals of Formulaic Language: An Introduction**, Bloomsbury Publishing, London.

- Wray, Alison (2000), "Formulaic sequences in second language teaching: principle and practice", **Applied Linguistics**, 21(4), 463-489.
- _____ (2002), **Formulaic Language and the Lexicon**, Cambridge University Press, New York.
- _____ (2013), "Formulaic Sequences", Carol A. Chapelle (Ed.), **The Encyclopedia of Applied Linguistics**, In (2200-2205), Blackwell Publishing Ltd.
- _____ (2017), "Formulaic Sequences as a Regulatory Mechanism for Cognitive Perturbations During the Achievement of Social Goals", **Topics in Cognitive Science**, 9(3), 569-587. doi: 10.1111/tops.12257
- Wray, Alison and Perkins, Micheal R. (2000), "The Functions of Formulaic Language: An Integrated Model", **Language and Communication**, 20(1), 1-28.
- Wulff, Stefanie (2019), "Acquisition of Formulaic Language from a Usage-Based Perspective", Anna Siyanova-Chanturia and Ana Pellicer-Sánchez (Eds.), **Understanding Formulaic Language: A Second Language Acquisition Perspective**, In (19-37), Routledge, Taylor and Francis.
- Wong, Ho Put (2012), **Use of Formulaic Sequences in Task-based Oral Production of Chinese**, Unpublished Doctoral Dissertation, Durham University.
- Xu, Qi (2016), "Formulaic Sequences and the Implications for Second Language Learning", **English Language Teaching**, 9(8), 39-45. doi:10.5539/elt.v9n8p39
- Qi, Yan and Ding, Yanren (2011), "Use of Formulaic Sequences in Monologues of Chinese EFL Learners", **System**, 39(2), 164-174. doi:10.1016/j.system.2011.02.003



APPENDIXES

Appendix 1: Rubric

Table 58: Rubric

	SCORE	LEVEL	CRITERIA
CONTENT		30-27	EXCELLENT TO VERY GOOD: knowledgeable • substantive • thorough development of thesis • relevant to assigned topic
		26-22	GOOD TO AVERAGE: some knowledge of subject • adequate range • limited development of thesis • mostly relevant to topic, but lacks detail
		21-17	FAIR TO POOR: limited knowledge of subject • little substance • inadequate development of topic
		16-13	VERY POOR: does not show knowledge of subject • non-substantive • not pertinent • OR not enough to evaluate
ORGANIZATION		20-18	EXCELLENT TO VERY GOOD: fluent expression • ideas clearly stated/ supported • succinct • well-organized • logical sequencing • cohesive
		17-14	GOOD TO AVERAGE: somewhat choppy • loosely organized but main ideas stand out • limited support • logical but incomplete sequencing
		13-10	FAIR TO POOR: non-fluent • ideas confused or disconnected • lacks logical sequencing and development
		9-7	VERY POOR: does not communicate • no organization • OR not enough to evaluate
VOCABULARY		20-18	EXCELLENT TO VERY GOOD: sophisticated range • effective word/idiom choice and usage • word form mastery • appropriate register
		17-14	GOOD TO AVERAGE: adequate range • occasional errors of word/idiom form, choice, usage but meaning not obscured
		13-10	FAIR TO POOR: limited range • frequent errors of word/idiom form, choice, usage • meaning confused or obscured
		9-7	VERY POOR: essentially translation • little knowledge of English vocabulary, idioms, word form • OR not enough to evaluate
LANGUAGE USE		25-22	EXCELLENT TO VERY GOOD: effective complex constructions • few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions
		21-18	GOOD TO AVERAGE: effective but simple constructions • minor problems in complex constructions • several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions but meaning seldom obscured
		17-11	FAIR TO POOR: major problems in simple/complex constructions • frequent errors of negation, agreement, tense, number, word order/function, articles, pronouns, prepositions and/or fragments, run-ons, deletions • meaning confused or obscured
		10-5	VERY POOR: virtually no mastery of sentence construction rules • dominated by errors • does not communicate • OR not enough to evaluate
MECHANICS		5	EXCELLENT TO VERY GOOD: demonstrates mastery of conventions • few errors of spelling, punctuation, capitalization, paragraphing
		4	GOOD TO AVERAGE: occasional errors of spelling, punctuation, capitalization, paragraphing but meaning not obscured
		3	FAIR TO POOR: frequent errors of spelling, punctuation, capitalization, paragraphing • poor handwriting • meaning confused or obscured
		2	VERY POOR: no mastery of conventions • dominated by errors of spelling, punctuation, capitalization, paragraphing • handwriting illegible • OR not enough to evaluate

Source: Brooks, 2013: 227-240

Appendix 2: Weekly Essay Topics

Table 59: Weekly Essay Topics

Essays	Essay Topics
Essay 1	Have people become overly dependent on technology?
Essay 2	Many teachers assign homework to students every day. Do you think that daily homework is necessary for students? Use specific reasons and details to support your answer.
Essay 3	Write about one of the following topics: 1. Most believe the university education aims to help graduates to find jobs that are better. However, there are also some people who think that the benefits of university education for both people and society are actually more than “helping to find better jobs”. Give your opinion regarding this situation and give reasons for your answer with examples. 2. Most believe that the best way to reduce crime is to give longer prison sentences. However, other people think that the alternative ways of reducing crime is also the case. Give your opinion regarding this situation and give reasons for your answer with examples. 3. Many people from different levels of the society think that animals should not be exploited by people. But other people think that people should use animals to for food and research purposes. Give your opinion regarding this situation and give reasons for your answer with examples. 4. Many people from different levels of the society think that the principal factors affecting the children`s development may be such things as television, internet friends, and music. However, there also some people who believe that “the family” is still a more important factor in children`s development. Give your opinion regarding this situation and give reasons for your answer with examples. 5. Some working parents think that childcare centres provide the best care for children who are still too young to go to school. Other working parents think that family members such as grandparents will be better carers for their children. Give your opinion regarding this situation and give reasons for your answer with examples.
Essay 4	Please discuss “Birth control in Turkey”. Agree or Disagree?
Essay 5	Because of the busy pace of modern life, many children spend most of their time indoors and have little exposure to the natural world. How important is it for children to learn to understand and appreciate nature? Give reasons for your answer and include any relevant examples from your own knowledge or experience.
Essay 6	How can people live together in peace and harmony?
Essays 7	It is generally seen that the latest developments in the field of artificial intelligence is likely to have a positive impact for our lives in years to come. Other people, on the other hand, are worried that they are we are not ready for a world, where computers will be more clever and dominant than us. Give your opinion about this topic by making your position clear to us. Don`t forget to mention “opposing ideas” and then try to refute this by your own ideas.
Essay 8	It is a fact that many people in our society think that “price” is the most important criteria while they are buying something in the market (cell phone, dress, shoes...) Agree or disagree?
Essay 9	Although it is fact that the contribution of women is great in various fields such as in education and employment, they are nevertheless treated unequally in the workplace when it comes to pay and promotion. Equality of opportunity should be promoted for men and women in the workplace? Agree or disagree?
Essay 10	Genetically modified food.

Appendix 3: Top 100 Frequent Formulaic Sequences in LOCNESS

Table 60: Top 100 Frequent Formulaic Sequences in LOCNESS

3-4 FSs	Raw Freq	Norm Freq	3-4 FSs	Raw Freq	Norm Freq	3-4 FSs	Raw Freq	Norm Freq
the fact that	162	22,46	for the best	48	6,65	can be seen	34	4,71
in order to	130	18,02	to have a	47	6,52	would not be	34	4,71
one of the	123	17,05	they do not	47	6,52	in the case	34	4,71
the united states	117	16,22	there are many	47	6,52	most of the	34	4,71
that it is	104	14,42	as a result	45	6,24	many of the	33	4,57
there is no	94	13,03	is not the	45	6,24	it can be	33	4,57
be able to	94	13,03	of the play	45	6,24	as it is	33	4,57
the right to	84	11,64	out of the	45	6,24	that there is	33	4,57
it is not	83	11,51	the invention of	44	6,10	is a very	33	4,57
due to the	82	11,37	the rest of	43	5,96	a part of	33	4,57
the end of	82	11,37	at the end of	42	5,82	is that the	33	4,57
because of the	80	11,09	that he is	42	5,82	as a whole	33	4,57
the idea of	77	10,67	to be the	42	5,82	I feel that	33	4,57
there is a	77	10,67	the question of	41	5,68	the introduction of	32	4,44
as well as	76	10,54	all of the	40	5,55	I think that	32	4,44
it is a	70	9,70	this is the	40	5,55	because they are	32	4,44
the use of	69	9,57	such as the	40	5,55	I believe that	32	4,44
this is a	68	9,43	is one of	39	5,41	a loss of	31	4,30
the end of the	67	9,29	is not a	39	5,41	as long as	31	4,30
in the united	67	9,29	this is not	39	5,41	would have to	31	4,30
should not be	66	9,15	it has been	39	5,41	one of the most	31	4,30
the number of	65	9,01	the case of	39	5,41	whether or not	30	4,16
in the world	64	8,87	the beginning of	38	5,27	of the European	30	4,16
of the world	64	8,87	aware of the	38	5,27	an example of	30	4,16
to be a	62	8,59	there would be	38	5,27	men and women	30	4,16
a lot of	61	8,46	the amount of	38	5,27	the only way	30	4,16
that they are	61	8,46	of the most	37	5,13	in the future	30	4,16
cannot be	60	8,32	would be a	37	5,13	is one of the	30	4,16
it is the	58	8,04	the invention of the	37	5,13			
in the united states	57	7,90	the majority of	37	5,13			
it would be	57	7,90	because it is	36	4,99			
part of the	57	7,90	in the past	36	4,99			
at the end	56	7,76	but it is	35	4,85			
have to be	51	7,07	some of the	35	4,85			
the people of	51	7,07	there has been	35	4,85			
on the other hand	50	6,93	more and more	35	4,85			

