AN ANALYSIS OF VERB-NOUN COMBINATIONS IN HIGH FREQUENCY VERBS IN ARGUMENTATIVE ESSAYS OF TURKISH ELT STUDENTS: THE CASE OF *MAKE* and *DO*

PhD Dissertation

Mehmet KAHRAMAN

Eskişehir 2021

AN ANALYSIS OF VERB-NOUN COMBINATIONS IN HIGH FREQUENCY VERBS IN ARGUMENTATIVE ESSAYS OF TURKISH ELT STUDENTS: THE CASE OF *MAKE* and *DO*

Mehmet KAHRAMAN

PhD DISSERTATION
Department of Foreign Language Education
Program in English Language Teaching

Supervisor: Assist. Prof. Dr. Gonca SUBAŞI

Eskişehir
Anadolu University
Graduate School of Educational Sciences
April 2021

JÜRİ VE ENSTİTÜ ONAYI

Mehmet KAHRAMAN'nın "An Analysis of Verb+Noun Combinations in High Frequency Verbs in Argumentative Essays of Turkish ELT Students: The Case of Make and Do" başlıklı tezi 19/03/2021 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği"nin ilgili maddeleri uyarınca, Yabancı Diller Eğitimi Anabilim Dalı İngilizce Öğretmenliği Programında, Doktora tezi olarak kabul edilmiştir.

	Ünvanı Adı Soyadı	İmza
Üye (Tez Danışmanı)	: ……..	
Üye	: ……..	
Üye	: ……..	
Üye	: ……..	
Üye	: ……..	

Enstitü Müdürü

ABSTRACT

AN ANALYSIS OF VERB-NOUN COMBINATIONS IN HIGH FREQUENCY VERBS IN ARGUMENTATIVE ESSAYS OF TURKISH ELT STUDENTS: THE CASE OF *MAKE* and *DO*

Mehmet KAHRAMAN

Department of Foreign Language Education
Program in English Language Teaching
Anadolu University, Graduate School of Educational Sciences, April 2021

Supervisor: Assist. Prof. Dr. Gonca SUBAŞI

It is often argued that word combinations, and, thus collocation knowledge is an essential part of the mastery in a foreign language. It is also argued that high-frequency verbs are tricky for the learners, especially when they are in combinations with other words. Comparing to native speakers, some learners tend to use high-frequency verbs more frequently while some others avoid using them considering their collocational properties or delexical meanings. The present study aims to reveal the use of high-frequency verbs *make* and *do* when they occur in a verb+noun combination in the argumentative essays of Turkish learners of English. In this context, this study investigated overuse/underuse, grammatical and semantic patterns and erroneous productions in the learner corpus. The investigation made use of LOCNESS as the reference native corpus for comparison purposes. The findings showed that learners tended to underuse *make*+noun combinations comparing to the native corpus; they also showed some similarities and dissimilarities with the native corpus in terms of grammatical and semantic properties in their productions. Finally, the error analysis revealed that learners have some problems with *make* and *do* verb+noun combinations. Based on the findings of the current study and previous studies, the present study shared theoretical and practical implications particularly for language teaching settings.

Keywords: Word combinations, High-frequency words, Collocations, Corpus

ÖZ

TÜRK İNGİLİZCE ÖĞRETMENLİĞİ ÖĞRENCİLERİNİN TARTIŞMACI YAZILARINDAKİ YÜKSEK FREKANSLI FİİLLERDEKİ FİİL+İSİM KOMBİNASYONLARININ BİR ANALİZİ: *MAKE* ve *DO* ÖRNEĞİ

Mehmet KAHRAMAN

Yabancı Diller Eğitimi Anabilim Dalı
İngilizce Öğretmenliği Programı
Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü, Nisan 2021

Danışman: Dr. Öğr. Üyesi Gonca SUBAŞI

Kelime gruplarının ve özellikle eşdizimsellik bilgisinin yabancı bir dilde yetkinliğin önemli bir parçası olduğu sıklıkla iddia edilmektedir. Ayrıca, yüksek-frekanslı fiillerin öğrenenler için özellikle diğer kelimelerle kombine edildiğinde yanıltıcı olduğu da iddia edilmektedir. Anadil konuşanlarıyla kıyaslandığında bazı öğrenenler yüksek-frekanslı fiileri daha sık kullanırken, bazıları ise bunların eşdizimsellik özelliklerini ve sözcüksel olmayan anlamlarını göz önüne alarak kullanmaktan kaçınmaktadır. Bu çalışma İngilizce öğrenen Türk öğrencilerin tartışmacı yazılarında *make* ve *do* yüksek-frekanslı fiilerinin bir fiil+isim kombinasyonundaki kullanımlarını ortaya çıkarmayı hedeflemektedir. Bu bağlamda, bu çalışma öğrenen derlemindeki aşırı kullanımı/ yetersiz kullanımı, dilbilgisel ve anlamsal örüntüleri ve hatalı kullanımları incelemiştir. Bu inceleme, karşılaştırma amacıyla, başvuru anadil derlemi olarak LOCNESS derleminden faydalanmıştır. Bulgular öğrencilerin *make*+isim gruplarını anadil derlemine göre yetersiz kullandığını; ayrıca dilbilgisel ve anlamsal olarak bazı benzerlik ve farklılıklar gösterdiğini ortaya koymuştur. Son olarak, hata incelemesi öğrencilerin *make* ve *do* fiil+isim kombinasyonlarında bazı problemleri olduğunu ortaya koymuştur. Bu çalışma ve önceki çalışmaların bulgularına dayanarak, çalışma özellikle dil öğretim ortamları için kuramsal ve uygulamalı öneriler paylaşmıştır.

Anahtar Kelimeler: Kelime grupları, Yüksek-frekanslı fiiller, Eşdizimliler, Derlem

ACKNOWLEDGEMENTS

Writing this dissertation represents not only my work at the keyboard. It would not have been possible to do without the guidance and support that I received from many people around me. I would like to thank all the people who contributed to the work described in this thesis in some way.

First and foremost, I am extremely grateful to my supervisor, Assist. Prof. Dr. Gonca SUBAŞI for her invaluable advice, continuous support, and patience during my PhD study. Particularly, the smooth communication with her and rapid feedback from her side on the way constructing this dissertation kept me more concentrated and motivated.

Also, I would like to thank my esteemed dissertation committee members Prof. Dr. Gül DURMUŞOĞLU-KÖSE and Prof. Dr. Işıl KABAKÇI-YURDAKUL. Thanks to their vast knowledge and experience, I felt more confident in each step of the study. Their continuous assistance and constructive feedback helped me a lot to walk in the right direction. I hope that I can "pay forward" what they have done for me by helping other people. I will use their selfless support as a model as I move forward. Besides, I would like to extend my grateful thanks to my defence committee members Prof. Dr. İlknur SAVAŞKAN and Assist. Prof. Dr. İlknur YÜKSEL for generously offering their time, support, guidance and good will through insightful comments and useful suggestions. Their suggestions brought in threads of thought that made my research much richer, and my dissertation something I can be proud of having written.

I would also like to thank my friends and colleagues İbrahim UZ and Caitlin Marie ANTAYA for their technical support in this work. Their native speaker judgment of Turkish and English languages and their experience in language teaching made my analyses much more reliable and objective.

Additionally, I wish to thank my beloved life companion, Rukiye, who has stood by me through all my travails, my absences, my fits of pique and impatience. She gave me help and support, discussed ideas and prevented several wrong turns. Along with her, I want to acknowledge my son Ahmet. He has been a great source of love and relief from scholarly endeavour. Also, my parents, my sister, my parents-in-law and my sister-in-law have all supported me wholeheartedly throughout the process. I feel lucky to have them beside me.

My appreciation also goes out to my friends for their encouragement and support all through my studies. Without their understanding and encouragement in the past few years, it would be much harder for me to complete my study.

Mehmet KAHRAMAN
Eskişehir 2021

..../..../20....

STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES

I hereby truthfully declare that this thesis is an original work prepared by me; that I have behaved in accordance with the scientific ethical principles and rules throughout the stages of preparation, data collection, analysis and presentation of my work; that I have cited the sources of all the data and information that could be obtained within the scope of this study, and included these sources in the references section; and that this study has been scanned for plagiarism with "scientific plagiarism detection program" used by Anadolu University, and that "it does not have any plagiarism" whatsoever. I also declare that, if a case contrary to my declaration is detected in my work at any time, I hereby express my consent to all the ethical and legal consequences that are involved.

…….

Mehmet KAHRAMAN

THE LIST OF CONTENTS

Page

Page

CHAPTER 4

CHAPTER 5

THE LIST OF TABLES

Page

THE LIST OF ABBREVIATIONS

C	: Combined Method
COCA	: Corpus of Contemporary American English
D	: Definition
DDL	: Data-driven Learning
EFL	: English as a Foreign Language
EI	: Explicit Instruction
ELT	: English Language Teaching
ESL	: English as a Second Language
FFI	: Form-Focused Instruction
GeCLE	: German Corpus of Learner English
ICLE-TRCU	: International Learner English- Turkish
ILCoWE	: Israeli Learner Corpus of Written English
L1	: First Language
L2	: Second Language
LOCNESS	: Louvain Corpus of Native English Essays
MFI	: Meaning-Focused Instruction
NA	: Not Applicable
P	: Pattern

CHAPTER 1

1. INTRODUCTION

1.1. Background to the Study

Examining the use of "make" and "do" across native and learner corpora is closely linked to two interrelated phraseological study areas; word combinations and high-frequency verbs.

Ever since Firth (1968) brought to the attention for the first time, word combinations have been subject to many studies (Deveci, 2018; Greenbaum, 1974; Halliday and Hassan, 1976; Koya, 2005; Laufer and Waldman, 2011; Richards, 1976; Sinclair, 1991) though the major emphasis is on collocations. Due to the rise of emphasis on pragmatic competence along with psycholinguistics since 1970s and availability of large corpora such as *British National Corpus (BNC)*, vocabulary teaching gained importance. However, many linguists focused on productivity rather than individual word in a sentence. This led the researchers to consider collocations at a greater extent (Koya, 2005). Brown (1974) for example, stated that vocabulary learning by word lists is dull, and also it might be dangerous due to the possible problems with concept formation and use; and course book exercises on the other hand, force students to write and speak in complete sentences. However, Brown (1974) suggested, choosing between the word or the sentence is not mandatory, we can choose collocational groups since we normally write and speak in chunks of language. Later studies, also, reinforced the significance of the word combinations, rather than individual words. For example, Howarth (1998) found in a 240000-word-corpus that there were 5000 verb-noun combinations and Nesselhauf (2005) argued that the large number of prefabricated units decrease the processing effort in human brain, and thus provide the speaker with fluency. Since then, many scholars compared collocation teaching to traditional vocabulary teaching, and they also found strong evidence indicating that collocation teaching resulted in better results in terms of number, retention or time spent (Chan and Liou, 2005; Gilquin, 2007; Kennedy, 2010; Şimşek, 2008; Vural, 2010)

High frequency verbs make up the other half of the background of the present study. According to Nation (1990), knowing a word in a productive sense requires knowing it

in full aspects such as spelling, pronunciation, grammatical patterns, frequency, context of use and possible collocations. High frequency verbs have certain characteristics which are common across the languages. For example, they dominate different semantic fields, they have equivalent matches in most of the languages, they have both universal and language specific meanings and they potentially create problems for the learners (Altenberg and Granger, 2001). The idea is also confirmed by Gilquin (2007), which argues that language learners are mostly ignorant of the collocates of high-frequency verbs, though they are highly familiar with the core meanings of those verbs. Interestingly, both overuse and underuse of high-frequency verbs are observed in learners' productions. Hugon (2008) concludes that learners overuse them because they feel safe when they use high-frequency verbs since they are well-learnt. However, some other learners tend to avoid using high frequency verbs and prefer using rarer words, which make the meaning awkward, especially when it comes to idiomatic phrases (Sinclair, 1991).

In sum, high-frequency verbs are tricky for the learners especially when they are in combination with other words. Some learners feel safer with them and ignore the properties of context and collocates, while some others avoid using them when it comes to idiomatic meaning.

1.2. Statement of the Problem

Lennon (1996) provides the evidence that although learners know the verb meaning, that knowledge is hazy when it comes to polysemy and grammatical, phrasal, contextual or collocational restrictions of a high frequency verb. The problem seems to be two folded. Firstly, high-frequency verbs such as *do*, *make*, *have*, *take*, *get*…etc. are among the most common verbs and usually learned at the early days of language learning, yet they are still difficult to gain full attainment because they often appear in various fixed collocational phrases in addition to their other properties (Liu and Lei, 2009). Altenberg and Granger (2001) state that high frequency verbs tend to be neglected by the teachers and learners once they have been taught. Since these verbs are extremely complex and learners cannot deeply understand their grammatical and lexical patterning, it is possible to observe overuse, underuse or deviant forms in learner language.

Secondly, many scholars argue that word combinations, and, thus collocational knowledge is an essential part of the mastery in a foreign language (Almacioğlu, 2018; Farooqui, 2016; Sinclair, 1991; Woolard, 2000; Wray, 1999). The proper use of collocations is regarded as one of the fundamental parameters that distinguish a native speaker from a non-native one (Durrant, 2014; Pawley and Syder, 1983). However, aside from their literal meaning, collocational phrases might have figurative or duplex meanings, which makes the case harder for the learners (Macis and Schmitt, 2016). Therefore, one can conclude that the studies on high-frequency verbs need to be blended with collocation studies so that they can produce fruitful implications pedagogically.

It is possible to see different types of lexical collocations studied in the field. Although there are various word combinations, verb-noun collocations have drawn more attention due to its higher frequency and effect size in communication (Chan and Liou, 2005; Howarth, 1998; Nesselhauf, 2005); and difficulty in learning (Howarth, 1998). Also, as elaborated by Nesselhauf (2005: 15), the term "collocation" is applied to all three restriction levels, *free combination, collocations* and *idioms by* different authors. Although they are theoretically located on the same line in the form of a continuum, the confusion of the terms might still produce confusing results for the researchers and teachers.

In the context of Turkey, it is possible to find many collocation studies, as well. However, a vast majority of them deal with the effects of collocation teaching through experimental studies (Akıncı, 2009; Kayıran, 2012; Qader, 2018; Vural, 2010). There are also some studies which aim to describe students' controlled collocational knowledge or awareness; or students' and teachers' perceptions of collocations (Almacioğlu, 2018; Koç, 2006; Mutlu, 2015). As for *make/do* collocations, Öztuna (2009) and Shibliyev (1993) have important studies, however they do not make use of any corpus.

Considering the aforementioned problems and facts, the present study focused on the use of high-frequency verbs *make* and *do* when they occurred in a verb+noun word combination. The term "*word combination"* was deliberately chosen due to the fact that the present study did not distinguish between word combinations in terms of their restriction levels. In other words, free combinations, collocations and idioms were all included in the current study. With regard to COCA (Davies, 2008), which consists of more than 600 million words, "*do*" is the third and "*make*" is the ninth most frequent verb in English. The present study dealt with *make* and *do* verb-noun combinations only, since

they are frequently confused by most of the Turkish EFL learners (Öztuna, 2009), as many other L2 English learners (Altenberg and Granger, 2001). Basically, a learner corpus was compared to a native corpus in this study. Since the comprehensive and authentic studies of language use cannot rely on small samples or anecdotes, the corpus-based approach was taken as a more feasible alternative to study large amount of natural data (Biber, Conrad and Reppen, 1998, p. 3). A comprehensive and authentic study of language use was necessary for creating a baseline since corpus-based approaches serve a transition to elaboration of better-quality learner input, and thus teachers and researchers are provided with a wider perspective of language as stated by Campoy, Belles and Gea (2010).

1.3. Aim

The main purpose of this study is to analyse "*make*" and "*do*" verb-noun combinations in argumentative essays of Turkish learners of English, who are at intermediate level. The natural uses of "*make*" and "*do*" verb-noun combinations in a learner corpus are investigated by taking a native corpus as the reference.

Within this framework, the study attempts to answer the following research questions:

1. What are the grammatical patterns in "*make*" and "*do*" verb+noun combinations produced by the learners and native speakers?
2. Among various dictionary definitions of the verbs "*make*" and "*do*", which meanings are considered by the learners and native speakers?
3. What are the possible error spots, error types and resorted strategies in the deviant "*make*" and "*do*" verb+noun combinations in the learner corpus?