Appendix 4: Individual Learners' Formulaic Sequence Tables

Table 61: Frequent Formulaic Sequences Produced by Gizem across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
as long as	2	81,10	noun+ who are	3	100,40	in my opinion	3	90,14	I cannot	4	96,29	in the world	6	96,81
and quality of	2	81,10	noun+ better than	2	66,93	a lot of	3	90,14	to buy a	3	72,22	some of the	5	80,67
data which are	2	81,10	most of these	2	66,93	and it is	3	90,14	there is a	3	72,22	I do not	3	48,4
far from the	2	81,10	because of the	2	66,93	to live in	3	90,14	we need to	3	72,22	have shown that	3	48,4
it should not be	2	81,10	in rural areas	2	66,93	one of the	3	90,14	but at least the	2	48,15	in higher positions	3	48,4
of our age	2	81,10	it is a	2	66,93	know each other better	2	60,10	and durability of a	2	48,15	the formation of	3	48,4
it may be	2	81,10	there are many	2	66,93	he claims that	2	60,10	for long term	2	48,15	the number of	3	48,4
the sense of	2	81,10	she can ask	2	66,93	you know what	2	60,10	cannot help	2	48,15	in the same time	3	48,4
which are learned in	2	81,10	shows that the	2	66,93	how many of you	2	60,10	are more important than	2	48,15	be able to	2	32,27
will turn into	2	81,10				there are a	2	60,10	it is not	2	48,15	by changing the	2	32,27
						to get outside	2	60,10	it will be	2	48,15	can be seen	2	32,27
						to know each other	2	60,10	made by human	2	48,15	and that in some	2	32,27
						we should take	2	60,10	of a product	2	48,15	according to their	2	32,27
						you have to	2	60,10	pay a lot	2	48,15	have not been able	2	32,27
						to protect their	2	60,10	in case of	2	48,15	I think that	2	32,27
						to teach the	2	60,10	for a product	2	48,15	if they are	2	32,27
						an opportunity to	2	60,10	according to reports	2	48,15	in my opinion	2	32,27
						a part of	2	60,10	states that these	2	48,15	in some parts of	2	32,27
									that it is	2	48,15	and this is a	2	32,27
									the end of	2	48,15	of the world it	2	32,27
									the lower is the	2	48,15	one of the	2	32,27
									the most important criteria	2	48,15	increase the number	2	32,27
									is not everything	2	48,15	reduce the number of	2	32,27
									in my opinion	2	48,15	should not be	2	32,27

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
									the ways to	2	48,15	is a serious	2	32,27
									is not the	2	48,15	are at the	2	32,27
									to make a	2	48,15	that they cannot	2	32,27
									which has no	2	48,15	it may be beneficial	2	32,27
									will be possible	2	48,15	may be beneficial to	2	32,27
									they do not	2	48,15	there is no	2	32,27
									it can be	2	48,15	they change their	2	32,27
									a product for	2	48,15	to be a	2	32,27
									because of their	2	48,15	to be taken	2	32,27
												to their skills	2	32,27
												transferred to the	2	32,27
												will increase the	2	32,27
												an important role to	2	32,27
												smb+ claims that	2	32,27

Table 62: Frequent Formulaic Sequences Produced by Elif across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSS	Raw Freq	Normed Freq	3-4 FSS	Raw Freq	Normed Freq	3-4 FSS	Raw Freq	Normed Freq	3-4 FSS	Raw Freq	Normed Freq	3-4 FSS	Raw Freq	Normed Freq
in this case	3	124,58	the fact that	3	115,92	due to the	4	133,87	of the person	6	167,22	in the world	13	204,21
for the students	3	124,58	due to the	3	115,92	on the contrary	3	100,40	the end of	4	111,48	I think that	7	109,96
can say that	2	83,06	day by day	2	77,28	inside than outside	2	66,93	the most important	3	83,61	he claims that	5	78,54
he or she	2	83,06	slowing down the	2	77,28	is difficult to	2	66,93	when people are	3	83,61	the increase of	5	78,54
it has a	2	83,06	cannot adapt the	2	77,28	is that the	2	66,93	end of the	3	83,61	the inadequacy of	4	62,83
they do not	2	83,06	due to the fact	2	77,28	more time inside than	2	66,93	due to the	2	55,74	more nutritious than	4	62,83
think that it	2	83,06	at least three	2	77,28	can be seen	2	66,93	of the person in	2	55,74	is not important	4	62,83
most of the	2	83,06	firmly believe that	2	77,28	the busy pace of	2	66,93	failed to present	2	55,74	that with the	4	62,83
according to the	2	83,06	they try to	2	77,28	the importance of	2	66,93	fails to mention	2	55,74	I believe that	3	47,13
a lot of	2	83,06	rate of the	2	77,28	the value of	2	66,93	as the end of	2	55,74	they think that	3	47,13
			that focused on	2	77,28	of nature on	2	66,93	he mentioned that	2	55,74	on the other hand	3	47,13
			have you ever thought	2	77,28	who live in	2	66,93	the price of	2	55,74	as a result of	3	47,13
			the difference between	2	77,28	one of the	2	66,93	when they are	2	55,74	who is a	3	47,13
			the pace of	2	77,28	in final consideration	2	66,93	on my research	2	55,74	as well as	3	47,13
			the rate of	2	77,28				one of the	2	55,74	he says that	3	47,13
			I firmly believe that	2	77,28				regardless of the	2	55,74	that it is	3	47,13
			the reduction rate of	2	77,28				I do feel that	2	55,74	that there is	3	47,13
			issues such as	2	77,28				in the article	2	55,74	I do not think	3	47,13
			to increase population	2	77,28				status of the person	2	55,74	the rate of	3	47,13
			which I mentioned above	2	77,28				such as the	2	55,74	in terms of	3	47,13
			who are cared by	2	77,28				the end of the	2	55,74	do not think that	3	47,13
			who are too young	2	77,28				is not a	2	55,74	will not be	3	47,13
			who is Turkish	2	77,28				the standard of	2	55,74	gives an example	2	31,42

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
			you ever thought that	2	77,28				to claim that	2	55,74	people who are	2	31,42
			they cannot	2	77,28				while I do feel	2	55,74	has the same	2	31,42
			an who is	2	77,28				who is a	2	55,74	according to the	2	31,42
									who is the	2	55,74	have a place in	2	31,42
									it is not a	2	55,74	have to get permission	2	31,42
									it is not	2	55,74	as they grow early	2	31,42
												he argues that	2	31,42
												he thinks that	2	31,42
												I do not agree	2	31,42
												and it is	2	31,42
												a solution to the	2	31,42
												a solution to this	2	31,42
												do not agree with	2	31,42
												do not have	2	31,42
												on this subject	2	31,42
												one of the	2	31,42
												instead of exploiting	2	31,42
												percent of the	2	31,42
												done was that	2	31,42
												who is an	2	31,42
												it is absolutely	2	31,42
												it is a	2	31,42
												it is an	2	31,42
												should be at	2	31,42
												so many people	2	31,42
												spend more time	2	31,42
												take care of	2	31,42
												it will not be	2	31,42

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSSs	Raw Freq	Normed Freq	3-4 FSSs	Raw Freq	Normed Freq	3-4 FSSs	Raw Freq	Normed Freq	3-4 FSSs	Raw Freq	Normed Freq	3-4 FSSs	Raw Freq	Normed Freq
												are a profit for	2	31,42
												that are so important	2	31,42
												you may not	2	31,42
												less salary than	2	31,42
												are not high	2	31,42
												even if they	2	31,42
												there are very serious	2	31,42
												there is a	2	31,42
												they are not so	2	31,42
												they cannot	2	31,42
												way to solve	2	31,42
												are not the	2	31,42
												while I do	2	31,42
												will be solved	2	31,42
												a lot of	2	31,42

Table 63: Frequent Formulaic Sequences Produced by Arda across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
the ones that	5	216,83	it is the	3	128,64	it is not	3	102,74	and it is	4	108,70	should be encouraged	4	100,35
because of the	3	130,10	it is a	3	128,64	won't be	3	102,74	since it is	3	81,52	the reason for	3	75,26
and the ones	3	130,10	would be the	3	128,64	if we are talking	2	68,49	at the same	3	81,52	it is not	3	75,26
the fact that	2	86,73	it is an	2	85,76	but it is	2	68,49	people to buy	3	81,52	as much as	2	50,18
do not like	2	86,73	but it is	2	85,76	because of the	2	68,49	in my opinion	3	81,52	for a long time	2	50,18
it helps us	2	86,73	be the better	2	85,76	different kind of	2	68,49	as soon as possible	2	54,35	because she is	2	50,18
for all of	2	86,73	for the world	2	85,76	for kids to	2	68,49	be able to	2	54,35	and there is	2	50,18
are the ones that	2	86,73	are three reasons why	2	85,76	and it is	2	68,49	and it is the	2	54,35	it is because	2	50,18
must be done	2	86,73	thing to do	2	85,76	is because there	2	68,49	also states that	2	54,35	we can see	2	50,18
most of the	2	86,73	to the life	2	85,76	is not important	2	68,49	not to be	2	54,35	less than men	2	50,18
the ones who	2	86,73	that it is	2	85,76	it is all	2	68,49	he is right about	2	54,35	one of the	2	50,18
there are also the	2	86,73	start to get	2	85,76	it is never	2	68,49	I do not	2	54,35	he is right about	2	50,18
there are some	2	86,73	the ones that	2	85,76	let go their	2	68,49	in the future	2	54,35	she did not	2	50,18
cannot be	2	86,73	in my opinion	2	85,76	since we are all	2	68,49	in this situation	2	54,35	in other words	2	50,18
a search for	2	86,73	grow up and	2	85,76	group of people	2	68,49	is not a	2	54,35	that we are	2	50,18
			the ones that are	2	85,76	are nothing but	2	68,49	is nothing but	2	54,35	is not because	2	50,18
			to be happy	2	85,76	the most important	2	68,49	cannot say	2	54,35	there are also	2	50,18
			in the world	2	85,76	the ones who	2	68,49	and there is	2	54,35	they are more	2	50,18
			we can prevent	2	85,76	I was a	2	68,49	more expensive than	2	54,35	they do not	2	50,18
			why it is	2	85,76	we are talking about	2	68,49	it is because	2	54,35	whether or not	2	50,18
			there are three reasons	2	85,76	there will be more	2	68,49	of a product	2	54,35	as a servant	2	50,18
			there is no	2	85,76	to understand and	2	68,49	of course we	2	54,35			
			is a must	2	85,76	and all of	2	68,49	it is the	2	54,35			
			as soon as possible	2	85,76				do not think	2	54,35			
			and it is	2	85,76				said that the	2	54,35			
									sample also states	2	54,35			
									that it was actually	2	54,35			
									the fact that	2	54,35			
									the ones that	2	54,35			

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
									there are also	2	54,35			
									they want to	2	54,35			
									to be fooled by	2	54,35			
									we are only	2	54,35			
									we cannot	2	54,35			
									we do not	2	54,35			
									when we are	2	54,35			
									it would be	2	54,35			
									a lot of	2	54,35			