1.4. Significance of the Study

The study is expected to provide a comprehensive understanding of *make/do* verb+noun combinations of Turkish learners of English with reference to native productions. Firstly, the cases of underuse and overuse are examined. This is supposed to give an overall idea about whether Turkish learners of English tend to use *make* and *do* as frequently as the native speakers of English. The relevant literature involves cases of both overuse and underuse. The findings of the present study are expected to contribute to the existing compilation of studies in the field. Secondly, the current study explores the

grammatical and semantic distribution of *make/do* verb-noun combinations. In many studies (*e.g:* Babanoğlu, 2014; Hugon, 2008; Kim, 2015), the same exploration is tried to handle by employing the categorization suggested by Altenberg and Granger (2001). However, this categorization is only applicable for "*make*", and thus not suitable for "*do*" as another high-frequency verb. Also, the same categorization is applied in an attempt to investigate both grammatical and semantic properties of both high-frequency verbs at the same time, which usually makes it harder to focus on each property individually. Therefore, the present study, unlike many other studies, examines the grammatical and semantic properties of *make/do* combinations separately using particular categorizations for each aspect. Such an analysis is expected to reveal the strengths and weaknesses of the learner language while giving an in-depth idea on which aspect of collation teaching should be emphasized more in ELT classes. Thirdly, since there are already existing studies suggesting the significance of collocation teaching, the current study is an attempt to show some caveats in collocation teaching by presenting a detailed error analysis. Unfortunately, the terminology and the way some other studies treated the errors do not follow a clear categorization. Where the error occurs, what is missing in a learner's interlanguage or which strategy is used by a learner are somehow elusive in the error analyses in some other studies. Lack of a clear categorization might have vague implications. The present study, however, deals with the errors under three sub-categorizations. They are *error spots*, *error types* and *strategies*. Investigating the deviant *make/do* combinations in an organized way, the present study is expected to yield diagnostic results which might move the discussion from "teaching collocations" to "how to teach collocations". Finally, exploring *make/do* verb+noun combinations of non-native speakers by referring to a native corpus, this study might be a unique one in Turkish context. Thus, the study is an attempt to fill in a gap in the literature, as well.

1.5. Definitions

Collocation: The combinations in which one part of the phrase is freely chosen and the other part is assigned considering the first one (Mel'čuk, 1998). A *make/do* verb-noun combination contains either *make* or *do* as the main verb and a following noun; e.g: *make an effort, do homework.*

AntConc concordance software: It is a freeware corpus analysis toolkit for concordancing and text analysis which is designed by Laurence Anthony (L. Anthony, 2019)

Corpus size: The condition that each corpus in comparison should be large enough to represent the distribution of linguistic features accurately. (Biber, 2010).

Corpus composition: "Composition" refers to the condition that each corpus must be sampled deliberately to represent the registers in use (Biber, 2010).

Delexical verbs: They are the verbs that contribute very little to the whole meaning in a verb+noun combination. In such a combination, typically, nouns carry the meaning rather than the verbs and therefore the verbs in this type of combinations are sometimes called "*light verbs*", as well. For example, in *make an arrangement,* the whole meaning greatly corresponds with *arrange.* (Nesselhauf, 2005: 20)

CHAPTER 2

2. LITERATURE REVIEW

The first part (2.1) of this review starts with fundamental issues and discussions in collocation and collocation studies in order to present a general framework of the topic. The second part (2.2) gives an idea of how collocations are dealt in educational settings in chronological order. The third part (2.3) outlines some studies on high-frequency verbs and *make/do* collocations, in particular. The last part (2.4) introduces the academic theses written on collocations at Turkish universities.

2.1. Fundamental Issues in Collocation Studies

First of all, collocations are divided into two main categories: grammatical collocations and lexical collocations. The former is defined as "a phrase consisting of a dominant word and a preposition or grammatical structure"; e.g.: *account for* (Benson, Benson, and Ilson, 1990: ix). The latter, "in contrast to grammatical collocations, consist of nouns, adjectives, verbs, and adverbs"; e.g.: *warmest regards*(Benson et al., 1990: ix)

Secondly, approaches to the collocations are another issue which needs a clarification in order to gain insight in collocation studies. There are two leading traditions in collocation studies: *frequency-based approach* and *phraseological approach*; thus, the definition of collocation is shaped according to these traditions (Nesselhauf, 2004). Within the scope of frequency-based approach, a collocation is defined as the "the occurrence of two or more words within a short space of each other in a text"(Sinclair, 1991: 170). The "*short space*" is defined as four words around the investigated word. According to this view, a collocation is composed of one "node", which is the investigated word, and "collocates", which are the other words in relation with the node within the range of four words to the left or right. Sinclair (1991) distinguishes between "casual" and "significant" co-occurrences, and only the significant ones are referred as collocations. Stubbs (1995), a follower of frequency-based approach, elaborates the idea by suggesting quantitative measurement of co-occurrences using corpora. He exemplifies his claim by searching the lemma "cause" in a 250-million-word corpus. As a result of the study, it is argued that the lemma "cause" is dominantly used in negative or unpleasant

contexts, and it is possible to observe and calculate the collocates of "cause" using large corpora. As to phraseological approach, it is considerably led by the works of Cowie (1981, 1998, 2001). Although there are certain variations in terminology or categorization across different phraseologists, collocation can be defined as lexical combinations which are relatively restricted and transparent in meaning. According to Aisenstadt (1979), all word combinations are either idioms or non-idiomatic phrases. The non-idiomatic phrases include restricted collocations and free phrases. Cowie (2001) argues that word combinations are divided into semantic combinations and pragmatic combinations. The former consists of collocations and idioms, while the latter consists of proverbs and routine formulae. The common ground of these two categorizations is the emphasis on the semantic aspect of the combinations, which is the basic difference between frequency-based approach and phraseological approach. According to phraseological approach, collocations are different from free phrases, due to their restricted word choice. For example, considering the phrase "*drink tea*", one can easily substitute "*tea*" with *water, beer* …etc. However, considering the phrase "*perform a task*", substitution of "*perform*" (*e.g.: make*) is not similarly possible. Collocations are also different from idioms since the elements of the phrase "*perform a task*" still has semantic relation with "*perform*" and/or "*task*" individually. However, "*blow the gaff*" does not keep semantic cord with the individual elements of the phrase (Nesselhauf, 2005).

Finally, in relation with the above mentioned categories and approaches, dictionaries have been shaped (Cowie, 1998). For example, *The Oxford Dictionary of Current Idiomatic English- Vol-1* (Cowie and Ronald, 1975) tended to provide grammatical description of the entries. However, *The Oxford Dictionary of Current Idiomatic English- Vol-2* (Cowie, Mackin, McCaig, 1984) included more idioms and collocations, which is inspired by the Russian-inspired phraseological types. In order to satisfy the need for more comprehensive collocational dictionary, *The Selected English Collocations* (Kozłowska and Dzierżanowska, 1982), and then *BBI Combinatory Dictionary of English* (Benson et al., 1990) were introduced.

2.2. Collocational Knowledge of L2 Learners

According to Hsu (2007), although there have been theoretical discussions on collocation ever since Firth (1968) introduced it to the field of theoretical linguistics, pedagogical investigation of the issue has come to the fore along with Lewis' (1993)

book, *Lexical Approach*. Based on the theoretical discussions in the field, many scholars have studied the significance of collocations and possible applications of collocation teaching in ESL settings.

One of the first studies on collocation knowledge was conducted by Zhang (1993). He gathered collocation knowledge tests and writing samples from 30 native and 30 non-native students. The findings showed that native writers had significantly higher collocation test scores, and in the same vein they had higher writing scores comparing to non-native peers. The study implied that variety and accuracy of the collocations were indicatives of better written communication. Soon after Zhang, the significance of collocations was investigated in Arabic region. Farghal and Obiedat (1995) asked college students to fill in the blanks by using twenty-two common collocations from English to Arabic, and English teachers to translate the equivalent collocations from Arabic to English. The results revealed that both groups had problems with collocational knowledge. The study also explored participants' strategies in the tasks. The most commonly used strategies were synonymy, paraphrasing, avoidance and transfer.

The literature also includes corpus-based studies which have given new directions to the field of language teaching (Conrad, 1999). For instance, one significant corpus-based study was on learners' difficulties with collocations which was conducted by Nesselhauf (2003). She collected data from German learners of English and attempted to explore erroneous collocation productions in their free writing essays. The results showed that although the degree of restriction on collocations predicted some portion of their mistakes, L1 interference was the most operative factor of their erroneous collocation productions. Due to the technological advance in all over the world, collocation teaching has been subject to technology oriented studies, too. Chan and Liou (2005), for example, conducted an experimental study. The researchers implemented a web-based bilingual Chinese-English concordance program for teaching collocations to thirty-two Chinese learners of English. They had only one group, which was treated as the experimental group. The implementation lasted for five units. The results of the pre-test and post-tense scores revealed that there was a significant improvement in collocation knowledge at the end of the treatment, yet a regression was observed in their delayed-post test scores. Nevertheless, the participants' performance was still better than their condition prior to the study.

Several years later, Hsu (2007) conducted a study similar to that of Zhang (1993), which is mentioned above. For his research, the researcher collected writing samples from sixty-two students in Taiwan and investigated their collocation productions in terms of frequency and variety by means of a web-based writing program. The result of the study indicated a positive correlation between students' collocational productions and their writing scores, which is hardly distinguishable from the findings of Zhang (1993). While Zhang (1993) and Hsu (2007) were interested in what collocation knowledge leads to, Shehata (2008), on the other hand, investigated what leads collocation knowledge to be better. Particularly, two factors, Arabic L1 and learning environment as ESL/EFL, were examined. The data were collected from sixty-two English majors in a university in Egypt (EFL) and thirty-five Arabic learners of English studying at a university in the United States (ESL). The study findings showed that both L1 (Arabic) and participants' learning environment were influential in collocation learning. The participants in ESL environment were more successful in collocation acquisition. Also, thanks to the data collection tools, it was possible to observe that the participants' receptive knowledge of collocations was broader than productive knowledge.

Aside from L1 influence and learning environment, Durrant (2008), in addition to some other points which are not directly related here, investigated the role of input frequency on adult learners' acquisition of collocations. The findings showed that lower input frequency resulted in lower acquisition level, while higher input frequency improved collocation learning. In their study, Laufer and Waldman (2011) focused on both the proficiency levels and erroneous/overused productions of learner. For this purpose, they compared three learner corpora to a native corpus as the reference. The study revealed that all three non-native corpora were below the native corpus in terms of collocation frequency and accuracy; also, although the advanced level corpus involved less errors comparing to other two non-native corpora, L1 based errors were persistent in all non-native corpora.

2.3. High-Frequency Verbs and Make/Do Collocations

Although collocational knowledge is a major component of the present study, it is also closely related to high-frequency verbs and *make/do* collocations, in particular. As for high-frequency verbs, one of the fundamental studies in the field is that of Lennon's (1996). The study revealed that high-frequency verbs are taught on the first days of

language learning, and then learners are supposed to gain new but low-frequency verbs. However, the study implies that teaching at advanced level should also aim to master the incomplete entry level high-frequency verbs. Murao (2004) investigated the verb-noun collocations of high-frequency verbs. The acceptability judgement tests were given to the intermediate and advanced level Japanese learners of English. The results showed that both intermediate and advanced learners have problems with the collocates of high-frequency verbs. The researcher draws attention to the need for negative evidence, contrastive approach as an implication of the study. Juknevičienė (2008), also, studied the high-frequency verbs, such as "*have, take, do, make, give*". The study showed that Lithuanian learners of English have problems with academic vocabulary with high-frequency verbs. Moreover, they tend to compensate their deficiency by resorting to their L1 translations, which, often, result in clumsy word combinations.

There are also some other studies trying to illustrate the semantic or syntactic difference among collocations and combinations with high-frequency verbs, in particular. Macis and Schmitt (2016), for example, differentiated between literal, figurative and duplex meanings of the collocations. The study contributed to the language teaching by bringing useful insights to the collocations in terms of their semantic nature. As for high-frequency verbs, Liu and Lei (2009) investigated a native corpus, COCA, and highlighted the deep semantic differences among the verb-noun combinations with *make, take, do* and *have.* Lantolf and Tsai (2018) focused only on *make* and *do* in semantic terms. They used SCBOAs to illustrate the deep semantic difference between make and do, then they asked the participant learners to draw their own SCOBAs for other combinations. The pre-test, post-test and delayed post-test scores revealed a significant improvement in using *make* and *do* verb noun combinations. The study is significant since it implements developmental education which is grounded in Vygotsky's *sociocultural theory* (Vygotsky, 1978) and Gal'perin's concepts of *materialization, verbalization and internalization* as defined in *systemic theoretical instruction* (Gal'perin, 1969).

Syntactic analysis of word combinations, also, has been another scope of studies in the field. Hiltunen (1999) examined a huge size of a corpus which was piled from a number of Early Modern English texts. The study shows various grammatical verb-noun phrases in detail, while shedding light on the verbs, verbal phrases and phrasal verbs. There are also some studies dealing with "make" only, due to its various uses. Gilquin (2007) compared advanced level French learners of English corpus to a native corpus;

and used elicitation tests to investigate *make* collocations. The study showed that learners used less *make* collocations comparing to the native corpus. They resorted to the collocations which have direct translations in their L1, which were safer for them. Although corpora comparison did not reveal significant deal of errors, the elicitation task revealed some deficiencies. Hugon (2008) examined the French learners of English in terms of their use of *make* in different semantic and syntactic categories through a comparison of corpora. The results showed that the learner corpus showed a varied degree of accuracy in terms of semantic categorization; however, delexical combinations of *make* had many deficiencies. Lareo (2009) concentrated on the noun used after *make* in verb phrases in scientific texts. The researcher compared a mathematic sub-corpus to a fiction text. The analysis revealed that, *make* combinations with specific nouns almost doubled the combinations with general ones.

More recently, the data-driven learning (DDL) has come to the fore due to the its close particular focus on vocabulary and collocation teaching. Huang (2014), for instance, conducted a study to test the effectiveness of DDL. The participants were 40 university students taking the course English for Business Purposes in China. The students were divided into one experimental and one control group randomly. Both groups were given a writing test before the experiment and the results showed that there was no statistically significant difference in English writing competence between the two groups. Both groups were asked to write on the same topics. They were required to use certain nouns in their argumentative essays on the tests, except for the delayed post-test. However, the control group consulted to dictionaries for the vocabulary items whereas the experimental group was provided with ten concordance lines of each target word during the corpus-based activities. The results showed that written productions of the experimental group contained a higher variety of collocational and colligational patterns comparing to the control group. In addition, the experimental group had fewer linguistic errors in using the target abstract nouns. Men (2020), also, tested the effectiveness of DDL model particularly contextualized for synonymous words in comparison with traditional techniques. The participants were 52 first-year university students majoring at English in China. The students were equally distributed into one experimental and one control group. The groups were at the same proficiency level based on their performances in their earlier test scores. As the pre-test both groups were given a test of twenty multiple choice items. The questions were testing the synonyms and their collocations which were taken from

their reading course textbook. After the pre-test, which took seven minutes online, the researcher employed a teacher's demo session with the groups in their assigned manners. The results suggested a significant increase in the experimental group's performance on collocation production thanks to the differentiation among the synonymous words by using extractions from the concordance lines. Moreover, the experimental group submitted the test in a shorter time than the control group. Finally, a great majority of the students found DDL model useful. Lay and Yavuz (2020) aimed to find out whether contextualized DDL targeting the interlingual interference helps reducing the errors in the written productions of low-intermediate students. The participants were 30 first year Turkish learners of English at a university. Their proficiency level was roughly B1. Prior to the DDL sessions, the researchers determined a list of 10 most common issues resulted by L1 influence in Turkish learners of English. This listing was gradually made possible by a literature review, and then asking the opinions of experienced teachers. For the next step, the participants were asked to write essays as responses to some basic prompts in which the researchers expected to observe the pre-determined target issues. Taking it as the pre-test, the students were given a treatment period for 11 weeks in a computer lab where each student can have their own computer. At the beginning of the treatment sessions, students were introduced and trained how to use TCSE corpus to sort out the target structures. Finally, the students were given the post-test and their written productions were analysed in terms of target issues comparing to the pre-test scores. The results showed that DDL activities worked at a moderate effect size.

In addition to comparison of DDL model to the traditional methodology, the literature includes some other studies keeping their scope within the DDL model. For example, Sun and Hu (2020) conducted a study in order to reveal the difference between students' direct and indirect exploration of the corpus within DDL model. The experiment took three weeks. The students completed the first writing before they were divided into two groups- direct DDL group and indirect DDL group. Then, they took their first and second instructional sessions in their assigned groups. Michigan Corpus of Upper-level Student Papers and the International Corpus Network of Asian Learners of English were used for instructional sessions. They were chosen for their suitable content and size. In the instructional sessions, the direct DDL group conducted guided searching activities in the corpora to explore the hedges and worked on them while the indirect DDL group just worked on the language samples which were selected and adapted beforehand. The same

teacher conducted all four sessions for the two treatment groups so that there were not teacher-related effects. Then, the students were given a questionnaire survey at the end of the second session. On the next day, both groups took the second writing test. Finally, they were given the delayed post-test two weeks later. The essay prompts were those requiring students to write argumentative texts. The concentration of the researchers was the frequency and variety of hedges in students' academic English writing. Only correct uses of hedges were counted for the analysis. The results showed that direct DDL treatment meaningfully affected the frequencies of the hedges in students' writing with medium effect size. However, the study still yielded both positive and negative aspects of both approaches. And the most recently, Otto (2021) focused on the word selection step, which is also an important part of the application of DDL model in language teaching settings. The researcher intended to identify better suited vocabulary to teach using DDL in an English for specific purposes courses. The motivation of the study was to show that it is possible to select DDL-friendly vocabulary which are important for a particular field-based English course and, at the same time, the vocabulary could be the ones which the students are not very proficient with. To this end, the researcher made use of three corpora- an expert civil engineering corpus compiled from professional documents written in the field, a student civil engineering corpus compiled from student writing assignments, and The Corpus of Contemporary American English (COCA). Firstly, a list of words particular to civil engineering was extracted, then the list was narrowed down through a comparative analysis based on certain criteria such as grammatical categories or frequency. Finally, a DDL suitability judgments were performed in order to decide if a word was suitable for teaching through DDL. Afterall, the analyses suggested 18 target words for the students of civil engineering department through DDL methodology. The system was effective since it provided better suited vocabulary which are potentially problematic for the learners. Furthermore, revealing how these words function contextually and how to teach these functions and usage, the system was highly advantageous over the traditional ways. At the end of the study, the researcher also stated some limitations of the system.