Table 64: Frequent Formulaic Sequences Produced by Lale across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
one of the	3	149,40	as a result	4	188,50	as long as	3	90,58	the end of	6	144,37	cannot be	4	103,95
given by the	3	149,40	more than one	3	141,38	in contact with	2	60,39	I do not	6	144,37	in the world	4	103,95
is not given	2	99,60	have to have more	2	94,25	in order to	2	60,39	can lead to	4	96,25	one of the	4	103,95
cannot achieve success	2	99,60	because they have	2	94,25	the strengthening of	2	60,39	are extremely useful	4	96,25	because of the	3	77,96
cannot distinguish	2	99,60	have to think about	2	94,25	in this way	2	60,39	of the human	4	96,25	the number of	3	77,96
in the house	2	99,60	to have more than	2	94,25	spend time in	2	60,39	an end to	4	96,25	the original ones	3	77,96
the subjects that	2	99,60	attention to the	2	94,25	to live together	2	60,39	can be the end	4	96,25	we try to	3	77,96
not understand the subject	2	99,60	on the contrary	2	94,25	such as +noun	2	60,39	the development of a	4	96,25	one of the biggest	2	51,98
of our lives	2	99,60	such as +noun	2	94,25	that there are	2	60,39	but the development of	4	96,25	in that point	2	51,98
prepared for the	2	99,60	very important for	2	94,25	live together in	2	60,39	of a full-scale	4	96,25	in this way	2	51,98
reinforce the subject	2	99,60	will be better	2	94,25	the opportunity to	2	60,39	be the end of	4	96,25	is the easiest	2	51,98
to be successful	2	99,60	the health of	2	94,25	to be able to	2	60,39	to the human	3	72,18	do not have any	2	51,98
and cannot	2	99,60	have more than one	2	94,25	we must ensure that	2	60,39	in contrast to	3	72,18	argued that they	2	51,98
						up in the	2	60,39	bring an end to	3	72,18	are the only alternative	2	51,98
						we must free	2	60,39	is actually in contrast	3	72,18	it can be	2	51,98
						we should not	2	60,39	contrast to an increase	3	72,18	because it is	2	51,98
						grow up in	2	60,39	contribute to people's	3	72,18	it is well known	2	51,98
									an increase of the	3	72,18	it may even	2	51,98
									it leads to the	3	72,18	longer than the	2	51,98
									do not agree with	3	72,18	according to the	2	51,98
									does not bring an	3	72,18	the problem of	2	51,98
									people think like a	3	72,18	make a career	2	51,98

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
									to people's thoughts is	3	72,18	should not be	2	51,98
									we have seen that	3	72,18	they produce more	2	51,98
									to an increase	3	72,18	they should not	2	51,98
									to the development	3	72,18	except for the	2	51,98
									why people think like	3	72,18	that even though	2	51,98
									an end to the	3	72,18	that there is a	2	51,98
									to contribute to	3	72,18	that we can	2	51,98
									I do not agree	3	72,18	the end of the	2	51,98
									in a similar way	2	48,12	by the way	2	51,98
									I do not support	2	48,12	try to keep	2	51,98
									in different areas of	2	48,12	way to produce	2	51,98
									cannot show	2	48,12	we can easily	2	51,98
									will also create a	2	48,12	not be in	2	51,98
									and even in	2	48,12	what kind of	2	51,98
									a similar way to	2	48,12			
									is the ability of	2	48,12			
									it is always	2	48,12			
									it is a	2	48,12			
									according to his	2	48,12			
									look at your	2	48,12			
									may be right	2	48,12			
									not support this idea	2	48,12			
									of all people and	2	48,12			
									a number of	2	48,12			
									of people such as	2	48,12			
									a problem of	2	48,12			
									that knowledge is	2	48,12			
									as a result	2	48,12			
									the ability of a	2	48,12			
									have no effect on	2	48,12			
									as a result of	2	48,12			
									will facilitate the	2	48,12			
									I disagree with	2	48,12			

Table 65: Frequent Formulaic Sequences Produced by Betül across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
in terms of	2	80,65	in terms of	4	151,52	in terms of	3	90,31	in terms of	5	125,06	in the world	17	279,88
it seems to me	2	80,65	has gone a	2	75,76	which is a	3	90,31	while a person	4	100,05	in terms of	7	115,25
seems to me that	2	80,65	in the matter of	2	75,76	for the sake of	2	60,20	people who live in	3	75,04	should not be	5	82,32
students who are	2	80,65	it appears that	2	75,76	as well as	2	60,20	a person is	3	75,04	I think that	5	82,32
success for students of	2	80,65	is connected with	2	75,76	it is an	2	60,20	not to mention	3	75,04	cannot be	5	82,32
who study by	2	80,65	it can be said	2	75,76	it is considered	2	60,20	it should be	3	75,04	the rate of	4	65,85
on technology in	2	80,65	can be said that	2	75,76	it is possible	2	60,20	firmly believe that	2	50,03	which has been	3	49,39
in the areas of	2	80,65	thanks to +noun	2	75,76	serve as a	2	60,20	have several issues	2	50,03	in same positions	3	49,39
			the data of	2	75,76	being in the	2	60,20	he claims that	2	50,03	does not cause	3	49,39
			one of the	2	75,76	so they may	2	60,20	he said that	2	50,03	cannot be ignored	3	49,39
			to grow their	2	75,76	to see that	2	60,20	according to the	2	50,03	the fact that	3	49,39
			where have the	2	75,76	to supply a	2	60,20	I mostly object	2	50,03	they want to	3	49,39
			point of view	2	75,76	when it is	2	60,20	I recognize him that	2	50,03	who work in	3	49,39
			cannot be	2	75,76	with each other	2	60,20	I think that	2	50,03	does not have	2	32,93
			according to the	2	75,76	in the same	2	60,20	if a person	2	50,03	even if the	2	32,93
						fact that the	2	60,20	in my opinion	2	50,03	I agree with him	2	32,93
						according to a	2	60,20	the cheapest thing	2	50,03	by the way	2	32,93
									is a tool to	2	50,03	in the same	2	32,93
									is clear that the	2	50,03	he ignores the fact	2	32,93
									I agree with	2	50,03	one of the	2	32,93
									and they will	2	50,03	rather than a	2	32,93
									recognize him that there	2	50,03	in addition to	2	32,93
									she should not	2	50,03	is being produced to	2	32,93

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
									should not be	2	50,03	is not the	2	32,93
									on the department of	2	50,03	is undeniable fact that	2	32,93
									that even if	2	50,03	it does not	2	32,93
									one of the	2	50,03	as being a	2	32,93
									I firmly believe that	2	50,03	according to the	2	32,93
									that there is a	2	50,03	it has right	2	32,93
									there is a point	2	50,03	it is quite	2	32,93
									undeniable fact that	2	50,03	it is undeniable fact	2	32,93
									which is used in	2	50,03	change the reality that	2	32,93
									I have several	2	50,03	when I consider	2	32,93
									who is an	2	50,03	did not have	2	32,93
									who is known	2	50,03	on the department of	2	32,93
									a person should be	2	50,03	on the journal of	2	32,93
									it is clear that	2	50,03	produced to increase the	2	32,93
												as well as	2	32,93
												the cost of	2	32,93
												the number of	2	32,93
												the only reason of	2	32,93
												the reason of	2	32,93
												getting a high	2	32,93
												do not want to	2	32,93
												be an obstacle	2	32,93

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
												to be a	2	32,93
												to increase the number	2	32,93
												has been performed	2	32,93
												who is a	2	32,93
												whose name is	2	32,93
												have an apprehension	2	32,93
												he claimed that	2	32,93

Table 66: Frequent Formulaic Sequences Produced by Ali across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
it is easy to	3	109,73	are the best	5	160,67	it should be	4	95,10	in the field of	4	91,58	in the world	9	168,79
are the most	3	109,73	is an undeniable fact	3	96,40	to create a	4	95,10	one of the	4	91,58	one of the	5	93,77
in order to	2	73,15	as they are	3	96,40	one of the	4	95,10	the most important	3	68,68	of the world	4	75,02
is easy to say	2	73,15	it should be	3	96,40	the most substantial	4	95,10	can be seen	3	68,68	in his book that	4	75,02
do my review	2	73,15	they can be	3	96,40	make people live in	3	71,33	it is not	3	68,68	in order to	4	75,02
get used to	2	73,15	according to this	3	96,40	is one of the	3	71,33	it is a	3	68,68	with the help of	3	56,26
claim that they	2	73,15	fond of to their	2	64,27	one of the most	3	71,33	it is the	3	68,68	can be a	2	37,51
how to use	2	73,15	be taken into consideration	2	64,27	in order to	3	71,33	would not be a	3	68,68	her study that	2	37,51
on the internet	2	73,15	an undeniable fact that	2	64,27	who live in	3	71,33	can eliminate that	2	45,79	cannot afford to	2	37,51
should be noted that	2	73,15	of this series	2	64,27	do not have	3	71,33	according to a	2	45,79	and they are	2	37,51
that they are	2	73,15	pay attention to	2	64,27	to make people	3	71,33	a person in	2	45,79	according to her	2	37,51
I strongly believe that	2	73,15	taken into consideration that	2	64,27	and it makes them	2	47,55	chance to save	2	45,79	cannot be a	2	37,51
as it is	2	73,15	have a chance to	2	64,27	have a problem of	2	47,55	do feel that	2	45,79	it is a	2	37,51
			have you ever	2	64,27	and make them	2	47,55	do not have	2	45,79	I strongly believe that	2	37,51
			I strongly believe that	2	64,27	according to this	2	47,55	fails to mention that	2	45,79	much more money	2	37,51
			according to a	2	64,27	how is it	2	47,55	at first appearance	2	45,79	I think that	2	37,51
			is necessary for	2	64,27	I strongly believe that	2	47,55	fill up one's	2	45,79	are equal in	2	37,51
			it is a	2	64,27	as a result of	2	47,55	be just a	2	45,79	a solution to	2	37,51
			it is an	2	64,27	a place where	2	47,55	an undeniable fact that	2	45,79	not only for	2	37,51
			can be a	2	64,27	a problem of	2	47,55	gives a chance to	2	45,79	on the same	2	37,51

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
			are the most	2	64,27	a big role in	2	47,55	going to be just	2	45,79	in a long term	2	37,51
			achieve their dreams	2	64,27	is an undeniable fact	2	47,55	is not a	2	45,79	parts of the	2	37,51
			it is an undeniable	2	64,27	is and how	2	47,55	is the most important	2	45,79	people who live in	2	37,51
			should be taken into	2	64,27	an undeniable fact that	2	47,55	it can be	2	45,79	point of view	2	37,51
			can be their	2	64,27	the most important	2	47,55	it helps to	2	45,79	possible effects of	2	37,51
			it is easy to	2	64,27	they do not	2	47,55	it is the most	2	45,79	rise in the	2	37,51
			there are some	2	64,27	of the most substantial	2	47,55	it may have some	2	45,79	based on my	2	37,51
			it is essential that	2	64,27	is the biggest	2	47,55	it should be taken	2	45,79	production of the	2	37,51
			as a result of	2	64,27	is the most	2	47,55	kinds of stuff that	2	45,79	same conditions in their	2	37,51
			although many people	2	64,27	create a place	2	47,55	not have any	2	45,79	in the same	2	37,51
						and can easily	2	47,55	has failed to	2	45,79	in the whole	2	37,51
						people cannot	2	47,55	have been made	2	45,79	be a solution	2	37,51
						plays a big role	2	47,55	of the most	2	45,79	that there are	2	37,51
						should be noted that	2	47,55	on the other hand	2	45,79	the cost of	2	37,51
						should be taken into	2	47,55	one of the most	2	45,79	the best way	2	37,51
						spend time in	2	47,55	point of view	2	45,79	there is a	2	37,51
						taken into consideration that	2	47,55	he fails to mention	2	45,79	there is no	2	37,51
						it is possible	2	47,55	and if you can	2	45,79	these kind of	2	37,51
						they do not have	2	47,55	but if one	2	45,79	there is no discrimination	2	37,51
						they have an	2	47,55	I do feel that	2	45,79	is not a	2	37,51
						to learn the	2	47,55	but it should	2	45,79	because of their	2	37,51

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
						to live in	2	47,55	taken into consideration that	2	45,79	a study from	2	37,51
						has a vital importance	2	47,55	I have several	2	45,79			
						a place that	2	47,55	the strongest claims	2	45,79			
									the study shows	2	45,79			
									the vast majority of	2	45,79			
									these thoughts of	2	45,79			
									to keep a person	2	45,79			
									to manage money	2	45,79			
									if it is	2	45,79			
									that is going to	2	45,79			
									vital importance in	2	45,79			
									while I do feel	2	45,79			
									who is a	2	45,79			
									who is an	2	45,79			
									if you can eliminate	2	45,79			
									a study from the	2	45,79			
									a chance to	2	45,79			

Table 67: Frequent Formulaic Sequences Produced by Ash across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
day by day	3	140,71	on the other hand	4	157,23	day by day	2	60,79	day by day	6	75,72	one of the	8	165,98
a result of this	3	140,71	day by day	2	78,62	even if they	2	60,79	in the future	6	75,72	in the world	5	103,73
the reason why	3	140,71	because of this	2	78,62	of nature for	2	60,79	I do not	6	75,72	this kind of	4	82,99
as a result of	3	140,71	one of the	2	78,62	on the contrary	2	60,79	people avoid to	6	75,72	in terms of	3	62,24
did not finish her	2	93,81	people cannot	2	78,62	on the other hand	2	60,79	is the most	6	75,72	the other biggest claims	3	62,24
had a daily	2	93,81	people have a	2	78,62	they do not	2	60,79	people do not	6	75,72	be a solution to	3	62,24
begin to use	2	93,81	level of their	2	78,62	who live in	2	60,79	in the same	6	75,72	I do not agree	2	41,49
and we begin to	2	93,81	reason is about	2	78,62	in ministry of	2	60,79	one of the	6	75,72	I have found that	2	41,49
we start to use	2	93,81	rising of personal	2	78,62	far away from	2	60,79	he said because of	4	50,48	can overcome all	2	41,49
is one of the	2	93,81	lots of children	2	78,62	because of this	2	60,79	he said that	4	50,48	cannot overcome	2	41,49
thanks to daily	2	93,81	most of people	2	78,62	as in the	2	60,79	a kind of	4	50,48	in some part of	2	41,49
of the most	2	93,81	stay with their	2	78,62				the reason why	4	50,48	claims of the	2	41,49
we can talk	2	93,81	think that grandparents	2	78,62				we are against	4	50,48	in the developing world	2	41,49
one of the most	2	93,81	who name is	2	78,62				in addition to	4	50,48	as a result	2	41,49
			what do you think	2	78,62				part of turkey	4	50,48	around the world	2	41,49
			cannot afford	2	78,62				cannot find a	4	50,48	the cause of	2	41,49
			both of them	2	78,62				whole of this	4	50,48	in the future	2	41,49
									are against each other	4	50,48	the most known	2	41,49
									people think that	4	50,48	is not a	2	41,49
									take care of	4	50,48	to the world's	2	41,49
									strongly believe that	4	50,48	it does not	2	41,49
									the end of	4	50,48	is the most	2	41,49
									it is obvious that	4	50,48	report concluded that	2	41,49

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
									because according to	4	50,48	who are the	2	41,49
									because of the	4	50,48	said that women	2	41,49
									on the contrary	4	50,48	she is the	2	41,49
									the number of	4	50,48	not want to	2	41,49
									start to give	4	50,48	because instead of	2	41,49
									be stronger than	4	50,48	of the report	2	41,49
									they do not	4	50,48	and they can	2	41,49
									when I look	4	50,48	the report is that	2	41,49
									he said because	4	50,48	of the world	2	41,49
									the rate of	4	50,48	who is a	2	41,49
									most people still	4	50,48	will be a	2	41,49
									in terms of	4	50,48	will be more	2	41,49
									according to his	4	50,48	will emerge in	2	41,49
									which is a	4	50,48	he does not	2	41,49
									will be the	4	50,48	on the contrary	2	41,49
									according to my	4	50,48	he supports these	2	41,49
									and he is aware	2	25,24	cannot be	2	41,49
									according to interview	2	25,24	are completely safe	2	41,49
									a research about which	2	25,24	a solution to the	2	41,49
									and I think he	2	25,24			
									and continue to	2	25,24			
									and for a while	2	25,24			
									and he adds	2	25,24			

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
									and he answered that	2	25,24			
									and he can understand	2	25,24			
									and he is	2	25,24			
									at the beginning	2	25,24			
									and he said that	2	25,24			
									and he warned	2	25,24			
									and I found	2	25,24			
									and I look a	2	25,24			
									and I suggest you	2	25,24			
									and I think	2	25,24			
									be the winner	2	25,24			
									and if it is	2	25,24			
									and if people do	2	25,24			
									and it is obvious	2	25,24			
									and it was	2	25,24			
									and make an agreement	2	25,24			
									and people became	2	25,24			
									again we are	2	25,24			
									a result of this	2	25,24			
									and start to	2	25,24			
									and people make a	2	25,24			
									against each other because	2	25,24			

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
									against each other he	2	25,24			
									and the most	2	25,24			
									and the result of	2	25,24			
									a big deal for	2	25,24			
									and they are	2	25,24			
									agree with them	2	25,24			
									and they cannot	2	25,24			
									and they do not	2	25,24			
									and they will not	2	25,24			

Table 68: Frequent Formulaic Sequences Produced by Zeynep across 5 Sub-corpora

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
to do their	6	281,43	they cannot	3	120,10	the rules of	14	481,76	the quality of	7	206,12	in the world	5	113,58
she cannot	4	187,62	they can learn	3	120,10	with each other	3	103,23	in addition to	4	117,79	one of the	4	90,87
his or her	3	140,71	as they want	2	80,06	another fact that	2	68,82	people think that	3	88,34	there is a	3	68,15
more than ever	2	93,81	due to the	2	80,06	are also improve	2	68,82	is not the	2	58,89	the genetics of	3	68,15
fed up this	2	93,81	of age and	2	80,06	and it makes	2	68,82	cannot find anything	2	58,89	because of the	3	68,15
he or she	2	93,81	effects on the	2	80,06	of people living in	2	68,82	content of the	2	58,89	answer to the	3	68,15
because of that	2	93,81	and some other	2	80,06	are the rules	2	68,82	anything to support this	2	58,89	the structure of	3	68,15
to sum up	2	93,81	the benefits of	2	80,06	rules to ensure	2	68,82	many people think	2	58,89	about this issue	3	68,15
not want it	2	93,81	things like that	2	80,06	in the street	2	68,82	determine the quality of	2	58,89	to the world's	3	68,15
they push their	2	93,81	to sum up	2	80,06	and with the	2	68,82	agree with this	2	58,89	are equal in	2	45,43
they are bad in	2	93,81	too bad for	2	80,06	show that the	2	68,82	at first appearance	2	58,89	are harmful to	2	45,43
time to do	2	93,81	and it is	2	80,06	some rules to	2	68,82	about this issue	2	58,89	people think that	2	45,43
to begin with	2	93,81				the ability to	2	68,82	a person cannot	2	58,89	are thought to be	2	45,43
too much and	2	93,81				the source of	2	68,82	think that the	2	58,89	as we know	2	45,43
to do her	2	93,81				to sum up	2	68,82	is sufficient or not	2	58,89	at first appearance	2	45,43
according to +smb	2	93,81				researches also show that	2	68,82	with the development of	2	58,89	is looking hard enough	2	45,43
						with the rules	2	68,82	may seem as a	2	58,89	according to +smb	2	45,43
						with the society	2	68,82	be an alive person	2	58,89	a person cannot	2	45,43
						a rule of	2	68,82	find anything to support	2	58,89	of the biggest problems	2	45,43
									because of the	2	58,89	of the products	2	45,43
									people believe that	2	58,89	on their own	2	45,43
									for me and	2	58,89	is one of the	2	45,43

Appendix 4: (Continue)

Sub-corpus 1			Sub-corpus 2			Sub-corpus 3			Sub-corpus 4			Sub-corpus 5		
3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq	3-4 FSs	Raw Freq	Normed Freq
									should be an	2	58,89	person cannot find	2	45,43
									support this idea	2	58,89	the prestige of	2	45,43
									but if someone is	2	58,89	issue may seem as	2	45,43
									in my opinion	2	58,89	seem as a senseful	2	45,43
									can help us	2	58,89	should not work	2	45,43
									we can provide	2	58,89	solution to this	2	45,43
									there are many	2	58,89	many people believe	2	45,43
									according to their	2	58,89	that they will	2	45,43
												may actually be	2	45,43
												the beauty of	2	45,43
												cannot be	2	45,43
												but if someone is	2	45,43
												the yield of the	2	45,43
												do not take	2	45,43
												too much work	2	45,43
												a person can	2	45,43

Appendix 5: Protocol Questions

1. What are the things that you pay the most attention while writing?
2. What difficulties do you have most while writing?
3. How important do you think “word choice” is while you are writing and how do you choose them?
4. Have you ever heard of “formulaic sequences”? Or collocations?
5. Do you pay attention to use formulaic sequences while writing?
6. What do you think of your essays in terms of usage formulaic sequences?
7. How did you think of using the formulaic sequences samples underlined in your essays?
8. In these essays, were there any FSs you used often since you knew well these sequences?
9. Have you noticed any change or improvement with regards to the use of formulaic sequences in your essays across two semesters? When? How?
10. In English lessons in previous years, did you ever encounter with the formulaic sequences?
11. How much do you think these encounters helped you increase your awareness towards them?