As can be seen in the literature, studies dealing with teaching vocabulary or teaching writing have covered high-frequency verbs or *make/do* collocations in some ways. More recently, DDL model has gained interest due to its effective use of corpora in language teaching.

2.4. Collocation Studies in Turkey

The theses on collocations in Turkey have been reviewed, and it is inferred that the researchers have approached the issue from three different aspects:

- Collocation teaching
- Error analysis of collocation productions
- Comparison of collocation use across genres

A considerable bulk of studies are those investigating the effectiveness of collocation teaching and; therefore, learners' gains in terms of collocation productions or collocation awareness. Gençer (2004), firstly, conducted an experimental study by participation of an experimental group and a control group. Both groups were given the same two texts to study on, but only the experimental group students were asked to pay attention to the collocations within the texts. Students were given two tests for each text requiring both receptive and productive skills to complete. The results showed that experimental group scored higher in both tests. The finding implied the significance of explicitness in collocation teaching, which results in higher awareness. Balcı (2006), also, conducted a similar study, in which traditional vocabulary teaching methods were compared to collocation teaching. The researcher focused not only on vocabulary learning, but also on vocabulary retention. The comparison of control and experimental group revealed that teaching vocabulary through collocations had positive effects on both acquisition and retention of the new vocabulary. Avcı (2006) explored whether paying extra effort in collocations would make a difference among the written productions of the pre-intermediate level students. To this aim, prior to writing, both the control group and experimental group read related texts and had brainstorming activities. In addition to this procedure, the experimental group was given possible collocations of the suggested words in reading texts, and during the brainstorming sessions, students' collocational mistakes were given corrective feedback. When the written productions of the students analysed, it was observed that the experimental group did not outperformed the control group in terms of writing scores, although they used higher frequency of collocations.

Similarly, in his study, Şimşek (2008) presented new vocabulary items to the control group by solely giving definitions, to the experimental group by collocations. The post-test scores showed no significant difference in their vocabulary gains, however

delayed post- test scores showed that the experimental group had higher retention rate comparing to the control group. This finding implied the significance of collocations in terms of vocabulary retention rather than acquisition. Öztuna (2009) adopted a narrower scope in her research. Instead of vocabulary gains, the researcher preferred to focus on *make/do* collocations. Negative evidence was taken as explicit treatment and input flood was taken as implicit treatment. Turkish learners of English at 7th grade were randomly divided into three groups as control group, negative evidence group and input flood group. After the treatments, performances were measured at both recognition and production levels by means of four different tests. The results showed that except for the input group having the similar score with control group at recognition level, both negative evidence and input flood groups got significantly higher scores in post-tests. When the delayed post-tests were concerned, it was seen that both treatment groups kept significant differences with the control group in all tests, though there was still a decrease in their scores. The researcher offered to employ both explicit and implicit treatments together in order to make maximum use of them.

Akıncı (2009) compared the effectiveness of three different treatments- data-drive learning (DDL), explicit instruction (EI) and combined method (C)- on verb-noun collocation development of one single group of participants. Aside from the performance tests, the researcher also involved participants' preference of treatment. The results showed no significant difference among three treatments in terms of performance on collocations, yet participants opted for DDL method, which refers to corpus consultancy in the study. Kayıran (2012) investigated the effects of collocation teaching on vocabulary development of the 9th grade students. After a five-week treatment to a single group, the comparison of pre-test and post-test scores showed that participant students improved their vocabulary knowledge at the end of the study. In addition to the test scores, the students were also interviewed at the end of the treatment, and the qualitative data revealed that the students gained an awareness of collocations. Öztuna (2014) attempted to test the computability of Laufer and Hulstijn's (2001) *Involvement Load Hypothesis* with collocation teaching. To this aim, the researcher created an online platform and supplied materials to it considering the steps as the hypothesis proposed (*need, research, evaluate*). The treatment was applied to one elementary and one pre-intermediate group for five weeks. However, the findings did not confirm the hypothesis. The researcher stated that although the hypothesis was based on both cognitive and affective factors,

learners' attitudes, cultural background and education policies might have caused this inconsistency between the theory and the practice. In fact, some of the students described the treatment as "boring", "confusing" and "too mechanical", which supported the researcher's opinion.

Mutlu (2015) compared average vocabulary knowledge, receptive verb-noun collocation knowledge and productive verb-noun collocation knowledge of the learners, and also, she analysed learners' and teachers' opinions about collocation learning/teaching. The findings, first, indicated that there was a positive correlation between learners' average vocabulary knowledge and collocational knowledge. Secondly, it was observed that productive verb-noun collocation knowledge of the learners was more limited than their receptive knowledge. Thirdly, both students and teachers agreed on the importance of collocations, however students thought the teachers did not allocate enough time for collocation teaching while the teachers thought they did. Finally, both students and teachers agreed that L1 was the main source of errors in collocation productions. Aydoğan (2016) made another comparison in her study. The researcher investigated the effects of input- oriented tasks and output-oriented tasks on collocation development. The measurement considered both active and passive recalls. Input oriented and output-oriented groups were assigned as control and experimental groups, respectively. However, the analysis of the pre-test and post-test scores yielded no significant difference between two groups in either recall type.

Qader (2018) examined the effects of teaching collocations on learners' speaking fluency. The researcher hypothesized that knowing collocations could enhance learners' oral proficiency. A total of 45 advanced level university students were divided into one control and one experimental group. The analysis of the pre-test and post-test scores of the groups showed that collocation-oriented teaching significantly increased the speaking proficiency of the students. The researcher concluded that mastering the collocations decreases the anxiety of the speaker and, thus help them speak more fluently. Sobucalı (2019) investigated the effects of meaning-focused (MFI) and form- focused instruction (FFI) on collocation teaching to Turkish learners of English. The researcher divided the learners into three groups as MFI, FFI and control group. 20 selected collocations were taught in line with each group's treatment. The results showed that experimental groups improved significantly comparing to the control group. Also, FFI group was more successful in form recognition than MFI group. Salihoğlu (2019) compared the implicit

collocation teaching to explicit collocation teaching. The participants were 61 university students. Before the experimental process, students were grouped into two (implicit group and explicit group) and pre-tested. Then, target collocations were extracted from a large corpus. After that, the implicit group was exposed to target collocations through reading and listening activities whilst the explicit group was given collocation learning tasks for 14 weeks. When the post-test was given, the results showed that the explicit group outperformed the implicit group.

Durmuş (2019) investigated the effectiveness of implicit and explicit teaching of adjective+ noun collocations to 42 secondary school students. Splitting the participants into two groups, the researcher conducted a pre-test, a teaching program, a post-test and a delayed post-test within 10 weeks. The results yielded significant differences across the tests within each group. However, there was no significant difference across the groups. The results implied that both explicit and implicit teaching programs helped students of both groups improve their collocation knowledge. More recently, Bozoğlu (2020) compared collocation teaching to the traditional vocabulary teaching techniques. The participants were a total of 52 seventh grade students. 16 of them were reserved as the experimental group. Following a pre-test, treatment, a post-test and a delayed post-test session, the researcher also integrated teacher's notes during the treatment, and a writing task and student interviews after the treatment into the research procedure. The results, overall, illustrated that the experimental group significantly outperformed the control group; moreover, the experimental group students reported that it was more enjoyable to learn vocabulary through collocations.

Research on collocations is not limited to collocation teaching. Thanks to the availability of corpus tools, error analysis type of research has become possible in Turkish context. Shibliyev (1993), for example, conducted a study by participation of 36 beginner level Turkic learners of English. The study employed a translation task and acceptability judgement test to investigate *make* and *do* combinations. The test results showed that students' unawareness of colloquial usage, high restrictedness of some collocations and collocations involving prepositions were the main cause of the errors. Yılmaz (2004) conducted a study focusing on the collocational errors of the advanced level learners. One control and one experimental group were pre-tested on collocations and idioms, and the experimental group was given eight weeks of treatment based on their errors on occurred in pre-test. Then, both groups were given a post-test. The analysis of the pre-test and post-

test revealed that collocation teaching based on error analysis made a significant improvement in performance of the experimental group students. Hama (2010) attempted to explore the learners' source of errors in their collocation productions. A total of 40 Kurdish learners of English were sampled for the study. The researcher adopted collocation completion test and think-aloud protocols as the data collection tools rather than designing an experimental study. The analysis of the tests and protocols showed that lower frequency of certain collocations and L1 interference had the strongest impact on learners' erroneous productions.

More recently, Bıçkı (2012) elaborated the error analysis by focusing on the structural details within erroneous productions of the adult advanced learners as well as source of errors. Using *ICLE-TRCU* corpus, which provides written productions of Turkish learners of English, the researcher analysed 177 essays. The results revealed that L1 had significant impact on errors. In particular, collocational errors were at higher rate when it comes to semi-restricted collocations and aspectual verbs. The learners were observed to have overextend the meaning of the light verbs, which caused unacceptable verb-noun combinations. Üstünalp (2013) aimed to investigate errors and error sources particularly in verb-noun collocations. Collecting 434 essays written by ELT department students, the researcher obtained a learner corpus. Then, the collocations were compared to BNC corpus, and referring to the *Oxford Collocation Dictionary* (McIntosh, 2009), students' productions were checked for accuracy. The analysis revealed that verb category and determiners were the most problematic parts. As to the source of errors, L1 interference was the most prominent one.

Nişancı (2014), as well, addressed to three major sources of errors (semantic transparency, restriction of the elements in word combinations and L1 transfer) in his study. Collecting 76 essays written by 11th grade Turkish learners of English, the researcher detected and categorized the erroneous productions of the learners. The results showed that collocations which were highly restricted and less transparent in meaning caused more problems, while L1 transfer had a poor effect on the erroneous collocations. Bağcı (2014) compared the learners' use of collocations across pre-intermediate and advanced proficiency levels considering both receptive and productive aspects. Data were collected through acceptability judgement and gap filling tests. The findings indicated that advanced level students performed significantly better than the pre-intermediate ones both in receptive and productive sections, which implies that proficiency level is an

important determinant in collocation knowledge. Also, while there was no difference in pre-intermediate students' use of lexical and grammatical collocations, advanced students were better in lexical collocations.

Demir (2016) investigated collocations, lexical hedges, and lexical boosters at the same time by comparing 100 academic articles of Turkish and the same number of articles by Anglophonic writers. The articles were all in the field of ELT. The analysis of the texts showed that Anglophonic writers showed more use of frequency and variety in their collocations, while Turkish writers overused certain collocations without showing variety, which is thought to decrease the native likeness. Finally, the researcher proposed a list of boosting and hedging collocations extracted from the articles of Anglophonic writers. More recently, Aktürk (2020) attempted to reveal the psycholinguistic reasons behind learners' non-native-like use of language. Basically, 71 participants were flashed some words on computer screens in a very short time such as collocations, non-collocations, fillers and non-words. Participants were asked to press certain buttons signalling their idea on the category of each item on the screen. Their reaction time and accuracy were tested based on a certain measurement. The results yielded that students' reaction times to collocation and non-collocations did not differ significantly. This result implied that learners could not prime collocations as a single unit in their mental lexicon.

Despite the dominance of collocation teaching and error analysis in the literature, it is still possible to come across one sample of thesis focusing on genre comparison in Turkish context. Ördem (2013) explored the similarities and differences across various disciplines -*health, physical and social sciences*- in terms of verb-noun collocation use. Adopting corpus-based approach, a total of 249 articles were sampled. As a result of analysis, the researcher detected 165 verbs used commonly across three disciplines. It was also observed that while health and physical science shared significant similarities in terms of collocations, pieces of works in social sciences followed a different route.

Considering the relevant literature in Turkey, one can see that most of the studies are centred around the effectiveness of collocation teaching, contrasting across corpora and error analysis. The current study is an attempt to integrate all these aspects by contrasting a learner corpus to a native corpus, highlighting the grammatical and semantic (dis)similarities and finally carrying out an in-depth error analysis. It is hoped that this study would be a promising one revealing some more elaborated implications thanks to its particular research procedure which will be detailed in the next chapter.

CHAPTER 3

3. METHODOLOGY

This chapter outlines the research design, the corpora used in the study and the research procedure.

3.1. Research Design

Collocations, as a part of formulaic language, have been given assiduous attention in corpus-based language learning studies in recent years (Gablasova, Brezina and McEnery, 2017) Considering the scope of the study, the present study, as well, was designed as a corpus-based analysis since it is a feasible way for describing and explaining variations and use in linguistic patterns. A corpus-based research does not aim to discover new linguistic features, rather it aims to discover how pre-recognized linguistic features govern the systematic patterns of use (Biber, 2010). The present study descriptively presented how collocations, as a recognized linguistic feature, appear in authentic use of learners and native speakers. A Chi-square test was used to find whether the learners overused or underused the target combinations. A Chi-square test is a statistical test used to interpret the relationship between two categorical variables and to determine whether a difference between expected data and observed data is due to chance or the relation between the variables (Field, 2005, p.682). Descriptive analysis such as frequencies and percentages were also used to explain the grammatical and semantic differences between two corpora. Also, deviant productions in the learner corpus were explained in an interpretative way considering the pre-organized classifications.

3.2. Research Corpora

Two different corpora were used for the study. The first one is the learner corpus which contains compilation of essays by Turkish learners of English at intermediate level. The reference corpus is a native corpus containing essays written by native speakers of English. In fact, both corpora are compilations of learners' essays written as a response to certain clue words in an argumentative way.

In corpus studies, corpus representativeness has been an important aspect of the linguistic studies (Biber, 1993; Biber, Conrad and Reppen, 1998; Hundt, Nesselhauf and

Biewer, 2007). Size and composition of corpora are two considerations for the representativeness in comparing corpora. Accordingly, "size" indicates the condition that each corpus in comparison should be large enough to represent the distribution of linguistic features accurately. "Composition" refers to the condition that each corpus must be sampled deliberately to represent the registers in use (Biber, 2010). In short, while size refers to the number of words in the corpora, composition refers to the compatibility of the corpora in terms of genre, dialect, history or individual speaker.

Although representativeness is the key point in corpus studies, it should be noted that no matter how big it is, no corpus can function more than a minuscule sample of a language. Therefore, the question is how many words are needed for descriptive adequacy. While studying rare linguistic structures needs larger corpora to examine, a smaller corpus is enough for high-frequency words (Kennedy, 2014; 66). Also, Koya (2006) reports that high-frequency collocations are common across two native corpora regardless of the topic. In fact, when learner data is concerned, the corpus sizes are typically smaller. For example, Nesselhauf (2005) used the GeCLE (German Corpus of Learner English), comprising 154,191 words, in her study. Laufer and Waldman (2011), also, used a 291,049 words of ILCoWE corpus, and they stated that it was a large corpus comparing to other learner corpora in the field (Laufer and Waldman, 2011). Finally, according to Davies (2015), although large corpora are favoured in most cases, accuracy of annotation is big challenge for the researchers. In other words, assigning the correct part of speech (word level annotation) and classifying texts according to certain criteria (document level annotation) get more likely to be mistaken when the corpora are larger than enough.

Taking the above literature into consideration, an optimum representativeness of the corpora was intended for the current study. To achieve this, several factors were considered such as the similarity of the essay topics and similarity of the corpora sizes. More detailed information for each corpus is given below.

3.2.1. The learner corpus

The learner corpus examined was a compilation of argumentative essays written by the first-year university students studying in ELT department at a state university in Turkey between the years 2009 and 2019. The students were those successfully completing preparatory English year prior to their first year at their departments. Based

on their end-term exam scores, the students were of intermediate proficiency level in English. A total of 166 essays, composed of 150,404 words, were included in the study. Although the clues words of the essays were various, the themes were education, technology and social life, health, media, and art. They were extracted from the Turnitin system in which the learners submitted their tasks as a part of their academic writing courses through the years. The ethical permission procedure was followed in this step of the study (see Appendix-3).