Appendix 6: Protokol Soruları

1. Essay yazarken en çok dikkat ettiğiniz şeyler nelerdir?
2. Essay yazarken en çok hangi zorlukları yaşıyorsunuz?
3. Essay yazarken “kelime seçimi”nin ne kadar önemli olduğunu düşünüyorsunuz ve bunları nasıl seçiyorsunuz?
4. "Söz dizinleri"ni hiç duydunuz mu? Veya eşdizimlerini?
5. Yazarken söz dizinlerini kullanmaya dikkat ediyor musunuz?
6. Söz dizinlerinin kullanımı açısından metinleriniz hakkında ne düşünüyorsunuz ve metinlerinizi nasıl değerlendirirsiniz?
7. Metinlerinizde altı çizili söz dizin örneklerini kullanmak nasıl aklınıza geldi?
8. Bu metinlerinizde, kullandığınız dizinler arasında hakim olduğunuz için diğerlerine göre daha sık kullandıklarınız var mıydı?
9. İki dönem boyunca kendi metinlerinizde söz dizinlerinin kullanımına ilişkin herhangi bir değişim veya gelişme fark ettiniz mi? Ne Zaman? Nasıl?
10. Önceki yıllarda İngilizce derslerinde söz dizinleriyle hiç karşılaştınız mı?
11. Bu karşılaşmaların söz dizinlerine olan farkındalığınızı arttırmada ne kadar yardımcı olduğunu düşünüyorsunuz?

Appendix 7: Sample Teacher Feedback

Essay 1
Participant 1

Have people become overly dependent on technology?

Intro { Exaggeration. It comes to human nature. We exaggerate love, anger, sayings... We exaggerate not only abstract things, substantial things too. Like technology. In fact, everything is essential moderate amount and make our life more facility. For example we use calculators, lifts, GPS's... We do shopping, chatting, etc. all of them makes our life more easy. But we've become lazy and addict.

B1 { Old is dependent on technology. For instance no power no work problem: in our work areas if a machine break down, our works are stop so we should pause it and that's mean is life is stop. Another example about our life's become complicated is in power cut conditions. Our machines, traffics, factories, lights, charge are stop. Especially our phone's charge run out! We are all beat up.

B2 { Our mobile phones, we think they are modern conveniences to make our lives easy, but there is a problem: they aren't our gadgets, we're their slaves. Unfortunately no one can say stop this. Of course everyone aware of this, there are many samples... One of them is a questionnaire in debate.org about technology addiction. The question is "When do you look your mobile phones? After wake up or later? This is astonishing 93% people say that not after wake up after I open my eyes so this is enough for explain. Furthermore, second question is "Do you want to give up your mobile phone?" they had told just seconds and said "No because they are our best friends." Can you imagine? They see them as a friend. A friend that can not feel and breath and haven't a body like yours. For them, it can be a perfect friend. Because their lovely friend do their all tasks. Make them lazy!

B3 { Humankind become fat, thoughtless, and lazy thanks to (!) their friends. There are many examples in the simplest term it calculate it's remember numbers all of them decline or mental exercises and you know our nerves can develop themselves only one condition: when we use them. When we tired them their capacities expand. This is mental laziness in other aspects there is physical laziness. Imagine a building has just two floors and it has an elevator. For an enabled person, it is really beneficial but if you healthy and you have time for use stairs there is no need an elevator. It's derange the health. It break the peace between you and your body.

Con. { To sum up, technology can entertain us, we learn really massive and complicated issues from it. It makes our lives more convenient. But we should be careful about losing ourselves, our control.

word choice

illy Comments

- 1) You need to be careful with word usages
 - 2) Almost no formulaic sequences are seen!
 - 3) Weak organization and arguments!
 - 4) Very short sentences and there are no signal words in between
 - 5) Lots of grammar problems can be seen!
 - 6) Turkish lexicalisation!
- thank you
A

Appendix 7: (Continue)

Essay 2

Participant 2

Many teachers assign homework to students every day. Do you think that daily homework is necessary for students? Use specific reasons and details to support your answer.

Intro
With the advancing technology, does the world and the human order deteriorate? Is technology just about making life easier and improving or does technology has not any harm to humans? In the past, the human order was more focused on production. This led to work both parents and their young children. Thus the outer life of the children started early. The children had more friends and many activities to spend their free time. They were more conscious about environment and nature where they live. Instead of spending time on a television or telephone, they would go to the park and spend time there. But this situation has changed with advancing technology, changing living standardts, and physical and pschological conditions.

B1
The damages of developing technology are definitely the most seen on children. Children who spend most of their time with technological devices become dependent on technology in a short time. Even if these technological devices provide children to have fun and educational time, these devices also affect their development in a bad way due to dependence. While they must learn their personal development and social lives through natural ways, they are deprived of many nature activities due to technology. Buğra Gültekin who study fifth grade student at Sakarya High School has a computer addiction that isolates him from surroundings. This dependence has begun at his young age and it is still continuing. Due to the phone giving his hand at the age of just learning the nature and his environment, he was deprived of all the outside activities and his environment. Thus because he grew without the nature benefits, Buğra is now one of those who could not be found in crowded places and he is a introverted person.

B2
Living standards changing from past to present, changed the way children evaluate leisure time. In the past, while children were running to nature to evaluate their every moment, now the children do not step out of the house. Instead of playing ball with friends in parks, men play online games from the internet. Instead of enjoying the clean and sunny weather, the girls are satisfied with their phones between the four walls of the house. The absence from this natural world not only isolates the person from the outside world, but also prevents him from becoming aware of his or her own self. Child Development and Expert Stefan Gilbert, in his speech at the Science and Health magazine, said "The impact of changing life standards on children is indisputable. The abstraction of children from the outside world has a very negative effect on their development. Families should not raise children with technology."

B3
Child's abstraction from the outside world, due to technology and changing living standards, has a negative impact as physical and psychological. Spending time inside the house in every day, and lack of friendship environment, even if the person does not notice at that moment, drag him or her to depression. The child feels psychologically lacking from his peers because of the lack of this natural life. According to the surveys conducted under the leadership of Istanbul University, 7 out of every 10 children who are deprived of the outside world have a concentration and communication problem. In another study, it is shown that children who were constantly in the home at 22 hours of the day are more prone to obesity and some diseases. Also, it is possible to see many examples of this. Because Hemra Bolut who seen recently on television news, spent 12 hours a day on the phone and the computer, he now suffer from obesity and neural eye disorders. That is why families should not give phone and tablet their children.

Con.
Shortly, considering the current situation of the world, it may be seen normal for children to spend more time at home than the outside. The problem is actually not the advancement of technology, but it is families introduce their children so early with technology. Living standarts changing completely destroyed their old habits and have brought rather different habits and results. Of course, the psychological of being deprived of the outside world should not be ignored. Put it in a nutshell, nature demands its existence to be use.

Comments

- 1) Few formulaic sequences were used!
- 2) Organisation and content is good.
- 3) Use more linking words next time!

Thank you
TA

? punctuation
word choice
grammar

Appendix 7: (Continue)

Essay 3

Participant 3

Some working parents think that childcare centres provide the best care for children who are still too young to go to school. Other working parents think that family members such as grandparents will be better carers for their children. Give your opinion regarding this situation and give reasons for your answer with examples.

Intro

With the developing of living standards, people gets more comfortable lives but while they are getting this, life is also wants something from people. For example, 50 years ago we had a different life style from now. In that time, only the father was working because families didn't have much necessities like today, so mothers were staying at home, and taking care of the children. But now, both of the parent are working, so the families who have children needs someone to take care of children while they are working. Some parents choose their own parents for this sensitive work, but on the other hand some parents thinks that is should be taken care of by child care centres. In my opinion, child care centres are the best option for this job because of these three reason: first one is education, the second is, beneficial activities and the last one is growing in more social environment.

B1

Firstly, while a children is growing up he or she effects by most of the thing which they see or hear. So, while they are growing up the carers should be careful about the environment of the children. Because of that, the persons who works at child care centres are selected and educated about child caring system. Thanks to their education they do not make any fail towards to the children which could harm the developing of children. According to the research, children who is bring up by care centres being more successful in their school years than who is bring up at home by his grandparents. With the help of that research, we can clearly see the benefit of child care centres in education of children.

B2

Secondly, while growing a children it is very important to doing activities which acquires to children the feeling of responsibility, the skill of leadership. They acquires these kind of skills in the child care centres by playing soccer together or making puzzle with their friends. Thomas Burgess, who is a Social Politics and Family Minister of U.S.A said that "The acquire of responsibility feeling is very important, If it wasn't acquired in childhood it is very hard to acquire later. Children can get this feeling by doing activities with their friends." As Thomas Burgess stated in his words it is very important to bring up a child with these skills, and children could easily get these skills in child care centres.

B3

Thirdly, in child care centres the carers sets up the nest social environment for children but some can ask that what is the advantage of this social environment? In this social environment children learns how to communicate with people, so in their next years they will not be shy persons who avoids from talking other people. Also, they gets a point of view owing to this social environment, they approach to the situations narrow minded. According to the research, %42 of children who brings up by his grandparents being shy and they are less talkative than children who brings up by child care centres. These research, also shows that, child care centres makes children more talkative because they experienced many thing than other child's, so they have stories to talk about it.

Conc.