3.2.2. The native corpus

For comparison purpose, the study made use of Louvain Corpus of Native English Essays (LOCNESS). The corpus was compiled by Granger (1998). The corpus is composed of 149,574 words of argumentative essays written by American university students, 18,826 words of literary-mixed essays written by American university students, 95,695 words of argumentative and literary essays written by British university students, and 60,209 words of British A-level argumentative essays. Overall, the corpus contains 324,304 words involving essays written by British and American university students. Only the argumentative essays written by native American university students were used in the present study so that the essay types and the number of words were compatible with the learner corpus. A total of 176 essays, composed of 149,574 words, were included in the study. A great majority of the essays were written in 1995. All of the essays were written by young adult NSs of English except for 4 students. However, their primary language of communication at home is English. The essays are the argumentative responses to the various clue words. Some of the clue words are as follows:

- gender roles in our society
- water pollution
- a man / woman's financial reward should be commensurate with their contribution to the society in which they live
- cheating in colleges
- great inventions and discoveries of 20th century and their impact on people's lives

Although the clue words show a wide range of variety, majority of the essay topics were education, technology and social life, health, media, art, sports, environment, politics and monetary issues.

A detailed information about both corpora are given below (see Table 3.1). Also, one sample essay from each corpus is provided in the Appendices part (see Appendix-1a and 1b).

Table 3.1. *The corpora used in the study*

	Learner Corpus	**LOCNESS (native)**
Contributors	Turkish university students at intermediate level of proficiency	American university students who are native speakers of English
Essay Genre	Argumentative	Argumentative
Number of Essays	166	176
Total Corpus Size	150,404 words	149,574 words
Average Essay Length	906 words	850 words
Topics	Education Technology and social life Health Media Art	Education Technology and social life Health Media Art Sports Environment Politics and monetary issues

3.3. Procedure

The study began with finding concordance lines with lemmas "*make"* and "*do"* with a following *noun* or *preposition/determiner +noun* combination (see Appendix-2 for the illustration of the procedural steps). All three types of combinations (free combinations, collocations, idioms) were included in the study regardless of the restriction level. Therefore, the term "*combination"* is used to refer to the examined verb+noun structures throughout the study. Also, irregular combinations of *make*+noun and *do*+noun such as passive structures were also included. The concordances were scrutinized and the lines which involved irrelevant cases such as "*make*" in a causative case and "*do"* in an auxiliary or emphatic case were eliminated. Also, verb+noun combinations with prepositions (as in phrasal verbs) and indefinite pronouns were excluded while verb+noun combination with determiners or modifiers were included. An illustration of the included and excluded verb+noun combinations in the current study is provided below (see Table 3.2).

Table 3.2. *Criteria for selecting make/do combinations*

	Included		Excluded
"*do*"+noun combinations	✓ "*do*" in combinations regardless of its restriction level (*free combinations, collocations, idioms*). ✓ "*do*" in both active and passive voice. ✓ "*do*" in all lemmatized forms (*do, does, doing, did, done*) ✓ "do" in combination with a pre-modified noun. *E.g: She did a great job yesterday.*	×	"*do*" as an auxiliary verb. *E.g: Do you speak Spanish?* × "*do*" as an emphatic verb. *E.g: I do like Maths*. × "*do*" in phrasal verbs. *E.g: She cannot do away with her problems.* × "*do*" in a combination with indefinite pronouns. *E.g: He cannot do anything now.*
"*make*"+noun combinations	✓ "*make*" in combinations regardless of its restriction level (*free combinations, collocations, idioms*). ✓ "*make*" in both active and passive voice. ✓ "*make*" in all lemmatized forms (*makes, makes, making, made*) ✓ "make" in combination with a pre-modified noun *E.g: She made a lovely cake yesterday.*	×	"*make*" in a causative structure. *E.g: That made me feel good*. × "*make*" as a component of a phrasal verb. *E.g: You can easily make up for missing classes.* × "*make*" in a combination with indefinite pronouns. *E.g: He can make everything he needs by watching videos.*

The AntConc concordancer program was used to extract "*make"* and "*do"* combinations within these criteria. It was used due to its functionality which are suitable for the study objectives. The AntConc is a freeware corpus analysis toolkit for concordancing and text analysis which is designed by Laurence Anthony (L.Anthony, 2019). It has basic functions such as wordlist, keyword, concordance, collocate, clusters and N-Grams. Using the concordance tab, all the occurrences of *do* and *make* were detected. Then, the output for each verb was saved as a text file. And for the next step, the text files were copied and pasted into an MS Word files. All the occurrences were reviewed considering the criteria mentioned above. Hence, some irrelevant cases were eliminated. Finally, a final list of *do*+noun combinations and *make*+noun combinations

were reached. The same procedure was applied to both the learner corpus and native corpus. An illustration of these steps was provided in the appendices (see Appendix 2).

When it comes to the data analysis, the final lists of the occurrences were worked on considering each research question. In order to give an overall idea, some basic descriptive analyses were given before moving to the research questions. Later, the research questions were examined. The research questions are restated below as a reminder.

1. What are the grammatical patterns in "*make*" and "*do*" verb+noun combinations produced by the learners and native speakers?
2. Among various dictionary definitions of the verbs "*make*" and "*do*", which meanings are considered by the learners and native speakers?
3. What are the possible error spots, error types and resorted strategies in the deviant "*make*" and "*do*" verb+noun combinations in the learner corpus?

As stated above, in order to present a general view, each occurrence of aforementioned *make/do* combinations was extracted in both corpora and they were compared in numbers through descriptive statistics by using a statistical analysis software package. This analysis aimed to determine whether the learners overuse or underuse these combinations, regardless of their syntactic and semantic properties.

For RQ-1, learner corpus was analysed syntactically. For this purpose, Hiltunen's (1999) grammatical patterns for high frequency verb-noun combinations were used. The patterning instructed by Hiltunen (1999) was followed in the present study, due to its sole concentration on high-frequency verbs. This patterning was also used by Lareo (2009). In accordance with the framework, the extracted verb-noun combinations were categorized into the corresponding patterns (P) below:

P1) Verb + a / an + (Modifier / s) + Noun

This patterning is the most common one in the Modern English. It gives the advantage of modifying the noun, through which the noun becomes a part of group verb and it is, then, easier to manipulate on it. *e.g., change* vs *make a dramatic change.* However, the syntax of the whole sentence is also considered while deciding what elements should intervene between the verb and the noun.

P2) Verb + (Modifier / s) + Noun

Zero article is the characteristics of this patterning. It is older than the P1 historically. It still survives today in verb + noun structures when the noun is abstract, uncountable or a generic name.

P3) Verb + the + (Modifier / s) + Noun

This patterning signals that the noun is contextually specific – e.g., *He did the experiment many times*- or grammatically constraint –e.g., *He has made the biggest mistake*-.

P4) Verb + (the) + (Modifier / s) + Nounplural

This patterning seems like a combination of the previous two patterns (P2-P3). The basic distinction is that noun is always plural with or without a definite article or a modifier (Hiltunen, 1999).

Grammatical typology of the verb+ noun structures has been used to make comparisons among various periods in the history of the English language or to analyse and compare different written genre in English (Hiltunen, 1999; Koskenniemi, 1977; Lareo, 2009; Nickel, 1968; Visser, 1963). However, the current study takes this analysis as another way to express the similarity/ dissimilarity between the native and leaner corpora, keeping the above-mentioned variables (historical period and genre) stable. This analysis aimed to reveal the structural differences among the native and non-native data. The erroneous productions of the learners were not excluded in the figures since their erroneous productions might still give an idea about the trend in grammatical patterning in their interlanguage.

For RQ-2, the online version of *Oxford Learner's Dictionary, Cambridge Dictionary* and *Macmillan Dictionary* were used. The definitions given in the dictionaries were synthesized by the researcher and one other language expert, who is an experienced English Language instructor. Thus, one single meaning categorization was created. The final categorization is presented in the table below (see Table 3.3).

Table 3.3. *Definitions of "do" and "make"*

No	Definitions of "do"	Definitions of "make"
1	to perform an action, activity, or job	to create/produce something
2	to clean something, or to make a place tidy	to cause to be formed by breaking, cutting, or tearing an object or by pushing one object into or through another
3	to have a good or harmful effect	someone performs the action referred to by the noun usually in fixed phrases
4	to study a subject	to arrange something
5	to spend an amount of time doing something	to earn/get money
6	to provide a service or product for customers to buy	to give the result of a mathematical calculation
7	to make something	to cause something to be successful
8	to move a particular distance or at a particular speed	to have right qualities for something/ to achieve something by reaching a necessary standard
9	to copy someone's voice, manner, or way of moving, in order to entertain people	to reach a place- to be able to be present at a particular event
10	to cheat someone	
11	to use illegal drugs	
12	to apprehend, arrest	
13	to visit a famous place as a tourist	
14	with some adjectives	

Various definitions of "*make*" and "*do"* were analysed and both native corpus and learner data were matched with the corresponding definitions given in the categories. This analysis is supposed to reveal whether more common, in the sense of simpler and more frequent, uses of "*make*" and "*do"* are used by the learners comparing to native speakers as argued by Hugon (2008) and Lennon (1996).

For RQ-3, all the extracted *do*+noun and *make*+noun combinations were examined considering their semantic and grammatical properties. In order to define a combination unacceptable, all the concordance lines were first examined carefully by the researcher and another language expert who is an experienced English instructor whose L1 is Turkish. Then, the problematic occurrences were checked up in the online versions of main stream dictionaries, namely *Cambridge Dictionary* and *Macmillan Dictionary*, and in order to be sure about the collocational errors *Online Oxford Collocation Dictionary* and *The BBI Dictionary of English Word Combinations* (Benson et al., 1990) were resorted. Afterall, there was still a need for a native speaker's judgement. Therefore, a list

of problematic combinations and the full text file of learners' essays were sent to a native speaker of English, an American, teaching English in the U.S. After some negotiations, a final list of unacceptable combinations were reached. Both dictionary entries and native speaker judgement were needed because sometimes there are contradictions among native speakers on the judgement of the same linguistic unit, which implies the need for resorting the dictionaries and native judgements as a two-step verification (Juknevičienė, 2008; Nesselhauf, 2005)

The deviant uses of word combinations and their possible reasons were investigated considering the approaches adopted by Blum and Levenston (1978), Farghal and Obiedat (1995) and Nesselhauf (2005). In her study, Nesselhauf (2005) investigated the deviant verb-noun collocation productions considering verb, noun, determiner and other more global aspects. Blum and Levenston (1978) and Farghal and Obiedat (1995), on the other hand, focused more on the strategies employed by the non-native speakers in their problematic productions of formulaic language units. However, since the scope of the present study is limited to *make* and *do* verb+noun combinations, and an attempt to analyse the error types and strategies used, the categories suggested in previous studies were tailored accordingly. For example, Nesselhauf's (2005) categories were taken not as the error types but the *error spots*, which imply what part of the combination is problematic (noun, verb…etc.).The term *error types* in this study imply what type of knowledge is missing (lexical, collocational …etc.). Also, the strategies highlighted by Blum and Levenston (1978) and Farghal and Obiedat (1995) are not completely possible to apply here since the current study deals with combinations, not non-combinations. Therefore, "*avoidance*", the strategy used by non-native speakers when a target combination cannot be produced, was ignored here. A final template of the error analysis is presented below (see Table 3.4).

Table 3.4. *The template of error analysis*

Error Spots	**Error Types**	**Strategies**
Noun	Lexical	L1 translation
Verb	Collocational	Synonymy
Determiner/Modifier	Grammatical	Paraphrasing
Whole Combination Error	Spelling	Not Observable

The current study is an attempt to handle the deviant uses in a more qualitative manner. Therefore, the deviant occurrences of the learners were scrutinized in details. As stated above, each deviant occurrence was examined considering three dimensions, *error spot*, *error type* and *strategy*. It should be noted that there were some guidelines about these dimensions to keep in mind during the analysis:

1. A typical combination is made up of a verb, a determiner or modifier and a noun; no determiner or modifier is also possible as well (Hiltunen, 1999).

2. The problem might be with one or more elements of a combination. Thus, some overlaps among the error spots are possible.

3. Although an erroneous combination in a sentence is possible to make acceptable by simply changing one or more elements of the combination; the sentence might be still unacceptable due to the combination preferred. Thus, in some cases, the whole combination might be problematic considering the context. The sentence might need another combination or no combination at all. In such cases, *whole combination inappropriate* code was assigned. The approach is adopted from Nesselhauf (2005).

4. *Error types* refers to the source of the errors. Considering the references (non-native judgement, native judgement and grammar resources), an error might stem from more than one origin. Therefore, it is possible to observe some overlaps among the error types.

5. Categories of *strategies* are fundamentally adopted from Blum and Levenston (1978) and Farghal and Obiedat (1995). Since asking students back about their essays is not possible in this type of corpus analysis (Nesselhauf, 2003), the interpretation of the strategies were determined by resorting to non-native judgement (L1 Turkish), native judgement (L1 English) and grammar and academic writing resources.

6. In the analysis of *strategies*, there was not a clear strategy observable in some cases. They were coded as "not applicable, (NA)"

This analysis is supposed to reveal the errors in learner corpus in a qualitative manner so that we can deeply understand the challenges with high-frequency verbs in verb+noun combinations.

All in all, the current study is supposed to shed light on the issues such as overuse/underuse, grammatical patterns, semantic transparency, restrictedness, and errors in association with the strategies employed by the learners. The results of the analysis are presented in the next section.

CHAPTER 4

4. RESULTS AND DISCUSSION

4.1. Results

In this chapter, the findings of analyses are provided considering the order of the research questions in the study. However, before moving to the research questions, occurrences of aforementioned *make/do* combinations are presented in both corpora regardless of their syntactic and semantic properties. They are presented through descriptive statistics to present the overall picture especially in terms of numbers (see Table 4.1 and Table 4.2)

Table 4.1. *Distribution of lemma "make" across the corpora*

	make	**%**	**makes**	**%**	**made**	**%**	**making**	**%**	**TOTAL**	**%**
LOCNESS	137	49.6	14	5.0	88	31.8	37	13.4	276	100
LEARNER	100	58.8	10	5.8	37	21.7	23	13.5	170	100

Table 4.2. *Distribution of lemma "do" across the corpora*

	do	**%**	**does**	**%**	**did**	**%**	**doing**	**%**	**done**	**%**	**TOTAL**	**%**
LOCNESS	37	37.3	8	8.0	13	13.1	16	16.1	25	25.2	99	100
LEARNER	51	50.0	4	3.9	2	1.9	15	14.7	30	29.4	102	100

In order to understand whether the difference in numbers across two corpora are significant, a 2X2 Pearson Chi-square test was conducted. Pearson's Chi-square test is of large interest in research studies due to its robustness and simplicity (Benhamou and Melot, 2018). Since the sizes of corpora are not exactly the same (LOCNESS 149574; Learner 150404), they were standardized to 250,000 words. A similar standardization methodology was followed by Laufer and Waldman (2011). Both raw numbers and relative frequencies for both corpora are provided below (see Table 4.3).

Table 4.3. *Standardized frequency distribution of "make" and "do" across the corpora*

	LOCNESS		LEARNER	
	raw freq.	relative freq.	raw freq.	relative freq.
make	276	461	170	283
do	99	165	102	170

Note: *Relative frequencies per 250,000 words are shown in the table.*

Results indicated that the number of *make*+noun combinations in native corpus was significantly higher than the those in learner data ($X^2(1) = 42.650$, $p < .001$). As for the number of *do*+noun combinations, although it seems that the number is slightly higher in the learner data, this difference is not statistically significant (X^2 (1) = .075, $p = .785$).

4.1.1. Grammatical patterns of "*make*" and "*do*" verb+noun combinations produced by the learners and native speakers.

The first research question aimed to find out whether the type of *make/do* collocations differs across the native corpus and learner data. In order to find out the distribution of aforementioned combinations, each *make*+noun and *do*+noun combination in both corpora was tagged according to the patterns suggested by Hiltunen (1999). The categorization is as follows;

Pattern 1: Verb + a / an + (Modifier / s) + Noun

Pattern-2: Verb + (Modifier / s) + Noun

Pattern-3: Verb + the + (Modifier / s) + Noun

Pattern-4: Verb + (the) + (Modifier / s) + Noun*plural*

The distribution of the *make*+noun and *do*+noun combinations according to the patterns (Ps) is tabulated for each corpus (see Table 4.4 and Table 4.5).

Table 4.4. *Grammatical patterns in LOCNESS*

	P1	**%**	**P2**	**%**	**P3**	**%**	**P4**	**%**	**TOTAL**	**%**
make	79	28.6	80	28.9	32	11.5	85	30.7	276	100
do	18	18.1	48	48.4	13	13.1	20	20.2	99	100

Table 4.4 shows that P4 (30.7%) is the most frequent *make*+noun pattern in the native corpus and it is followed by P2 (28.9 %) and P1(28.6%). On the other hand, P3

(11.5%) is the rarest syntactic pattern found in the native data. As for *do*+noun combinations, P2 (48.4 %) dominates the *do*+noun combinations used in the data. P4 (20.2 %) is the second highest pattern and it is followed by P1 (18.1 %). Finally, P3 (13.1%) becomes the rarest *do*+noun pattern followed in the native corpus.