To sum up, child care centres provide the best care for children in many ways. They grows the children expertly owing to their education skills. Also, they makes the children acquired the feeling of responsibility and leadership by the beneficial activities. Lastly, the children cared by caring centre will not be shy and narrow minded, with the help of social environment which was created by child care centres. I want to finish my writing very related proverb with this subject, that is; "Give your job to the expert and don't look back."

- my comments*
- 1) Your word usages good. Underlined sequence are good.
 - 2) Good organization and content
 - 3) there are some grammar problems!

Thank you

Appendix 7: (Continue)

Essay 4

Participant 4

Please discuss "Birth control in Turkey". Agree or Disagree?

Do you know what birth control means? Birth control is the use of a variety of methods, tools, or medicines to temporarily or permanently prevent pregnancy or reduce the possibility of pregnancy. In most countries, there is some precaution about pregnancy. When you ask yourself whether there is any birth control in Turkey, could you say yes or no? Following the searches I have done, I firmly believed that there is a visible birth-control deficiency in Turkey. Educational level, cultural differences, and shyness shape my thoughts about birth control in Turkey.

Intro
good intro

The first important reason cause to think that there is a birth control deficiency in Turkey is educational level. People are classified as a primary school, secondary school, high school, and university graduate according to their educational level. According to "Condition of Women in Turkey" report by Directorate General on the Status of Women Last 10 years, among the women, the percentage of literacy increased to 80,4 % from %76,9. Yet still, one out of every five women, indeed 5 billion 732 thousand women, is not illiterate. To connect the education level and birth-control, lack of education causes lack of consciousness-raising; therefore, the number of the child mother is more than 20 thousand a year. If the education level was high, this number would be less than 10 thousand. Also, Megan Elliot who is from Marie Stopes International Charitable Foundation says "Every year 21 billion women have an unhealthy abortion, and 70% of them are younger than 20 years old." To sum up, in Turkey, if we want birth-control, the educational system should be improved.

✓ B1

good example

In addition to educational level, cultural differences have a bad effect on birth control in Turkey. East and West regions' cultures are so different that it also affects the number of people in the family. According to data from The Turkish Statistical Institute, in 2017, the fertility rate in Şanlıurfa was 4,29; in Edirne was 1,46. As you can see, cultural differences influence the birth control. For instance, Zelaya who lives in Şanlıurfa has 7 children because of the cultural pressure and in addition to 7 children her mother-in-law always wants one more child especially boy. If her bride has fewer children than the other brides in the village, her mother-in-law will feel twist slowly in the wind. On the contrary to Zelaya, Işık who lives in Edirne has only 2 children. Her mother-in-law supports Işık and generally says "You have 2 children, you have 2 important responsibilities, and you have so pretty little family." Işık's mother-in-law thinks about grandchildren's education and growth. In brief, in Turkey, there is a problem about birth-control due to cultural differences.

✓ B2

good example

The last reason is shyness. In my opinion, people want to hide not only pregnancy but also some important diseases. Is it a true solution? Why they study for 6 or 7 years to become a doctor? If everybody hides their health issues, what will a doctor do at the hospital? Also, do not forget "There is no shame in medicine." In Trabzon, pregnant woman tend to go "woman" gynaecologist. Thus, in our country, first, we need to overcome feelings such as fear and shyness. Then birth control will be provided but if everything continues with the same way, birth control is impossible. To sum up, shyness affects my thoughts about birth control in Turkey.

✓ B3

Briefly, birth control is to temporarily or permanently prevent pregnancy or reduce the possibility of pregnancy. Most countries have some precautions on birth control. To focus on Turkey, I strongly believed that educational level, cultural differences, and shyness lead to lack of birth control. In the introduction paragraph, I asked "When you ask yourself whether there is any birth control in Turkey, could you say yes or no?" Again I ask, yes or no?

✓ Conc.

my comments

- 1) You have used several word combinations underlined, Good.
- 2) Good organization and content!
- 3) Several examples you have given, Good.
- 4) few grammar problems - Good

thanks
A

Appendix 7: (Continue)

Essay 5
Participant 5

Because of the busy pace of modern life, many children spend most of their time indoors and have little exposure to the natural world. How important is it for children to learn to understand and appreciate nature? Give reasons for your answer and include any relevant examples from your own knowledge or experience.

✓
h+co
Have you ever heard about the term cosmopolitan? According to Oxford Learner's Dictionary, it means "containing people of different types or from countries, and influenced by their culture". Nowadays, there are so many countries that have cosmopolitan cities such as Paris, Los Angeles, London, New York, Amsterdam et cetera. Although it is seen as a positive situation by most people because of its advantages like learning about other cultures and having wider perspectives, it is also hard for governments to keep a peaceful environment between that many people from that many different cultures. Thus, for everyone to live in peace and harmony, the government of a cosmopolitan city must follow three certain rules: keeping religion away from itself, giving education about varied cultures in schools, having a solid constitutional law.

good points

✓
B1
First of all, the government must keep religion away from itself. Religion is something that every individual chooses for themselves, therefore; it is not appropriate for a government to have a specific religion. Also, it is a fact that there is chaos in countries and cities whose governments picked a religion. For example, in 2016, The New York Times asked women from Saudi Arabia, which has a cosmopolitan city named Jeddah, how their lives were, and the answers shocked the world. A woman named Juju (21) said that it was like as if she was in hell and that she had no freedom. Another woman, Rulaa (19), explained how she almost died in an accident because the ambulance did not take her to the hospital without a male guardian. Other one, Al Qahtaniya (28), stated that her husband, father or brother were allowed to beat her up, lock her in the house for months if they wanted to, decide her clothes, friends, travel activities for her et cetera. However, as Ghadah (27) said, if they went to police for protection, they would go to prison as if they were criminals. Unfortunately, the state of Saudi Arabia has not changed since these interviews. In brief, it is clear that it is highly unlikely to have peace, freedom and equation in a country or a city that has an official religion.

good et.

good.

✓
B2
Secondly, giving education about varied cultures in schools is as important as dividing religion and the government. In a cosmopolitan city, it would be harmful to focus on just one culture because obviously, to require a healthy atmosphere, equality is needed. Moreover, inequality would make people who are from other cultures feel like outsiders, which would be a reason of deterioration of peace. For instance, Adanna Akams, an African woman who writes for a blog called A Medium Corporation, shared her experiences in one of the most multinational countries in the world, America, in 2017. She explained that she has been bullied for her cultural dissimilarity like any other black people. She said she was being made fun of, abused for her hairstyles such as dreadlocks and cornrows, which were a part of black culture. Moreover, because the US government only focuses on white culture, she was not allowed in the class with her traditional clothes. Ms. Akams also stated that even though it has been an issue for decades, white people are still uneducated about racism, cultural appropriation and black history because of the lack of information given in schools. As it is as plain as a pikestaff, if the government of a cosmopolitan city does not care to give information on every culture, it is impossible for people to live in an order.

✓
B3
The last but not the least, the government must have a solid constitutional law. It is known by everyone that the substances of laws that a country has are the most important things for that country to be considered livable. When it comes to a country or a city with diversity of ethos, one of the best options is to have rules that can be accepted by every kind of people. For example, Naime Öztekin, an infamous writer who experienced Mustafa Kemal Atatürk's revolutions, the founder and the first president of

My Comments

- 1) Several formulaic sequences you have used. Good.
- 2) Good content and organization.
- 3) Good examples ↓ ↓ but you can give just one example in each paragraph.

Appendix 7: (Continue)

Republic of Turkey, explained that removal of the injustice of the system changed her life for better. Because she was born and raised in another country, she was unfamiliar with the culture of people in Turkey, which was heavily influenced by their religion. Moreover, Mrs. Öztekin said that she felt anxious about the former regulations in the country, and that she should not have had to follow someone else's religion's or ethos's rules. In short, as Mrs. Öztekin's experience shows, the government should never build their laws under the impact of one culture.

✓
conc.

All in all, for a society that works in peace and harmony, the government of a cosmopolitan city should consist of the things that are given in the paragraphs above. It is well known that if there is any injustice to some of the people in a country or a city, disorder and chaos are inevitable. For that reason, the government should fulfill its responsibilities for the people and create a healthy environment. As Morgan Scott Peck said: "The key to community is acceptance, in fact the celebration of our individual and cultural differences. It is also the key to world peace."

✓
My Comments

4) this is a good essay but try to extend the scope next time!

thank you
✱

Appendix 7: (Continue)

Essay 6

Participant 6

How can people live together in peace and harmony?

Living in a metropolitan city without any problems? It does not sound persuasive. From the industrial revolution, cities become growing day by day, and people leave their towns and start move on metropolitan cities. Because of this moving, cities have twice as many problems as they have before. In that time, governments should stand strong if they do not, it will be chaos. So, there is no doubt that governments should take some precaution for living peace and harmony in metropolitan cities. There is a lot of issues that governments should make some differences or renew but traffic, education and healthcare centers are more prior than others.

Firstly, traffic is one of the most essential parts of each metropolitan city. If big cities have not an improved traffic system, governments will fail. The reason why traffic so important is about people's new habit. In today's world, every person has own car and they usually prefer to use a private car for their daily transportation. Because of this prefers, traffic problems come out. So, governments should take this issue seriously. For example, Mevlüt UYSAL who is Ministry of İstanbul Metropolitan Municipality (İBB) make an agreement which is about using public transportations with private companies and government offices. According to the agreement, İBB gives transportations cards which are twice as cheap as other cards to workers and thanks to this cards workers can take advantage of all of transportation vehicles such as metro, bus, tramway, marmaray, and as a consequence of this, people who use their car just for daily work transportation give up using own car. After six months, Mevlüt UYSAL announced that intensity on highway %13 decrease and on the other hand using public transportations %18 increase. So, that's show İBB's cards attack pointed. And if governments take a logical decision like this, traffic can be calmer.

Another significant reason is about education. People do not just live themselves also nearly every family has children and these children must go to school when they are 6 years old. As time goes by, the number of children increase, and governments should make something to provide welfare education system for school-age children. Because generally in the big cities number of students, in a just one class, is too much. In this problem having foresight is quite important. For instance, Mustafa Sabri HOŞOĞLU who is principal of education in Samsun made a population research for learning number of children who are going to or will go to school in Atakum. According to his research school-age children's numbers twice as much as in the past. It shows that schools which are located in Atakum will not enough for the next generation. When he saw this result, he began to meet with The Ministry of Education and he wanted to money supply for building new schools. Then 2 years later 3 schools which primary, middle school and high school were constructed. Thanks to Mustafa Sabri's insight into lack of school issue, Atakum does not fail about welfare education opportunities and also classes do not be crowded. So, every government takes a precaution like this about education.

Health care centers are last prior subject among others. From the child to the elderly, every human has some health problems. These problems solutions are in the hospitals but health care centers usually so crowded and complicated. That's mean health care centers, especially in the metropolitan cities, do not enough support to patients wishes. So, in this case, governments make something. According to Doctor Sevim YILMAZ, works in Ministry of Health, observations; because of high demand, patients spent so much time in the hospital and this situation directly to people become to be nervous and angry. And also, people behaviour harshly and even if they can fight with doctors. Then, she wrote to Ministry and offer to them built new health care centers every street in the whole of Turkey to decrease intense of patients in hospitals. After her suggestion, Ministry applied this suggest in some part of Turkey, Ankara, İzmir and Adana, and it worked. That show common of count health care centers is also essential for harmony in big cities.

Finally, what makes government elevated in terms of the public is that it always works and take decision for the public level of life with harmony and peace. And if the governments pay attention to

Appendix 7: (Continue)

suggestions, such as about education, traffic, and hospital, the cities will be more peaceful and harmonic area for living.

My comments

- 1) Several good combinations/formulaic sequences you have used! Good
- 2) Good organization & content
- 3) Your sentences are rather short!
- 4) Try to extend the scope of examples next time.
- 5) Good grammar/word usages/essay content
- 6) Specifications of the examples are extra good.

Thank you
A

Appendix 7: (Continue)

Essay 7

Participant 7

It is generally seen that the latest developments in the field of artificial intelligence is likely to have a positive impact for our lives in years to come. Other people, on the other hand, are worried that they are we are not ready for a world, where computers will be more clever and dominant than us. Give your opinion about this topic by making your position clear to us. Don't forget to mention "opposing ideas" and then try to refute this by your own ideas.

Intro ✓
Mark Zuckerberg, who is an American technology entrepreneur and philanthropist and founder of Facebook, is leading the improvement of artificial intelligence and maintaining that it is a light for future. Zuckerberg expresses his thoughts about artificial intelligence on his Facebook wall that from diagnosing illnesses to keep a person healthy, to improving self-driving cars to keep a person in safe, and from showing better content in the news feed to delivering more relevant research consequences it will help to develop systems across so many different fields. At first appearance, his thoughts about artificial intelligence may seem as a sensible one, but if someone digs down deeper, a person will fail to find any sensible fact. Based on my research and my own personal experience, Zuckerberg is wrong because artificial intelligence may seem more harmless than it is. It is highly apparent that Zuckerberg's thesis does not sort together with the real life as artificial intelligence is more trouble than it's worth. ✓

B1 ✓
One of Zuckerberg's strongest supporting claims is how artificial intelligence keeps a person healthy with its upgraded testing ability and ways of treatments. The CEO of Facebook, Zuckerberg, made evident the whole questions via Facebook Live Broadcast. He uttered the health-wise benefits of artificial intelligence by stating that it can diagnose a disease too much earlier than the human brain and treat it. He added that with the development of artificial intelligence, the average lifetime will increase. However, he fails to mention that the ways of diagnosing a disease with artificial intelligence might be unhealthy. For example, a study from Chicago University, conducted by Angelo Olinto who is a Dean of Physical Sciences Division, reveals that treatments being done with robots may damage a person's brain and neural system because of its radioactive rays and radiant density that spread from it. I find this to be quite more persuasive to not premeditate the artificial intelligence and from my standpoint, a human brain can do better of diagnosing diseases and curing them healthfully and harmlessly. It can be seen that it helps for making human life longer, but it is clear that the human body absorbs harmful rays which do not kill today but tomorrow. ✓ good example

B2 ✓
One of the most striking supporting argument that suggested by Mark Zuckerberg is developing self-driving cars. Zuckerberg mentions that self-driving cars will reduce car accidents largely and it makes the daily life easier as it presents a remote control system that gives a chance to do double duty. He also noted on his Facebook Live Broadcast that a clear majority of death causes for people are still car accidents and if you can eliminate that with AL, that is going to be just a dramatic development, Zuckerberg said. One of the top causes of death for people is car accidents still and if you can eliminate that with AI, that is going to be just a dramatic improvement. I have several issues with these thoughts of Zuckerberg: first, self-driving cars are more tend to have an accident as they are all robotic vehicles and do not have a human brain which makes them unable to react on sudden moments. As self-driving cars don't have any reflex, it is more dangerous for other drivers and the traffic. For example, according to a news from Euronews.com that reported by a journalist, Kerem Congar, a self-driving uber car hit a woman and cause her death while doing its test driving in Arizona, U.S.A. Zuckerberg's fault can be seen about this statement that if it substantially keeps a person in safe with its high technology and marvellous intelligence, one can easily see that there would not be a car accident even in test driving. Thus, these data suggest that being in the car and holding the wheel is safer than trusting any human-made brain as it's not impeccable. ✓ good examples

good-
While I do feel that Zuckerberg has failed to offer the strongest claims and benefits of artificial intelligence, it is not a suitable declaration to say artificial intelligence do not have any useful sides

Appendix 7: (Continue)

either. Although many studies and researches have been made about harms of artificial intelligence to society, it may have some hard-hitting advantages such as in the field of health, military and economy. In fact, I would love to say with artificial intelligence, dozens of life can be saved and cures might easily come up with a solution. For example, according to a study from the Department of Engineering Science in Oxford University conducted by Mark Cannon, who is an associate professor, shows that after artificial intelligence coming into force in the field of health, %79 of diseases can be diagnosed in a shorter time and it is a great number for the health sector. On the other hand, Cannon added that artificial intelligence has made a great impact in the field of the military; with the using of artificial intelligence and with the developments of unmanned aerial vehicle and self-driving tanks in the field of military, most of the deaths and physical injuries started to decrease. For this reasons, claiming that artificial intelligence is completely harmful to today's world would not be a good statement.

B3

✓

In this paragraph
you used
visibility
items.

Artificial intelligence has a vital importance in some case, but it should be taken into consideration that one should be aware of its harms based on misappropriation of artificial intelligence and incidents happened because of careless use of artificial intelligence. However, it's an undeniable fact that artificial intelligence, by making a person healthier, reducing deaths because of car accidents with the autonomous car, gives some opportunities. What is artificial intelligence in real? Is it helpful or harmful? If it is not being used in a proper way, it can be called canary in a coal mine.

conc.

✓

My comments

- 1) You have used the underlined formulae sequences.
- 2) Good organization of ideas and content.
- 3) Your examples are good.
- 4) You have used several visibility items ... (what I do ...)
- 5) You can give just one exemplar for each paragraph and extend it!
- 6) This is a good essay.

thank you
/

Appendix 7: (Continue)

Essay 8

Participant 8

It is a fact that many people in our society think that "price" is the most important criteria while they are buying something in the market (cell phone, dress, shoes...) Agree or disagree?

✓ Intro
Warren Buffet who is known as one of the most successful investors for the twentieth century has several principles to spend his money. "Buy the cheapest thing, don't keep close anything which is not necessary for you, and the difference in income of your society does not affect your expenditure approach" are basically three mottos which are claimed by Buffet. According to the first of his main mottos, Buffet identifies that when a person goes shopping or invests, the most logical decision that is preferred the cheapest thing by oneself. Thus, even if some amount of his/her money is expended, the person may have enough fund for potential occasions. However, I firmly believe that the determination of Buffet rules out the magnitude of quality. Unlike Buffet, Joseph Pine who is an experience economy and supports businesses to create modern purchasers argues that money is a tool to buy not only a product but also a class in his book of which is Infinite Possibility. Of course, just as high quality does not mean expensiveness, poor quality is not cheapness. Nevertheless, in my opinion, while a person is buying something, he/she should want to know whether there is rational stability between its price and its quality. I strongly believe that a conscious consumer should pay attention high-class rather than cheapness, so his/her money does not go up in smoke. → very good

✓ B1
Another attitude which is asserted by Buffet is buying solely necessities. In his interesting conviction, Buffet says that a person should avoid from unnecessary shopping. Moreover, he claims that even if a person purchases a thing which is no necessity for him/her no matter how cheap it is, the product which is bought is an exorbitant price in fact for him/her. However, these explanations are not very fascinating for me. In a study conducted by Fatma Oren who is an academician on the department of economics at Çukurova University, it was shown that when several discounts or special offers are made in the retail sector, consumers take into account their forthcoming necessities and do shopping accordingly. In addition, when the corporations offer some special options for their devoted shoppers, it appears that the customers buy extra things which are not needed. It is an undeniable fact that sale has a charming effect for everyone, not to mention women. Certainly, if businesses use as a trump the influence, they may gain profits. Thus, buyers have additional expenses or debts as well. proof

✓ B2
The other distinct point which is emphasized by Buffet is about income groups in the society. As noted by Buffet, people who live in the same society prefer similar choices to do shopping regardless of their income. What's more, he says that there exists an interaction among them in terms of consumption addictions. Furthermore, he claims that the situation relates to the community psychology. However, I have several issues about the one-sided argument, also I think that income level has a role more significant than social psychology as regards consumption expenditures. In contrast to Buffet, the thesis of Umut Yıldız who is a journalist and a grad student on the department of sociology in Istanbul University reports that no essential resemblance in society between a high-income group and a low-income group in terms of their expenditures. In his remarkable analysis of garbages where are in different neighbourhoods in Istanbul, Yıldız identifies that whereas people who live in Etiler, which receives wealthy groups, buy daily and glass bottle milk, people who live in Esenler, which hosts poor groups, purchase long-life and box milk. It is clear that the reason of their different selections is connected with their revenues. Meanwhile, I think that a conscious consumer should cut his/her coat according to his/her cloth regardless of other people. good example

✓ B3
While I mostly object to Buffet's opinions, I recognize him that there is a point I agree with him. According to Buffet, a person should be aware of money's value, also he/she should not pay overmuch for anything. It is undeniable fact that earning money is not an easy task, and it should not be squandered. In addition, Buffet says that if a person spends money for the non-essential thing, he/she can be obliged to sell his/her essential goods for one day. Certainly, a person should not bring about the impasse, while he/she doing shopping. Actually, credit cards are the most appropriate samples for the situation. While a person is buying something, he/she cannot be economical due to necessity (real)

Appendix 7: (Continue)

credit cards. Moreover, he/she can feel free in terms of money and spend extremely. Eventually, he/she can find herself/himself in a debt binge. On account of these, a person should be attentive about shopping, but also he/she should not avoid seeking quality.