Table 4.5. *Grammatical patterns in the learner corpus*

	P1	**%**	**P2**	**%**	**P3**	**%**	**P4**	**%**	**TOTAL**	**%**
make	31	18.2	67	39.4	6	3.5	66	38.8	170	100
do	12	11.7	42	42.1	7	6.8	40	39.2	102	100

According to the Table 4.5, in terms of *mak*e+noun combinations, the learner data were dominated by P2 and P4 (39,4%; 38,8%). P3 (3,5%) is followed at a very low frequency in *make*+noun combinations in the learner data. As for *do*+noun combinations, a similar trend is observed here. P2 and P4 (42,1%; 39,2%) are the most frequently used patterns in *do*+noun combinations while P3 (5,8%) is, again, the rarest used pattern in *do*+noun combinations in the learner data.

Consequently, the results revealed that learners followed P2 and P4 more frequently in both *do*+noun and *make*+noun combinations. On the other hand, they used P1 and P3 less frequently again in both *do*+noun and *make*+noun combinations. As for natives, comparing their within group *do*+noun and *make*+noun combinations, they seemed to have followed divergent patterns. In *do*+noun combinations, they used P2 at the highest and P3 at the lowest frequency. In *make*+noun combinations, native students followed a fairer distribution among the patterns. They used P4, P2, and P1 at a similar rate. Yet, they still underused P3. The overall results showed that P2 was the pattern observed at the highest frequency in all four cases (both *do*+noun and *make*+noun combinations in both corpora), which is a commonality between two corpora. However, learners did not make a discrimination in their use of *do*+noun and *make*+noun combinations regarding the grammatical patterns they followed. They followed a similar grammatical patterning in their productions regardless of *do*+noun or *make*+noun combination. However, the findings in the LOCNESS corpus revealed that the native students had a tendency of adjusting the grammatical patterning considering the high-frequency verb in the combination.

4.1.2. Semantic distribution of "*make*" and "*do*" verb+noun combinations produced by the learners and native speakers

The second research question aimed to find out which meanings of *make* and *do* are considered among various dictionary definitions. The analysis here only focuses on the *make* and *do* in combinations with a noun. In order to use as the reference, *Oxford Learner's Dictionary, Cambridge Dictionary* and *Macmillan Dictionary* were examined due to their wide use among language learners. Considering the criteria of selecting *make/do*+noun combinations at the beginning of the analysis, irrelevant meanings such as causatives or phrasal verb structures were eliminated. Consulting with a language expert, who is an experienced English teacher and an academic at a university, given definitions of *make* and *do* were put into categories. Since it seemed more practical to assign numbers to each individual definition category (D) and explanations were clearer, the style of *Macmillan Dictionary* was followed by taking the other dictionaries into consideration, as well. The final meaning categorization for *make* and *do* is presented below (see Table 4.6 and Table 4.7).

Table 4.6. *Meaning categories for "make"*

No	Definition	Example
D1	to create/produce something	*Let's make coffee.*
D2	to cause to be formed by breaking, cutting, or tearing an object or by pushing one object into or through another	*The rain made a hole on the ground.*
D3	someone performs the action referred to by the noun usually in fixed phrases	*We couldn't make a progress yesterday.*
D4	to arrange something	*I want to make an appointment*
D5	to earn/get money	*She makes 75 dollars a day.*
D6	to give the result of a mathematical calculation	*Five and two makes seven*
D7	to cause something to be successful	*His songs made the show.*
D8	to have right qualities for something/ to achieve something by reaching a necessary standard	*A good story makes a good film*
D9	to reach a place- to be able to be present at a particular event	*We cannot make the conference hall on time.*

Table 4.7. *Meaning categories for "do"*

No	Definition	Example
D1	to perform an action, activity, or job	*I do karate at the weekends.*
D2	to clean something, or to make a place tidy	*I will do the bedroom after lunch.*
D3	to have a good or harmful effect	*Sunlight will do good for your body.*
D4	to study a subject	*She is doing chemistry and biology.*
D5	to spend an amount of time doing something	*I did three years in New York*
D6	to provide a service or product for customers to buy	*We do sandwiches for parties.*
D7	to make something	*The paintings were done by him.*
D8	to move a particular distance or at a particular speed	*They did 500 km last night*
D9	to copy someone's voice, manner, or way of moving, in order to entertain people	*He did Michael Jackson at the party.*
D10	to cheat someone	*You paid £50 for this? You have been done!*
D11	to use illegal drugs	*She doesn't do drugs like other guys.*
D12	to apprehend, arrest	*He was done for shooting a guy at the bar.*
D13	to visit a famous place as a tourist	*We can go back to hotel after we do the museum.*
D14	with some adjectives	*I always advised her, but she did the opposite.*

Taking given definitions in Table 4.6, the distribution of the meaning categories within the extracted *make*+noun combinations are given in numbers and percentages below (see Table 4.8).

Table 4.8. *The distribution of the meaning categories across corpora for "make"*

No	Definition	LOCNESS		LEARNER	
		N	**%**	**N**	**%**
D1	to create/produce something	32	11.5	26	15,3
D2	to cause to be formed by breaking, cutting, or tearing an object or by pushing one object into or through another	2	0,7	0	0,0
D3	someone performs the action referred to by the noun usually in fixed phrases	182	65.9	129	75,9
D4	to arrange something	3	1.0	1	0,6
D5	to earn/get money	47	17.0	14	8,2
D7	to cause something to be successful	2	0.7	0	0,0
D8	to have right qualities for something/ to achieve something by reaching a necessary standard	8	2.8	0	0,0
D9	to reach a place- to be able to be present at a particular event	0	0,0	0	0,0
	TOTAL	**276**	**100**	**170**	**100**

According to the Table 4.8, there are some common points to be highlighted among the native and learner corpora. Firstly, it is observed that a great majority of *make*+noun combinations in both the native and learner corpora are used in the meaning D3 (65,9 % and 75,9 %, respectively), which is defined as "*someone performs the action referred to by the noun usually in fixed phrases*". Finally, the meaning D6 and D9 are not referred in *make*+noun combinations either in native or learner data (0,0%).

The table also reveals some discrepancies across two corpora in terms of the semantic aspect of the *make*+noun combinations. Firstly, the category D1, which is defined as *"to create/produce something",* comprises 11,5% of all the *make*+noun combinations produced by the native students whereas the figure is 15,3% when it comes to the learner corpus. Secondly, the meaning D5, which is defined as "*to earn/get money*" makes up the 17,0% of all occurrences of *make*+noun combinations in the native corpus, whereas the same meaning category comprises only the 8,2% of the *make*+noun combinations in the learner corpus, which is less than a half of the figure in the native corpus. Thirdly, the table shows that there are only two categories -*D6 and D9*- not observed in native corpus while there are five categories not observed -*D2, D6, D7 and D9*-. Finally, although the occurrences of D3 and D4, which correspond to the delexical "make" as defined in Altenberg and Granger (2001), seem to be similar in percentages, it is clear in the frequencies that native students used them more often than the learners (182; 129 respectively). These differences between the native and learner productions seem to be in accordance with the idea that more superficial meanings appear in learners productions (Hugon, 2008; Lennon, 1996).

Ultimately, it can be concluded that both the natives and learners refer to the same meaning category (D3) as the primary meaning of "*make*". Nevertheless, they still differ in the frequency count. Also, the secondary meanings referred to "*make*" are divergent. Learners tend to use "*make*" in D1 as the secondary meaning attributed to it, while the secondary meaning attributed to "*make*" is the one defined in D5 for the native students. Also, the learners do not present variety of meaning categories in their essays as much as native students do.

Taking given definitions in Table 4.7, the distribution of the meaning categories within the extracted *make*+noun combinations are given in numbers and percentages below (see Table 4.9).

Table 4.9. *The distribution of the meaning categories across corpora for "do"*

No	Definition	LOCNESS		LEARNER	
		N	%	N	%
D1	perform an action, activity, or job	65	65.6	96	94.1
D2	to clean something, or to make a place tidy	0	0.0	0	0.0
D3	to have a good or harmful effect	9	9.0	1	0.9
D4	to study a subject	0	0.0	0	0.0
D5	to spend an amount of time doing something	0	0.0	0	0.0
D6	to provide a service or product for customers to buy	1	1.0	0	0.0
D7	make something	4	4.0	2	1.9
D8	move a particular distance or at a particular speed	0	0.0	0	0.0
D9	to copy someone's voice, manner, or way of moving, in order to entertain people	0	0.0	0	0.0
D10	to cheat someone	1	1.0	0	0.0
D11	to use illegal drugs	0	0.0	0	0.0
D12	apprehend, arrest	0	0.0	0	0.0
D13	to visit a famous place as a tourist	0	0.0	0	0.0
D14	with some adjectives	19	19.1	3	2.9
	TOTAL	**99**	**100**	**102**	**100**

Table 4.9 shows some similarities among the native and learner corpora in terms of the meaning referred in *do*+ noun combinations. First of all, it is seen that both the native students and the learners use "*do*" in D1, which is defined as "*perform an action, activity, or job*", at the highest frequency comparing to the other meanings (65,6% and 94,1%, respectively). Secondly, the category numbers D2, D4, D5, D6, D7, D8, D9, D10, D11, D12 and D13 are referred a few or no times by both the native students and learners (4,0 % - 0,0%).

On the other hand, the table reveals a difference between two corpora. D3 and D14, which are defined as "*to have a good or harmful effect*" and "*with some adjectives"* are referred only a few times by the learners (0,9%; 2.9%, respectively). However, those categories are mentioned a lot more in native students' essays (9%; 19.1%, respectively). This difference between the native and learner productions, as well, seems to be in accordance with Hugon (2008) and Lennon (1996).

Overall, it can be concluded that both native students and learners take D1 as the primary meaning of "*do*" in their *do*+noun combinations. In fact, this category dominates over all other possible meanings in both corpora. On the other hand, D3 and D14 reveal

a difference between two corpora. Native students consider those categories as the secondary meanings of "do", while the same categories are hardly ever referred in learners' essays.

4.1.3. Error analysis of "*make*" and "*do*" verb+noun combinations produced by the learners

The third research question aimed to find out the possible sources of deviant "*make*" and "*do*" verb-noun combinations in the learner corpus. To this end, as detailed in the methodology section, the deviant productions of the NNSs were determined by resorting to a Turkish instructor of English, certain dictionaries and a native speaker. Finally, a list of deviant combinations were reached. The deviant uses were examined considering three aspects, *error spot, error type* and *strategy used*. It is worth reminding that some of the errors overlap among the categories, *i.e:* a combination is erroneous both grammatically and collocationally, or there are mistakes in verb and determiner within the same combination. An overall illustration of the errors and percentages of both *do*+noun and *make*+noun combination*s* are shown below (see Table 4.10)

Table 4.10. *Overall distribution of the deviant uses in the learner corpus*

	Total Occurrences	**Deviant Occurrences**	**%**
do+noun	102	17	16.6
make+noun	170	28	16.4
TOTAL	272	45	16.5

Firstly, Table 4.10 shows that the learners produced a total of 102 *do*+noun combinations. Of all these combinations, 17 productions were deviant, which makes up 16,6 % of all *do*+noun combinations. Secondly, the learners produced a total of 170 *make*+noun combinations. Of all these combinations, 28 productions were deviant, which makes up 16,4 % percent of all *make*+noun combinations. When the deviant *do*+noun combinations and *make*+noun combinations were compared quantitatively, the percentages were similar. (16,6 % and 16,4 %, respectively). However, the errors needed to be elaborated for the *do*+noun combinations and *make*+noun combinations individually in terms of *error spot, error type* and *strategy used*.

The detailed categorization of the errors in *do*+noun combinations are given below (see Table 4.11).

Table 4.11. *The detailed categorization of the errors in do+noun combinations*

Error Spots	Verb	Noun	Det/Mod	Whole Comb.
	11	3	1	4
Error Types	Lexical	Grammatical	Spelling	Collocational
	2	3	0	13
Strategies	L1 Translation	Synonymy	Paraphrasing	NA
	5	9	3	0

According to Table 4.11, when the *error spots* are concerned, 11 of the combinations were unacceptable due to problems with *verbs*, which accounts for the great majority of the deviant uses locationally. There were also 4 occurrences in which *whole combination* was inappropriate and 3 occurences in which *nouns* were not acceptable. Finally, there was 1 occurrence in which *determiner or modifier* was problematic, which was also problematic in terms of its verb, as well. In order to illustrate the analysis more clearly, one sample occurrence from the learner data for each error spot is provided below (see Table 4.12).

Table 4.12. *Sample occurrences for the error spots in do+noun combinations*

Sample Case	**Error Spot**	**Suggested Correction**
one of the most important mistakes we do while teaching English is that … .	Verb	*… mistakes we make*
government knows the events before done	Whole comb.	*Government knows the events before they happen*
if you do not do illegal works, ….	Noun	*… do illegal things*
it is unnecessary to do such change	Det/Mod	*... make such a change*

As for the *error types*, Table 4.11 shows that collocational errors were responsible for the greatest portion of the errors with 13 occurrences. There were also 3 occurrences of grammatical error and 2 occurrences of lexical error. Finally, there was no spelling mistake in the *do*+noun combinations of the learners. One sample occurrence from the learner data for each error type is provided below in order to make the points more clear (see Table 4.13).

Table 4.13. *Sample occurrences for the error types in do+noun combinations*

Sample Case	Error Type	Suggested Correction
they believe that while doing material, they waste time in vain	Collocational	*… designing a material*
It has been done some researches	Grammatical	*Some research has been done*
they do their works that they suppose to do with internet	Lexical	*… do their things*
-	Spelling	-

As the third dimension, the study addressed the *strategies* employed by the learners. Table 4.11 reveals that synonymy was the most resorted strategy with 9 times of occurrence. There were 5 cases of L1 translation and 3 cases of paraphrasing. Finally, there was no utterance for which a strategy could not be assigned. One sample occurrence from the learner data for each strategy used by the learners was illustrated below (see Table 4.14).

Table 4.14. *Sample occurrences for the strategies in do+noun combinations*

Sample Case	Strategy	Explanation/ Correction
We had to paint, do music and learn role-playing	Synonymy	"*do*" was used instead of "*make*"
They can not study the next day lessons and do homeworks (ödevleri yapmak)	L1 translation	"*homework*" is a singular noun in English, but both ways are possible in Turkish. (*ödev-ödevler*)
It can be deterrent to stop people from doing criminal events	Paraphrasing	*Commiting crimes*
-	NA	-

The following table shows the detailed categorization of the errors in *make*+noun combinations (see Table 4.15).

Table 4.15. *The detailed categorization of the errors in make+noun combinations*

Error Spots	Verb	Noun	Det/Mod	Whole Comb.
	17	2	4	6
Error Types	Lexical	Grammatical	Spelling	Collocational
	6	9	1	17
Strategies	L1 Translation	Synonymy	Paraphrasing	NA
	16	7	0	4

According to the Table 4.15, when the *error spots* are concerned, there were 17 combinations unacceptable due to the *verbs* in the combinations. As in the *do*+noun combinations, *verbs* were the parts of the combinations where the number of errors were the highest. There were 6 occurrences of *whole combination* errors and 4 occurrences of *determiner/modifier* error. The number of errors in *nouns* was only 2. In order to illustrate the analysis more clearly, one sample occurrence from the learner data for each error spot is provided below (see Table 4.16).

Table 4.16. *Sample occurrences for the error spots in make+noun combinations*

Sample Case	**Error Spot**	**Suggested Correction**
while trying to escape, she made an accident and died	Verb	*... she had an accident...*
Many inventors, before inventing, they made imagination.	Whole comb.	*.... they imagined.*
…it made great impact to our social life	Det/Mod	*…it made a great impact…*
one has been enabled to make video callings through the world	Noun	*…make video calls …*

When the second dimension, *error types*, was regarded, Table 4.15 shows that *collocational* errors were responsible for the greatest portion of the errors with 17 occurrences. There were also 9 occurrences of *grammatical* error and 6 occurrences of *lexical* error. Finally, there was 1 *spelling* mistake in the *make*+noun combinations of the learners. One sample occurrence from the learner data for each error type is provided in the table below to better illustrate the error types in the data (see Table 4.17).

Table 4.17. *Sample occurrences for the error types in make+noun combinations*

Sample Case	**Error Type**	**Suggested Correction**
Exams should be made in formal classrooms	Collocational	*Exams should be taken ….*
They find their old friends by looking other people's information and make contact them	Grammatical	*….. making contact with them*
World revolves, time flies, human changes and make change everything on the earth.	Lexical	*… changes everything*
… therefore, they make no personal efford to help the environment	Spelling	*effort*

The current study also addressed the *strategies* used by the learners. Table 4.15 reveals that *L1 translation* was the most resorted strategy with 16 times of occurrence. There were 7 cases of *synonymy*. The strategy of *paraphrasing* was not observed in this

part of the analysis. However, there was 4 occurrences for which any strategy could not be assigned (*NA*). One sample occurrence from the learner data for each strategy used by the learners is provided in the table below (Table 4.18)

Table 4.18. *Sample occurrences for the strategies in make+noun combinations*

Sample Case	Strategy	Explanation/ Correction
To raise a healthy children, their vaccinations should be made regularly	L1 translation	It is "*aşı yapmak*" in Turkish, which entails "*make*" as the equivalent for "*yapmak*". "*Given*" should replace for "*made*" as the correction.
make strong relationships with real individuals not with virtual and fake ones	Synonymy	"*build strong relationships …*" *make* was used for *build* as a perceived alternative.
Moreover, Facebook makes antisocial person.	NA	*"makes a person antisocial"* could be suggested as a correction. No strategy was observed here.
-	Paraphrasing	-

In conclusion, these results indicate that there are some commonalities observed in both *do*+noun and *make*+noun combinations, and some differences, as well.