Conc. ✓ Consequently, the price should not be the most significant criteria while a person is doing shopping, and he/she should consider not only economical option but also quality standards. Furthermore, the fine line and between the cheapness and the class should be taken into account by consumers. Also, it is logical that buying used regularly goods thanks to several discounts and campaigns, although users do not need them at that moment. In addition, it should be known that an individual's consumption routines become different him/her income level, not according to the social impact. Despite all these, it should be kept in mind that money is a tool to obtain our desires, and it is substantial value for everyone in all the world. While a person who is spending money should consider all of these, but also it should be regarded the unforgettable expression which is used in almost every culture: "I am not rich enough to buy cheap things!"

My Comments

1. You have used many formulaic sequences, good.
- 2) Very good organization and content
- 3) the underlined formulaic sequences are frequent in English
- 4) the examples are good and clear and well-designed.
- 5) the grammar is okay / punctuation is good.
- 6) thank you for this paper which has a ^{high} quality

Thank you
✓

Appendix 7: (Continue)

Essay 9

Participant 9

Although it is fact that the contribution of women is great in various fields such as in education and employment, they are nevertheless treated unequally in the workplace when it comes to pay and promotion. Equality of opportunity should be promoted for men and women in the workplace? Agree or disagree?

✓ Intro
Through long ages, women have been a significant part of the workforce but they have been exposed to discrimination in a lot of ways. Yet, for some people, this situation has been denied or underestimated. In 2015, Danny Masterson interviewed with a member of parliament whose name is Luke Wilson and he spoke about the situation of women in the business sector on BBC News. In his interview Danny Masterson, Luke Wilson said that he does not think that women are treated unequally because he claims that according to a survey in 2015 women employees do not feel that they are treated unequally in terms of promotion or pay. Also, he said that the difference in wages between genders is not a big deal anymore because the difference in wages is gradually decreasing. However, I do not agree with his opinion about because based on my research, recent studies argue against his thoughts. Therefore, even though he tried to create a positive atmosphere about gender equality, in my opinion, he denied the fact that women are still treated inequality in the workplace. ✓ ok

✓ B1
First of all, one of Luke Wilson's strongest supporting claims is that according to his research women do not feel that they are treated unequally and because of that he does not think that there is not an issue about gender inequality. As the interview continued he said that women are not exposed to discrimination in terms of promotion contrariwise women make progress in the profession easier. I do not agree with his statement because multinational professional services network PwC asked female employees whether they have exposed to discrimination in the workplace in terms of promotion or not. According to the research of PwC in 2016, fifty-five of women thinks that they have treated unequally by institutions and employers in terms of career advancement. In the research, more than half of female employees stated that even though they have the same educational level and same qualities for the job as their male colleague, they cannot make progress in their career as their male colleague. Thus, this study demonstrates how wrong his opinion is about women not feeling that treated unequally in terms of promotion. ✓ ok

✓ B2
Secondly, Luke Wilson's other underestimating gender inequality statement was that he does not think that gender inequality is not a big deal anymore since the wage gap between men and women is gradually decreasing. However, I strongly believe that women are still exposed to discrimination when it comes to paying. Moreover, Lauren Holter published a report called "The Gender Wage Gap Is Way Bigger Than You Thought" in 2018. According to her report, a new study by The Institute for Women's Policy Research shows that the difference in wages between women and men is very high. The study indicates that a woman earns 70 cents for every dollar earned by a man. So, it shows that the gender wage gap is not decreasing, on the contrary, the gender wage gap is increasing. Also, he mentioned that according to his research, women do not feel that they treated unequally in terms of pay either. Yet, according to the research called "Gender Equality In Working Life" of PwC in 2016, Fifty-nine percent of women think that the wages between women and men are not equal. Therefore, new studies show his fallacy about this issues. ✓ v. good counter argument

✓ B3
While I mostly do not agree with his ideas, his statement about women does not feel that they treated unequally might be right because according to research agency called Emnid, one-third of female employees do not even know that they being paid less than their male colleague. Also, according to news published by The New York Times in 2017, When it was found that a textile company paid female workers a dollar lower than the wages paid to men, a female employee named Lisa Danielle applied to the court and she had won the case. In this case, it is not a surprise that some women do not feel that they are treated unequally. For this reason, his statement might not be completely wrong. ✓

✓ C
To put it briefly, as in every field, the presence of women in the field of work has always been important. However, women have been faced a lot of challenges in the workplace such as the gender wage gap and getting promoted. Unlike Luke Wilson, I firmly believe that these problems still a big deal for us since recent researches indicate that working women still face these issues in the working ✓

Appendix 7: (Continue)

life. Therefore, in my opinion, to prevent these problems authorized people such as a member of parliament should attach importance to gender inequality in the workplace instead of denying or ignoring these problems since it will have no contribution to solving these problems.

My comments

- 1) This is a very good essay.
- 2) You have used many frequent formulaic sequences.
- 3) Very good grammar and word choice.
- 4) Strong examples you have used.
- 5) B3 shows visibility ✓

Thank you
A

Appendix 7: (Continue)

Essay 10

Participant 10

Genetically modified food

Intro ✓
WFP, the World Food Programme reports that around 800 million people worldwide are starving, while the world population continues to grow. By the year 2050, estimated ten billion people will live on Earth. To provide them, agricultural production would have to double, according to the UN Food and Agriculture Organization. But crises and conflicts, natural disasters, and the consequences of climate change pose major challenges in the fight against hunger. And that is where the genetic engineering company Monsanto comes into play. Group CEO Hugh Grant promises to double crop yields thanks to green genetic engineering and claims that this biotechnology is the only way to satisfy future generations and to control the entire food chain. However, while some see this US group as a rescue of the food crisis, it is predominantly criticized. In a superficial approach, this concept seems fine, but I do not think that the greedy company is really concerned about the influence of genetically modified plants on human health or the food quality of the future generations. Unfortunately, genetically modified food that is generally said to have no side effects, as a matter of fact can cause various harms to human body, nature, and ecological system. There is no point in harming people and the world they live in, in order to feed them. good

B1 ✓
"In order to meet growing demand as well as changing eating habits, farmers will have to produce more food over the next few decades than they did over the past 10,000 years", says the company Monsanto. Therefore, they are working to double the yields of corn, soybeans, cotton and rape until 2030. The fact that genetically modified plants can contribute to world nutrition, and even solve the problems that have been aggravated by climate change, has been Monsanto's main argument for their commitment for many years. However, in order to solve the world's hunger problem, they involuntarily endanger people's health. In experiments on animals, it has been determined that GMF products cause organ destruction, irregularities in digestion and immune system, and infertility. In the United States, health problems have increased after the emergence of GMOs in 1996. The proportion of Americans with chronic diseases increased from 7% to 13% in just 9 years, according to Health Research Funding. Food allergies have suddenly increased, with diseases such as autism, irregularities in reproduction, and obvious increases in other discomforts. Soya, which is more than 90% genetically engineered, carries our young children under great danger. If it is consumed too much it causes hormonal disorders. For example, in the case of male children, feminine features; in girls, negative effects such as early adolescence or sound thickening can be observed, according to the internet newsletter Healthline. So, the damage of human body outweighs in contrast to its nutrition. strong example

B2 ✓
In order to increase yields, it is important to protect the crop from harmful external and environmental factors, and to immunize them against pests. For this, Monsanto has infiltrated the goods with a gene that repels insects. In addition, they claim that the production of salt-, dry- and cold resistant plants is their main objective. According to Monsanto's homepage, already in 1998, they proudly showed plants that can grow with minimal use of water and even from saline soils to a group of German journalists. All this makes it possible for plants and animals to grow up in areas apart from their natural habitat, become resistant to insects, and to obtain more crops to provide larger crowds. But all these measures that sound nice have not only negative effects on human body but also on our ecosystem. GMO crops and their associated insect killers harm insects, amphibians, the marine ecosystems and soil organisms. These detrimental chemicals reduce biocompatibility and pollute water resources. For example, GMO crops have annihilated the life habitat of some butterfly species. The amount of king butterflies is reduced by 50% in the United States, according to report of BBC news. As we have seen, products with GMOs have become a threat not only to human beings but to all beings of nature. In general, it does not matter which different organisms are damaged, because ultimately the human body is affected, since at the end of the chain there we are. Therefore, it is vital not to consume GMO products. v. good

B3 ✓
Nevertheless, the world's hunger problem is a very serious topic and we need several alternatives to solve it. In this regard, Monsanto's objectives and solution proposals are a good way to prevent major crises and create a stable state. With their biotechnologies, they are able to produce much more and to protect these products from damage factors. However, we have to wake up and see the other, dark side of Monsanto and its policy. This company is entering into the agricultural market, especially in the countries of south america, as a competitor against domestic, traditional, natural agriculture. The price of corn produced by chemistry and genetically modified corn is much cheaper than the ecological price of the market, so the natural farmer can not stand this competition. Monsanto is exploiting the world's hunger problem to introduce itself to people as

Appendix 7: (Continue)

a savior. But actually never before has humanity produced more food than today. More than a third of these products is thrown away in our countries. If food were fairly distributed, no one would go hungry. Yet one billion people are starving today, more than ever before on this planet. 70% of all hungry people live in the countryside - especially in Asia and Africa. People's lack there is the access to land, water, seed, and to practical know-how. It is more rational to solve these problems instead of poisoning people with genetically modified food.

Conc. ✓ In summary, there must be found an answer to the world's hunger problem, but it should be not as harmful for people and nature as genetically modified food. The reason for that is the fact that it makes no sense to find a solution which causes bigger problems. We can not put people's health in danger in order to feed them.

My Comments

- 1) This is another very good essay with high frequency formulae sequences
- 2) The arguments are strong and backed by good examples
- 3) The organization, content and other technical issues are well-done
- 4) The conclusion part is not conclusive enough OR it can be revised!

thank you
A

CURRICULUM VITAE

Hanife ÖZTÜRK.. She graduated from Güzelyurt Primary and Secondary School in 2009; Yunus Emre Anatolian High School in 2013; Karadeniz Technical University – Faculty of Letters, Department of English Language and Literature in 2018. She started master’s degree in 2018 at Karadeniz Technical University – Institute of Social Sciences, Department of Western Languages and Literature, the program of Applied Linguistics. She works as a “research assistant” at the Department of English Language and Literature at University of Eurasia.

ÖZTÜRK is single and speaks English.