One shared result revealed here is that a great majority of the errors occurred in verb part the combinations in both *do*+noun and *make*+noun combinations. Secondly, considering both *do*+noun and *make*+noun combinations, the errors were majorly collocational. As a final commonality, grammatical problems were observed more frequently than the lexical ones.

Do+noun and *make*+noun combinations differed in terms of the strategies used by the learners. The most distinctive one is that the learners used synonymy as the primary resort dealing with *do*+noun combinations and L1 translation as the secondary strategy. However, when it comes to *make*+noun combinations, the learners were observed to follow the opposite direction. They used L1 translation as the primary strategy and synonymy as the second resort.

4.2. Discussion

4.2.1. Summary of the findings

The results section revealed that the number of *make+noun* combinations were significantly higher in native corpus in comparison with the learner data. As for the *do+noun* combinations, the frequency was higher in the learner data than the native corpus. However, this difference was not a significant one.

The first research question aimed to determine the grammatical structures underlying the verb+noun combinations. It was observed that learners and native students showed both similarities and dissimilarities in terms of their verb+noun combinations' grammatical patterns in their written essays. That the P2 was the most frequent pattern observed in both corpora and both combination type (*do*+noun and *make*+noun) was a commonality between two corpora. On the other hand, the results also yielded that learners followed the patterns at a similar proportion regardless of the combination type while the native students followed a relatively different patterning regarding the combination type.

The second research question was about the dictionary meanings of the *make+noun* and *do+noun* combinations. The results yielded both similarities and dissimilarities among the native and learner data. When *make+noun* combination is concerned, in the majority of the cases, the verb "*make*" is used in the meaning *someone performs the action referred to by the noun usually in fixed phrases*. However, when the other meaning categories are concerned, it was clear that the learners do not present variety of meaning categories in their essays as much as native students do. As for do+noun combinations, the verb "*do*" was majorly used in the meaning *to perform an action, activity, or job* in both the learner and the native data. However, native corpus seemed to present more variety of meaning categories comparing to the learner data.

The third research question addressed the deviant *do*+noun and *make*+noun combinations produced by the learners. As detailed in results section, of all the components of verb+noun combinations (verb, determiner/modifier, noun), the most problematic part for the learners was the *verb*. When the error types are considered, the collocational errors were the most frequent type of errors in the learner data. However, students' errors regarding *do*+noun and *make*+noun combinations differed in terms of the strategies they employed in their unacceptable productions. While the learners used synonymy as the primary way of tackling with the problematic *do*+noun combinations, they used L1 translation as their primary resort in *make*+noun combinations.

4.2.2. Discussion of the findings

The current study is intended to provide an overall picture of the two high-frequency verbs (*make* and *do*) in verb+noun combinations. Before moving to the research questions, the issue of overuse/ underuse was examined. The results revealed

that learners underused "*make*+noun" combinations significantly, and they slightly overused "*do*+noun" combinations, which was not significant. The results corroborate the findings of a great deal of the previous works (Borgatti, 2006; Gilquin, 2007; Hugon, 2008; Juknevičienė, 2008; Kaszubski, 2000).

Borgatti (2006) examined two L1s (Dutch and French) at the same time and again the data revealed that "*make*" was underused in both learner data in comparison with the native data. Gilquin (2007) also reported that French learners of English underused "*make*" collocations in their writing but they still overused the ones which have direct equivalents in French. Hugon (2008) also reported a significant underuse of "*make*" in the written learner data (French L1) noting that the same type of difference was not signifciant in the spoken data. In her study, Juknevičienė (2008) compared the written productions of Lithuanian learners of English and native English speakers. She examined some other high-frequency verbs in collocations like *have, give* and *take* besides *make* and *do*. The frequency of "*do*" collocations were slightly higher, yet not significant, in the learner corpus while the frequency of "*make*" collocations was signficantly lower in the learner corpus. Kaszubski (2000) analyzed some high-frequency verbs in a learner and a native corpora. The learner corpora (Polish L1) involved more than one proficiency levels. The results revealed that "*make*" was used at a significantly higher rate in the native corpus, and the frequency decreased as the proficiency level decreased. Interestingly, for the other high-frequency verbs, the case was just the opposite.

The current study might discuss some of the ideas mentioned above. Firstly, as the learners were from different L1 background in the above mentioned studies (Polish, Dutch, French and Lithuanian), and they all followed a similar route in their written productions, this consistency implies that L1 is not a significant factor in the use of high-frequency verbs. The current study provides a further evidence for this idea. Secondly, as implied in Gilquin (2007), the learners' tendency to avoid making a mistake might explain their underuse of "*make*" collocations to some extent. However, since the current study also examined "*do*" collocations, as well, and the results showed no significant difference between the learners and the natives, it is clear that avoiding a mistake cannot account for underuse in this case. Therefore, the current study is in line with Gilquin (2007) with regard to "*make*". However, it also provides a counter evidence for the idea of avoidance when it comes to "*do*" combinations. Thirdly, Hugon (2008) made a distinction between written and spoken data, and reported an underuse in the written data but no significant

difference in the spoken data. It was further stated that native speakers also used less high-frequency verbs in conversation than in academic writing, but learners cannot calibrate their written and spoken productions. Therefore, learners write as they speak. Juknevičienė (2008), which is highly in accordance with the current study, also ascribes the underuse of "*make*" to lack of academic writing skills of the learners. The current study, being an academic writing analysis, provides a partial evidence for this argument, since it does not contain any spoken data. Finally, Kaszubski (2000)'s findings revealed that the frequency of "*make*" increased as the proficiency level increased, and it was significantly the highest in the two native corpora used in the study. However, for the other high-frequency verbs, the trend was just the opposite. The current study supports this finding by providing the cases of "*make*" and "*do*" collocations in a single proficiency level, though.

On the other hand, the results of this study are not consistent with some other studies (Babanoğlu, 2014; Hasselgren, 1994; Ringbom, 1998; Shaw, 2001). Babanoğlu (2014) compared two learner corpora to a native corpus in terms of the frequency of "*make*" in various lexical and grammatical patterning, and when the overall frequency of "*make*" is regarded, the study did not reveal any significant difference among the learner corpora and the native corpus. However, Babanoğlu (2014) used Altenberg and Granger (2001)'s categorization. The categorization suggested by Altenberg and Granger (2001) covers the causative and phrasal verb uses, as well. The current study, on the other hand, focused on *make*+noun and *do*+noun combinations excluding phrasal verb structures and causatives. This tendency might explain the origin of the different findings between these two studies. The same reason might be responsible for the great portion of the divergence divergence between the current study and that of Shaw (2001) and Ringbom (1998). Additionally, Shaw (2001) excluded erroneous productions in his study unlike the current study. To conclude, the current study revealed that learners do not stick to lexical "teddy bears" (high-frequency verbs) as argued by Hasselgren (1994).

4.2.2.1. Grammatical patterns of "make" and "do" verb+noun combinations produced by the learners and native speakers.

The current study used the grammatical patterning summarized by Hiltunen (1999) in order to show the similarity or dissimilarity among the native and learner corpora in

terms of *make*+noun and *do*+noun combinations. The patterning used in the study is given below as a reminder:

Pattern 1: Verb + a / an + (Modifier / s) + Noun

Pattern-2: Verb + (Modifier / s) + Noun

Pattern-3: Verb + the + (Modifier / s) + Noun

Pattern-4: Verb + (the) + (Modifier / s) + Noun*plural*

As detailed in the above section, the most frequent pattern observed in both corpora (native and learner) and both combination type (*do*+noun and *make*+noun) was P2. On the other hand, it was also evidenced that the native students followed a relatively different patterning regarding the combination types whereas learners were stick to similar patterns (P2 and P4) in their combinations regardless of the combination type.

It should be noted that many other studies on the grammatical or semantic aspect of high-frequency verbs (Altenberg and Granger, 2001; Babanoğlu, 2014; Kim, 2015; Laporte, 2012; Lin, 2019) mostly used the categorization suggested by Altenberg and Granger (2001), so that they could present a single picture depicting the grammatical and semantic properties at once. The current study, within the framework of the first research question, focused on the grammatical aspect individually in order to find a more robust answer. Therefore, Hiltunen's (1999) categorization of the grammatical patterns was used unlike many other studies. In essence, the categorization used in the current study has already been used to identify the English language across genres and historical periods (Hiltunen, 1999; Koskenniemi, 1977; Lareo, 2009; Nickel, 1968; Visser, 1963). Learner language was not addressed in those studies. Thus, the current literature does not provide any previous study using Hiltunen's (1999) categorization comparing the native English and learner language in terms of collocations or high-frequency verbs. Only that of Lareo (2009) might be compared to the current one regarding its concentration on the verb+noun collocations. Comparing a Maths corpus to a fiction corpus, Lareo (2009) reported that P2 was the most or one of the most frequently used pattern in both corpora, which is in line with the results of the current study. When the learner corpus is taken into consideration in isolation, one can infer that the learners produced *do* and *make* combinations which were very similar to the native speakers in terms of grammatical patterning.

Interestingly, Hiltunen (1999) argues that P1 is the most common pattern today, but both the current study findings and that of Lareo (2009) yielded conflicting results with

this argument. One possible explanation for this conflict might be that the genres analysed in the above-mentioned studies were different from each other. The current study made use of the argumentative essays written by university students in various topics and Lareo (2009) used fiction writings and science (Mathematics) articles. This diversity of the genres might explain the differing grammatical patterns in the corpora. At the same time, it might also imply that grammatical patterning is not governed by genre of the writing.

Taking only the learner data into consideration, the discussion might be further elaborated by recalling the specifications of P2. As stated above, P2 is a pattern in which one can use a noun without any definite or indefinite article. In fact, it is only possible for generic, abstract or non-countable nouns in English. Emphasising on the abstract nouns, Hiltunen (1999) explains this as an effect of French language on English in the course of time. Considering that the learners' L1 was Turkish, this might bring forward the issue of L1 effect on learners' verb+noun combinations. In Turkish, a verb+noun combination is possible with zero article as well as the accusative case (definite article "the") and indefinite article case. However, it is not acceptable in English.

E.g.:	*Ali bir kek yaptı*	*Ali made a cake*
	Ali keki yaptı	*Ali made the cake*
	Ali kek yaptı	**Ali made cake*

As can be seen in the above examples, Turkish learners of English might have produced some erroneous *make/do* combinations without any definite (*a/an*) or indefinite article (*the*) since it is acceptable in their L1. Thus, the frequency of P2 -*Verb + (Modifier / s) + Noun*- might have increased (Üstünalp, 2013). The mismatch between the languages might account for the use of P2 pattern more than other patterns. This point will be reconsidered in the Research Question-3, which deals with the errors in the learner corpus, in this study.

4.2.2.2. Semantic distribution of "make" and "do" verb+noun combinations produced by the learners and native speakers

The study findings revealed that both natives and learners assigned one common primary definition category (D) for each of *do* and *make* in their essays. For *do*+noun combinations, this common ground for both corpora was D1 - *perform an action, activity, or job.* This meaning category can be regarded as the core meaning of the verb "*do*". As for *make*+noun combinations, the highest frequency was observed in D3 - *someone*

performs the action referred to by the noun usually in fixed phrases- in both native and learner corpora. Although native and learner corpora yielded seemingly similar results, the results are more remarkable in terms of differences. Firstly, it was observed that these common primary Ds were used relatively at higher percentages in the learner corpora and D1 and D3 (for *do* and *make* respectively) were so frequent in the learner corpus that the other meaning categories were hardly considered, which decreased the diversity in their productions. Native corpora, on the other hand, showed more diversity in terms of definitions attributed to *do* and *make* in their combinational uses. Secondly, while D1 for *do*+noun combinations can be regarded as the core meaning of the verb *do*, D3 for *make*+noun combinations is not the core meaning of the verb *make* but is a delexical (*e.g., make a judgement*) definition of it. Comparing the percentages of *make* productions within groups, one might, deceptively, argue that the learners showed more examples of delexical *make* in their writings. However, when the frequency counts are reconsidered, the figures show that native speaker used higher number of *make*+noun combinations either in total or D3 in particular. It means that native students showed more examples of delexical *make* in their writings while presenting diversity at the same time. Finally, some senses of the high-frequency verbs were used at a relatively lower frequency in the learner corpus. For example, D3 and D14 for *do*, and D5 and D8 for *make* were rare in the learner corpus in comparison with the native corpus.

The results are partially or completely in line with some of the significant studies comparing the learner language and the native language in this regard (Allerton, 1984; Altenberg and Granger, 2001; Cobb, 2003; Hugon, 2008; Lennon, 1996; Sinclair, 1991). Although detecting the pitfalls of the learner language and making generalizations over the differences from the native English is not focused majorly, the current study still provided some evidence for the differences between the learner language system and native language within overlapping issues. In an effort to discover the semantic diversity produced by the learners in high-frequency verbs, the current study revealed that although "*make*" and "*do*" are typically learned at the beginning of the EFL instruction, the learners (mixture of pre-intermediate and intermediate proficiency) seemed to fail in presenting various dictionary meanings attributed to these verbs in their academic writing. In a more global perspective, Cobb (2003) stated that even advanced learners of English have difficulty in discovering the full phrasicon in English, and they tend to repeat the same phrases whereas the native speakers implicitly know it, and thus they show more diversity

in their language use. The current study presented evidence for this statement in terms of semantic diversity. Also, Sinclair (1991; 79) argued that learners avoid using common words and instead "they rely on larger, rarer, and clumsier words which make their language sound stilted and awkward". This tendency was particularly observable in two instances in the current study. In the case of "*do*", the learners used D3 -*to have a good or harmful effect*- only once, however it was possible to find D3 9 times in the native corpus. Moreover, the learners unacceptably used the lemma forms of "*give harm*" instead of "*do harm, cause harm* or *harm*". Doing this several times, the learners decreased their D3 frequency while making their productions awkward. As for "*make*", learners seemed to have preferred "*earn*" as an alternative to "*make*" in the sense of D5 –*to earn/get money*-. Although the same interchange was observed in the native corpus (7 occurrences), the learners used that alternative more often (11 occurrences). Therefore, it can be said that Sinclair's argument was confirmed in the current study.

A similar claim was made by Lennon (1996). It was claimed that although learners have a broad idea of verb meaning, they have a more limited knowledge of some other important aspects, such as polysemy, semantic boundaries or collocational restrictions. Thus, their productions are mostly based on the core meaning of the verbs, though the verbs "*make*" and "*do*" have broader boundaries and collocational possibilities. As exemplified above, the current study shows that although learners are well familiar with the words "*make*" and "*money*" separately, they still prefer "*earn money*" at a higher frequency but "*make money*" at a lower frequency in comparison with the native corpus. A similar underuse of *make* in the sense of *make money* was also found in some other studies (Altenberg and Granger, 2001; Babanoğlu, 2014; Hugon, 2008). Accordingly, it can be concluded that learners do not prefer *make* as an alternative to *earn* in the sense of *make money*. As stated by Lennon (1996), the problem is not just a verb-choice error, in fact "*earn money*" in the current study is not erroneous at all. The essence of the problem is that learners have a great tendency of sticking to the core verb meanings and they are unable to extend their knowledge to the delexical usage of a verb. Although high-frequency verbs allow for various uses, learners still feel restricted to the core meanings of them. Nevertheless, it should be noted that Shaw (2001) reported contrary findings in this regard. Comparing the writings of Chinese learners of English and a combination of two native sub-corpora, the researcher evidenced that the learner corpus contained higher number of *make* in the sense of *make money* than the native corpora. In the current study

and above mentioned concordant studies dealt with Turkish, Japanese, Swedish and French learners of English, while Shaw (2001) examined Chinese learners in his research. Although, there is not an explanation for this contrast in Shaw (2001), the difference among the findings might imply an L1 effect or the language instruction.

As cited in Altenberg and Granger (2001), Allerton (1984) argues that although there are grammatical, syntactic and semantic restrictions which have already been defined, it is still worth considering some sort of arbitrariness in the selection of some words such as high-frequency verbs. According to Altenberg and Granger (2001), learners might be aware of this arbitrariness, and thus they avoid using semantically unmotivated high-frequency verbs particularly when a high-frequency verb does not match with its L1 equivalent. The case can be exemplified with D8- *to have right qualities for something/ to achieve something by reaching a necessary standard* - of the verb *make*. This meaning category was not observed at all in the learner corpus while there were 8 occurrences of it in the native corpus. One plausible explanation of this divergence might be the fact that this sense of *make* in English does not make a similar sense in the L1 of the learners, Turkish. A similar motivation can account for some other differences across the corpora, such as underuse of make in the sense of D5 – *make money*- or *do* in the sense of D3 – *do good/bad*-. Therefore, it can be concluded that the current study confirmed the argument (Allerton, 1984) and explanation (Altenberg and Granger, 2001) by previous studies.

4.2.2.3. Error analysis of "make" and "do" verb+noun combinations produced by the learners

The results section revealed that the percentages of the deviant *do*+noun and *make*+noun combinations in the learner corpus were almost equal (16,6 %; 16,4%, respectively). According to Kurtböke (1998), bilingual dictionaries are away from providing target delexical meanings to the Turkish equivalents (most potentially *yapmak* or *etmek*); thus learners are supposed to find the best option considering the context on their own. Mostly due to the Persian and Arabic words, it is even difficult to find a constant match within Turkish. For example, *tamir (repair)* is used with *etmek* while the plural form *tamirat* is used with *yapmak.* (Kurtböke, 1998). Therefore, it can be concluded that "*make*" or "*do*" might not make individual delexical senses to the Turkish learners of English, which makes them potentially and equally error prone. Juknevičienė (2008)

also provided evidence for this in her study while Altenberg and Granger (2001) found evidence both for and against this malformed L1 matching.

Another finding revealed in the study was that total deviant uses of *do*+noun and *make*+noun comprised 16,5% of the total productions of the learners, who were at intermediate proficiency. Contrary to the current study findings, Nesselhauf (2005) examined the advanced German learners of English in her study and 1/3 of the total combinations were deviant. Also, Bıçkı (2012) found more than 75 % of the collocations problematic in a corpus of Turkish learners of English at advanced proficiency. Remembering that the current study examined intermediate level learners' productions, one can conclude that the current study yielded highly contradictory results. However, it should be noted that above mentioned two studies were not restricted to *make* and *do* combinations only, rather they took all kinds of verb+noun combinations into account. Hugon (2008) focused on *make* collocations only and found 17,2% of the collocations problematic, which seems in accordance with the current study findings. The discrepancy among the findings might be ascribed to the nature of high-frequency verbs. High-frequency verbs such as *make* and *do* are learnt at the very beginning of the instruction and thus they are perceived relatively more usable and safe for the students (Hasselgren, 1994).

The current study examined the deviant uses in three dimensions. They are *error spots*, *error types* and *strategies*. Discussing the study findings with relevant literature, it should be noted that the high-frequency verbs analysed across studies, the categorization of the errors or the terminology used do not entirely overlap across the studies mentioned here, so some sort of adjustment and recalculation should be taken into account.

The first dimension of the errors was the analysis of the *error spots.* The analysis revealed that in both *do*+noun and *make* noun combinations, it was verbs where the great majority of the deviant uses observed (nearly 58,3%). This result confirmed some previous studies (Altenberg and Granger, 2001; Bıçkı, 2012; Boers, Demecheleer, Coxhead and Webb 2014; Nesselhauf, 2003; Üstünalp, 2013). Nesselhauf (2003) explains this by referring her definition of collocation in which the restrictedness on the verb choice is emphasized. Bıçkı (2012) supported this argument in his study by stating the fact that learners overextended the verb meanings especially in the light verbs. Üstünalp (2013) also reported 68,88% of verb errors and the problem was ascribed to learners' ignorance of restrictedness. Although there is not an error analysis in their study,

Altenberg and Granger (2001) emphasized the treacherous side of high-frequency verbs by taking *make*+noun collocations as an example. Their study drew attention to the fact that although high-frequency verbs usually share a semantic common ground across the languages, they still develop some language specific characteristics. Boers et al. (2014) reinforced this argument by stating that when a high-frequency verb is employed in a verb+noun combination, the noun carries most of the meaning in the combination. Secondly, the verb seems well-familiar to the learners, and they concentrate more on the noun element of the combination. Therefore, they often fail producing the correct verb. Considering all the explanations above, it can be argued that learners might be unaware of the restrictions with high-frequency verbs, *do* and *make*, and they might tend to apply an L1 motivated semantic framework in their mind onto the target language ignoring the language specific features with the high-frequency verbs.

Boers et al. (2014) can also be confirmed by looking at the errors with nouns in the current study. The current study revealed that only 10,4 % of the errors were related with *noun*s, which is equal to the *determiner/modifier* errors and just a half of *whole combination* errors. Thus, it can be concluded that learners might have paid more attention to the nouns in the combinations. Üstünalp (2013) contrasted the determiner errors in his study to the Nesselhauf (2003) and Nesselhauf (2005). It was hypothesized that since English has more commonalities with German language in terms of determiner usage than that of Turkish, the Turkish learners of English might be more error prone comparing to the German learners. When the ranking of the error spots are concerned, the hypothesis might be approved because determiner were the second most frequent error spot in Üstünalp (2013) and the current study while it is the fifth and sixth in Nesselhauf (2003) and Nesselhauf (2005), respectively. However, when the percentages of the determiner errors within the total erroneous productions, the hypothesis is not confirmed because determiners account for the 16,9 % in Nesselhauf (2003), 13,8% in Üstünalp (2013) and 10,4 % in the current study. Therefore, in a closer statistical view, these inconsistent results can only tell us that the distribution of error spots are not homogeneous across the error spots in the writings of Turkish learners of English unlike the German counterparts.

Error types is the second dimension examined in the study. The term *error type*, as detailed in methodology section of this paper, refers to what type of knowledge is missing or problematic within the learner corpus, *i.e: grammatical, lexical, spelling or collocational*. The study findings yielded similar results in terms of error types for

do+noun and *make*+noun combinations. *Collocational* errors were the most dominant error type observed in both *do* and *make* combinations. The number of *collocational* errors were even higher than total number of other error types. The frequency of the other types of errors were ranked as *grammatical, lexical* and *spelling* errors in both type of combinations in the learner corpus. Collocational aspect was brought about in many studies previously and the restriction was highlighted as a problematic issue (Bıçkı, 2012; Boers et al., 2014; Henriksen, 2013; Howarth, 1998; Laufer and Waldman, 2011; Lennon, 1996; Nesselhauf, 2003; Nişancı, 2014; Shibliyev, 1993; Yılmaz, 2004). Howarth (1998) stated that when the restriction in combination is relatively low as in free combinations or relatively high as in idioms, the learners can deal with it. However, when the restriction is unpredictable as in collocations, the learners face the greatest challenge because they need to be aware of the collocability of the items in their combinations. Nesselhauf (2003) and Laufer and Waldman (2011) also evidenced this argument with German and Hebrew learners of English. Similar results were found in some other studies with Turkish learners, as well (Bartan, 2019; Nişancı, 2014; Shibliyev, 1993; Yılmaz, 2004). The current study is in line with these studies in terms of collocational errors. Nevertheless, the current study does not distinguish among free combinations, collocations or idioms. Thus, it is not possible to observe error types considering the restriction levels, however it can be clearly seen that collocational errors caused a big problem to the learners.

Grammatical errors ranked the second in the learner corpus. Especially with *make*+noun combinations, the problem with the grammatical aspect was much higher. It should be noted that most of the grammatical errors are caused by the deviant use of determiners and/or resulted in whole collocation inappropriate. This problem was observed in previous studies even though the learner corpora in those studies were of advanced level learners of English. Considering that the learner corpus used in the current study is of intermediate level learners, the grammatical problems are more understandable. However, it should still be noted that malformed L2 collocations place greater processing burden on the native readers of the language (Millar, 2010). Finally, there were a few *lexical errors* in the learner corpus. Such errors were found in nouns which do not signal a collocational deficiency. According to Lennon (1996), verb choice errors are just the tip of iceberg. In fact, wrong verbs in a combination stems from lack of semantic boundaries and collocational restrictions. Therefore, a more delicate examination was conducted to detect sole lexical errors in the corpus. Comparing *make*

and *do* combinations in terms of grammatical and lexical errors, the learner corpus contained such errors more in *make*+noun combinations than in *do*+noun combinations. It can be concluded that *make* is more problematic than *do* in combinations with a noun. In fact, Kaszubski (2000) highlighted the speciality of "make" previously. Among six high-frequency verbs, including make, it was observed that higher proficient learners used make more frequently and accurately in their writings whereas for the other five verbs the trend was just the opposite. Kaszubski explains this by referring to the higher percentage of restricted patterns with make. The current study confirmed this unique aspect of make.

Strategies was the third dimension examined in the current study. When *do* and *make* combinations are considered together, the most frequently used strategy was L1 transfer, followed by Synonymy as the second. When they are examined individually, *do* and *make* combinations showed a discrepancy. In *do*+noun combinations, *synonymy* was the most frequently used strategy, whereas L1 translation was used at the highest frequency in *make*+noun combinations. The use of synonymy as a strategy was highlighted in some previous studies (Blum and Levenston, 1978; Farghal and Obiedat, 1995; Lennon, 1996; Nesselhauf, 2005; Yan, 2010; Yılmaz, 2004). According to Blum and Levenston (1978) synonymy can be observed in different ways. One of these ways is that learners use true synonyms disregarding the collocational restrictions, which results in deviant collocations. Farghal and Obiedat, (1995) conducted a translation study with Arabic learners of English. Although the translation task required adjective+noun collocations, the results were similar to the findings of the current study. They found that learners employed synonymy as the most frequent strategy, and it was argued that learners' using this strategy heavily implies their unawareness of restrictions of lexical items in collocations. The same problem was also observed in other studies under different names, as well. For example, it is termed as semantic similarity in Nesselhauf (2005), synonym copying in Yan (2010) and false concept hypothesizing in Yılmaz, (2004). As stated in error spots analysis, Lennon (1996) accounts for such deviant uses as a result of violation of semantic boundaries due to the lack of collocational knowledge. This explanation is in line with that of Farghal and Obiedat (1995) and well applicable to the current study finding.

As for *make*+noun combinations, *L1 transfer* was the most commonly used strategy followed by synonymy. When only make combinations are regarded, L1 transfer strategy

was observed in the 59 % of all erroneous productions in the current study. The overall occurrences of L1 based errors (*do* and *make* combinations together) was 36%. It can be concluded that make+noun combinations forced students to take L1 transfer more frequently in *make*+noun combinations. The literature provides several findings showing L1 effect on learners' interlanguage (Altenberg and Granger, 2001; Bartan, 2019; Bıçkı, 2012; Blum and Levenston, 1978; Gilquin, 2007; Gyllstad and Wolter, 2016; Juknevičienė, 2008; Laufer and Waldman, 2011; Nesselhauf, 2003, 2005; Nişancı, 2014; Shibliyev, 1993; Yan, 2010; Zhou, 2016). Laufer and Waldman (2011) reported half of the deviant productions, of Hebrew learners of English at three different proficiency levels, stemmed from L1 effect. Nesselhauf (2003) made an analysis of overall verb-noun combinations, and she found a 45 % of L1 effect in a corpus of German learners of English at advanced proficiency level. Another general verb-noun combination analysis was carried out by Nişancı (2014) and the participants were Turkish learners of English. The findings yielded only 23,3 % of L1 effect. In a translation task study, Bartan (2019) detected that 47 % of the errors, of Turkish translators of English, were L1 motivated. It seems that L1 based errors vary at a great extent among the studies and one can argue that learners' native language is not a significant factor in this case. Moreover, the current study made the range of L1 effect even greater by *make* combinations particularly. It can be suggested that "*make*" is a marginal verb in this perspective and thus requires a closer analysis. Another explanation for this relatively high influence of L1 could be learners' being intermediate level unlike many of the other studies in the field. Although the current study does not make a comparison among proficiency levels, comparing other study findings and the current study findings, it can be argued that lower level of proficiency results in more prone to L1 influence.

There are also some more cautious approaches towards L1 transfer strategy. Firstly, although the majority of the related literature provides evidence for L1 influence in some way, there are still studies showing little or no effect of L1 such as Leśniewska and Witalisz (2007) and Yılmaz (2004). Leśniewska and Witalisz (2007) conducted an acceptability test to advanced Polish learners and found no influence of L1. The result was ascribed to the advanced proficiency level of the learners. Yılmaz (2004) also detected a little influence of L1 in her translation task study. The explanation for this result was the idea that learners were of advanced proficiency level; therefore, they were more aware of the influence of L1, and instead of giving an unacceptable response, the

students preferred not to answer tricky questions. This argument also implies that L1 influence can be controlled by an advanced learner. In fact, together with that of Blum and Levenston (1978), this idea was adopted in the current study methodologically. Thus, L1 influence was taken as a strategy (not a source of error) which is used (or not used deliberately) by the learners. Wang and Shaw (2008) supported this argument by stating that learners can have a say on the effect of L1 by considering the perceived similarities/dissimilarities between their L1 and TL. Moreover, L1 influence is not necessarily negative. Both positive and negative transfer are possible (Gilquin, 2007; Zhou, 2016).

The current study, first of all, cannot clearly show whether the learners used L1 transfer strategy consciously or not since it is not possible technically (Nesselhauf, 2003). However, considering that the learners were intermediate level learners, unlike the participants in Yılmaz (2004), and they had too many errors especially in make combinations in the current study, one can assume that they relied on their L1 more than enough. It might be so due to the fact *make* is a high-frequency verb and thus learners did not consider the collocational aspect of it carefully although they had the access to resources (dictionary, internet, reference books…etc.) and no time pressure. Taking Wang and Shaw (2008) into consideration, it might even be argued that they underestimated "*make*" and they did not perceive much difference between Turkish and English in that sense, which implies that L1-L2 congruency was delusional to the learners. Previous literature evidenced that the learners have difficulty in dealing with collocations which are incongruent with their L1 (Gyllstad and Wolter, 2016; Nesselhauf, 2005; Yamashita and Jiang, 2010). However, taking the current study findings on one hand and Yılmaz (2004) on the other hand, one can observe that the same L1 (Turkish)-L2 (English) matching revealed conflicting results. It leads us to consider perceived congruency rather than congruency alone. Therefore, it can be concluded that an L1, itself, cannot be the sole responsible for the erroneous productions. The perception of the congruency by the learners are more important. Since the participants in Yılmaz (2004) are advanced learners and the learner corpus used in the current study is of intermediate proficiency level, it can be argued that proficiency level is a factor on this case (Bağcı, 2014). That is to say, the perception of L1-L2 congruency grows more mature as the proficiency level increases. This argument is also in line with the findings in Wang and

Shaw (2008), which suggested that the factors within L2 are as important as interlingual ones.

CHAPTER 5

5. CONCLUSION

5.1. Summary of the Findings

The current study has five main chapters. The first chapter, *Introduction*, includes a background to the study, the problem statement, aim and the significance of the study and, finally, introduces the research questions. The second chapter, *Literature Review*, starts with fundamental issues and discussions in collocation and then gives an idea of how collocations are dealt in instructional settings. The same chapter also outlines some studies on high-frequency verbs and *make/do* collocations including the academic thesis written on collocations at Turkish universities. The third chapter, *Methodology*, elaborates the research design and corpora used in the study. The fourth chapter, *Results and Discussion*, provides the findings regarding the research questions and the results obtained in the study are discussed with references to the relevant literature. Hereby, the fifth chapter, *Conclusion*, summarizes the results of the study. This summary is followed by implications which outlines the pedagogical implications referring to the present study and some other studies in the field. The chapter, and the whole paper as well, ends with limitations of the current study and some suggestions for further research.

The finding of each research question is briefly given below:

1. *What are the grammatical patterns in "make" and "do" verb+noun combinations produced by the learners and native speakers?*

The distribution of the grammatical patterns in two corpora was similar in terms of frequency. However, the variation between *make* and *do* was limited in the learner corpus.

2. *Among various dictionary definitions of the verbs "make" and "do", which meanings are considered by the learners and native speakers?*

The learner corpus showed a smaller scale of variety in terms of meaning categories in comparison with the native corpus.

3. *What are the possible error spots, error types and resorted strategies in the deviant "make" and "do" verb+noun combinations in the learner corpus?*

Errors were mostly on *verbs*. They were majorly *collocational*, and the learners seemed to resort *L1 translation* and *synonymy* as the primary strategies in their written productions.

In detail, the study showed that when *do+noun* combinations are regarded; the frequency was not significantly different between the learner data and the native corpus. However, the number of *make+noun* combinations were significantly lower in the learner corpus in comparison with the native corpus. This underuse is in consistence with some previous studies while it is, at the same time, contradictory to some others. This fundamental outcome made it easier to analyse the following research questions. When grammatical patterns are regarded, although P2 and P4 were the most frequent patterns in both corpora, learners seemed to use almost the same portions of the patterns regardless of their being *make* or *do* combinations. However, the native corpus revealed somewhat varying degrees of frequency considering the difference between *make* and *do* combinations. It implies that learners do not distinguish between *do* and *make* in this regard and they apply the same grammatical patterning not considering these items individually. Also, the high frequency of P2 in the learner corpus could be attributed to use of erroneous zero article nouns under the influence of L1 (Turkish). This point is not made clear in the study, though. When the semantic properties are concerned, the study revealed that although both *do* and *make* have various dictionary meanings, the learner corpus did not show variety in using them in combination with nouns. They tended to stick to core meanings of them. These findings imply that learners are not much aware of the collocational possibilities or extended meanings of *make* and *do*, which make their writing clumsier as argued by Sinclair (1991, p.79). Finally, the study dealt with the deviant uses in the learner corpus. The errors were majorly on the *verb* part of the combinations and *collocational* errors were dominant in both *do* and *make* combinations. The error analysis also focused on the strategies used by the learners. It should be noted that the strategies were distilled from the deviant uses. Therefore, the strategies mentioned in the study were the ones which did not work well. In deviant *do*+noun combinations, the students used *synonymy* as the major strategy, followed by *L1 translation*. In deviant *make*+noun combinations, *L1 translation* was the most dominant strategy observed in the learner corpus. The error analysis, overall, implies that learners have problems with collocational uses of high-frequency verbs, *do* and *make*, and these problems are observable mostly in these verbs rather than nouns, determiners or modifiers of combinations. These findings yielded that learners are not much aware of restrictedness and collocability of the high-frequency verbs examined. Also, learners seemed to be

misguided by their L1 very frequently, which brings forward the issue of congruency and/or perceived congruency.

All in all, the current study detected some problematic areas in the learners' use of high-frequency verbs by focusing on *do* and *make* in noun+verb combinations. It was evidenced that high-frequency verbs are starting block and stumbling block at the same time, as suggested by Hugon (2008). Considering the study findings, it can be concluded that collocations, particularly with high-frequency verbs, stand as a serious problem in language learning. Therefore, a special care should be given to high-frequency verbs and their collocational uses, rather than underestimating them. The following part shares some pedagogical implications from the literature which might be useful in language instruction.

5.2. Implications

The study attempted to reveal the written productions of Turkish learners of English in terms of verb+noun combination with the high-frequency verbs, *do* and *make*. Underuse, grammatical and semantic variation, restrictedness, collocational errors, synonymy and L1 congruency were some of the hot words of the paper. In this chapter, some implications are given considering the current study findings in the light of some previous studies in the field. Since there are various issues to be taken account, the implications will be given under three sub-categories: *approach level implications*, *method and technique level implications* and *implications for textbooks and dictionaries*. According to E.Anthony (1963), an approach is the widest category in which beliefs and assumptions about language learning are determined; a method is the circle in which these assumptions and beliefs are put into practice and content is defined and; finally, a technique is the smallest circle where the classroom procedures and activities are concretely specified.

5.2.1. Approach level implications

Considering the current study findings and the relevant previous studies in the field, one major conclusion to be drawn is that knowing a word involves collocational uses of the word as well its core meaning and grammatical properties attached to it in usage (Thornbury, 2002, p.16). This brings the implication that the teachers, learners and material designers should have the awareness of collocations in the target language

(Babanoğlu, 2014; Bıçkı, 2012; Cobb, 2003; Gilquin, 2007; Nesselhauf, 2004; Wray, 1999). Secondly, high-frequency verbs should not be ignored as they are seemingly easier. This study evidenced that although *make* and *do* are both high-frequency verbs which the learners are well-familiar with, they failed to use them accurately mostly due to the lack of collocational knowledge. In fact, the current study did not deal with the difference between learners' receptive and productive performance but some other studies in the field (Bağcı, 2014; Gitsaki, 1996) showed that receptive knowledge is broader than productive knowledge in many cases. Therefore, since the high-frequency verbs have a treacherous nature, they should be approached in various ways beyond their core meanings in language learning settings. As stated by Lennon (1996), not only quantitative but also qualitative vocabulary gain should be addressed in language teaching. It means instead of continuously teaching new but undigested vocabulary, imperfectly acquired vocabulary items should be better consolidated so that the learners are able to use their pile of vocabulary productively (Cobb, 2003). Another implication can be about the different instructional approaches adopted in different parts of the world as suggested by Hugon (2008). Hugon (2008) argued that Japanese learners of English performed better in written production while Western learners of English use a more speech like language in their writing; and this is attributed to the emphasis on spoken language in Western countries and written language in Japan. Since the language is a composition of written and spoken; and reception and production, the overall implication can be adopting a more comprehensive approach to the language teaching, which encompasses language as a whole.

5.2.2. Method and technique level implications

As stated above, giving more room for collocation teaching is the major implication of the current study. Methodologically, the current study revealed some points to be considered in language teaching, as well. The study yielded underuse of *make* as a high-frequency verb, lack of grammatical and semantic variation, limited collocational knowledge and semantic transparency and negative L1 influence. Considering the above-mentioned approach level implications, the methodologies in ESL/EFL settings should put more emphasis on the literal and figurative meanings of the target vocabulary, not the core meanings only. Especially, extended meanings and collocational uses of the words should be explicitly taught due to the fact that mere exposure has little or no effect in this

respect (Nesselhauf, 2003). Since the learners failed to show productive variation both in grammatical and semantic aspects, the explicit teaching should keep a good balance of form-focused and meaning-focused activities. Also, in collocation teaching, congruency among the native language and the target language should be considered, as well. Some collocations the in the target language can be directly translated into the native language keeping the same semantic property. However, it is not possible for incongruent ones. Thus, they create a big problem for the learners especially in productive sense. Therefore, the teachers should give priority to the incongruent collocations by explicitly contrasting the literal, figurative or register specific meanings in language teaching (Bahns, 1993). The current study evidenced that the learner corpus involved many occurrences of L1 translation strategies which did not work. Finally, academic writing should be given more space in ESL/EFL settings as a medium to teach collocations better. When the oral communication is more emphasized over the written production, there arise certain problems. Since the incorrect use of collocations sound odd in communication but does not spoil the communication necessarily, the learners and teachers are satisfied with conversational language, and thus ignore the language accuracy, which potentially impair the collocational learning (Laufer and Waldman, 2011). As evidenced in other studies, *i.e.*: Hugon (2008), when it comes to production, learners cannot distinguish between written and spoken languages. As stated earlier in the current study, some erroneous productions of the learners might have been resulted from the same problem. Therefore, the problematic collocational structures should be highlighted by negative evidence as a part of academic writing lessons when needed (Murao, 2004).

As for the classroom procedures, there are some practical implications with some caveats. Firstly, it should be remembered that classroom instruction might lead learners to produce grammatically correct but unidiomatic utterances due to the lack of sensitivity of collocational associates (Wray, 1999). Although there are some basic formulaic expressions covered in many course books, collocations with high-frequency verbs have a wider coverage in the language. It takes years for learners to learn these seemingly easy verbs, especially in delexical sense. As evidenced in the current study and many other studies, even upper-intermediate and advanced level learners cannot deal with high-frequency verbs when it comes to collocational uses. Therefore, the teachers should go beyond the course books and try to expose students to more real-life examples of the language. This is highly possible thanks to the concordance software nowadays.

Especially, web-based concordance programs make it possible to observe a particular word with its numerous collocates in various authentic texts and evidently increase the vocabulary development (Akıncı and Yıldız, 2017; Conrad, 1999; Daskalovska, 2015). For example, one high-frequency verb can be picked for a lesson and the most frequent collocates of it can be retrieved using a concordance software and these collocations can be analysed in a phraseological manner dealing with the lexical units in a collocation regarding its restriction level, semantic boundaries and register specific characters. While selecting the target collocations, learners' L1 should be considered and non-congruent ones should be highlighted through various examples. Doing this as a part of language teaching would move the lessons beyond the course book and provide the learners with a really an authentic language. This kind of a teaching can be integrated into the regular vocabulary or academic writing lessons. It would give the learners an awareness of collocation and show that the high-frequency verbs are always with them through the way they learn the language. Another idea about collocation teaching is that the collocations with high-frequency verbs can be distinguishable semantically though it is seemingly impossible. To this end, Lantolf and Tsai (2018) tested the applicability of the SCOBAs (Schema for the Orienting Basis of Action) in teaching the difference between *make* and *do,* and when the deep difference between the two verbs are visualized or materialized, the learner could semantically understand the difference instead memorizing them as chunks. The technique is very innovative especially for adult learners as they are potentially better at conceptualizing the verb meanings. However, it seems not much applicable to the young learners.

Testing students' collocational knowledge is also important since it is closely linked to processing, comprehension and use of language (Almacioğlu, 2018). It should be noted that collocation knowledge was evidenced to be in correlation with both vocabulary score (Mutlu, 2015) and writing score (Hsu, 2007). Hence, testing collocation knowledge might give a novel and better understanding of learners' vocabulary and writing development, which is sometimes not much observable by repeating the same traditional tests in the classroom. Therefore, allocating time and energy for collocation cannot be regarded as an extra burden for teachers or students. One technical way of testing collocational knowledge of the learners can be the CollGram technique suggested by Bestgen and Granger (2014). Basically, each contiguous word pair in a learner's writing text is assigned and compared to a large reference corpus. Each collocation is scored and the

text is given an overall score for the frequency and accuracy of the collocations in it (*see* Bestgen and Granger, 2014 f*or more detail*). This technique promises a good alternative for assuring the text quality, and thus overall writing score. Another thing to be considered in testing or exercising collocations is the matching activities. Although matching exercises are one of the most common way of drilling and testing, it might be risky due to the risk of potentially distractive mismatches. According to Boers et al., (2014), intact wholes of collocations should be used in the exercises instead of partial combinations.

Finally, it should be noted that DDL has gained importance recently. Since it gives the students more opportunity of engaging in the corpus, teachers can make use of DDL model. Hence, both teachers and learners feel more confident in detecting the needed vocabulary, differentiating the nuances among similar vocabulary, eliminating the negative languages transfer and teaching/learning them more permanently as claimed in the literature (Huang, 2014; Lay and Yavuz, 2020; Men, 2020; Otto, 2021; Sun and Hu, 2020). For example, one or a few target words for each lesson can be determined by analysing the concordance lines extracted from a relevant corpus. This determining process can be directly performed by the students as implied in direct DDL activities. Also, the teacher can make the selection beforehand considering the time, physical conditions, students' proficiency levels and their ability to use concordance programs on a computer. Having determined the concordance lines, teachers can show the concordance lines of the target word or combinations on a screen or distribute the printed-out concordance lines to the students. The numbers of the target word or word combinations and concordance lines to be shown can be adjusted by the teacher considering the course objectives and variations in the uses of the target word(s). Especially, teaching of collocations which are incongruent between the students' native language and the target language would be more effective by this way.

5.2.3. Implications for textbooks and dictionaries

Textbooks and dictionaries are also at importance for the collocational development of the learners. They have been usually criticized for not being sufficient or delicate about collocations (Ergül, 2014; Molavi, Koosha and Hosseini, 2014; Peksoy, 2013; Vassiljev, Skopinskaja and Liiv, 2015; Wray, 1999). As stated above, learners' native language should be considered while selecting the target collocations. Mainstream textbooks are, in nature, unable to consider numerous native languages in the world. Since one size-fits-

all approach is not very helpful in collocation teaching, teachers should take more responsibility so that learners are exposed to the exclusively selected collocations rather than random ones. Alternating textbooks with corpus-based collocation teaching seems a very effective technique as stated above. As for dictionaries, Hugon (2008) suggested that they should provide the learners with more contextual information (formal/informal, frequency… etc.) about high-frequency verbs. Although it is labour-extensive for dictionary authors, finding such information in a dictionary, of course, would have benefits for the learners. However, it should be noted that the current study implied that learners did not even looked up in the dictionaries as they were not hesitant using erroneous combinations although they had the opportunity. It can be concluded learners did not even have an awareness of collocations as stated above. Thus, it can be said that raising an awareness comes first. Although textbooks might have a role in raising an awareness of collocations, dictionaries seem to have relatively minor effect.

5.3. Limitations and Suggestions for Further Studies

The present study investigated the high-frequency verbs, *do* and *make* in verb+noun combinations through a corpus-based analysis. The data were obtained from a mini corpus composed of the argumentative essays of Turkish learners of English who were in the first year of English Language Teaching department in state university. Regarding their proficiency exam results, their proficiency levels were intermediate. The reference corpus used in the study was a small portion of LOCNESS, composed of argumentative essays written by American college students.

Although the present study aimed to present a clear picture of the *make/do* verb+noun combinations in the written productions of the Turkish learners of English, there are still some limitations to be considered in this study. One major limitation of the current study was on the proficiency levels of the students who are the contributors of the non-native learner corpus used in the present study. As mentioned above, the students were at intermediate level regarding their base passing scores at the end-term exams of English preparation class in previous year. However, it should be noted that since the learner data were compiled from year 2009 to 2019, the students contributing to the learner corpus in the present study might not be very homogenous in terms of proficiency level because the exam format changed a few times in this period and thus the student profile, too, might have changed in the course of time. Therefore, the proficiency level of

intermediate might be mixed with some higher achievers such as upper-intermediates or advanced students, though they are very limited in number. Secondly, the time period in which the students (both native and non-native) wrote the essays were not the same. Majority of the native essays were written in 1995 but as for the non-native learners' essays, they were written between the years 2009 and 2019. Although, the present study focused on *make*/*do* verb+noun combinations, which are high-frequency verbs and highly possible to occur in all kinds of texts, it might be still considered as another limitation of the current study. Finally, it should be also considered that both corpora used in this study are compiled in classroom settings. Both the native and non-native essays were written as a response to pre-set clue words. In this kind of essays, the writers (either native or non-native students) might be worried about the writing performance rather than arguing their sincere point. Thus, it might influence the frequency or accuracy of the target word combinations. Therefore, this can be taken as another limitation of the study.

This study focused on one single proficiency level. A further study can focus on more than one proficiency level such as intermediate and advanced levels. Comparing two learner corpora among each other and to one native reference corpus might yield important results on the developmental factors in collocations with high-frequency verbs. Alternatively, the current learner group might be asked to write argumentative essays parallel to the current ones in terms of representativeness again in their fourth year at the department and their development in terms of collocations with high-frequency verbs can be observed in a longitudinal way.

The present study considered *make* and *do*, only due to the potential problem they create for Turkish learners. A further study can focus on other high-frequency verbs such as *take, have, get*… etc. The high-frequency verbs can be studied all together, individually or selectively based on certain criteria. Also, the current study considered only verb+noun combinations regardless of their restriction level. A further study can focus on other types of combinations such as adjective+noun combinations. Even, a certain restriction level can be focused such as *idioms, collocations* and *free combinations*.

As evidenced in some previous study findings, learners, unlike native speakers, cannot calibrate between written and spoken language and thus they write as they speak, and they speak as they write. The current study could partly contribute to this previous argument since it did not provide any spoken data. A further study can add spoken data

along with written data; and thus, it becomes possible to see whether learners' written and spoken data reflect the same performance in terms of collocation use.

In the error analysis part, the current study assumed the possible strategies used by the learners in each of their production. Although reference books and native and non-native judgments were resorted to define the learners' strategy in each case, it was never possible to ascertain it by asking back to each learner for each case. In fact, it is the nature of corpus-based studies. Therefore, a qualitative further study can be designed by participation of limited number of definite learners so that their strategies in mind can be explored by asking back to the individual student in an interview or using think-aloud protocols.

REFERENCES

Aisenstadt, E. (1979). Collocability restrictions in dictionaries. *International Journal of Applied Linguistics, 45*(1), 71–74.

Akıncı, M. (2009). *Effectiveness of corpus consultance in teaching verb-noun collocations to first-year ELT students*. Unpublished MA Thesis. İstanbul: Boğaziçi University, Institute for Graduate Studies in the Social Sciences.

Akıncı, M., & Yıldız, S. (2017). Effectiveness of corpus consultation in teaching verb + noun collocations to advanced ELT students. *Eurasian Journal of Applied Linguistics, 3*(1), 91–109.

Aktürk, A. (2020). *Investigation of collocational priming in tertiary level Turkish EFL learners' mental lexicon*. Unpublished MA Thesis. Trabzon: Karadeniz Technical University, The Institiute of Social Sciences.

Allerton, D. J. (1984). Three (or four) levels of word cooccurence restriction. *Lingua, 63*(1), 17–40. https://doi.org/10.1016/0024-3841(84)90030-5

Almacioğlu, G. (2018). Examining Turkish ESL learners ' receptive collocational knowledge. *Gaziantep University Journal of Social Sciences*, *17*(3), 796–812.

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of make in native and non-native student writing. *Applied Linguistics*, *22*(2), 173–194.

Anthony, E. (1963). Approach, method and technique. *English Language Teaching*, *17*(2), 63–67.

Anthony, L. (2019). *AntConc* (3.5.8). Wasada University. https://www.laurenceanthony.net/software

Avcı, G. B. (2006). *Yazma çalışmaları ile desteklenen eşgörüntüleme öğretiminin öğrencilerin kelime edinimine etkileri*. Unpublished MA Thesis. İstanbul: Marmara Üniversitesi, Eğitim Bilimleri Entitüsü.

Aydoğan, S. (2016). *Task orientation and vocabulary knowdge types: How do they relate to collocational knowledge?* Unpublished MA Thesis. İstanbul: Boğaziçi University, Institute for Graduate Studies in Social Sciences.

Babanoğlu, M. P. (2014). A Corpus-based study on the use of make by Turkish EFL learners. *International Journal of Education and Literacy Studies*, *2*(2), 43–47. https://doi.org/10.7575/aiac.ijels.v.2n.2p.43

Bağcı, N. D. (2014). *Turkish university level EFL learners' collocational knowledge at receptive and productive levels*. Unpublished MA Thesis. Ankara: Middle East Technical University, The Graduate School of Social Sciences.

Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal*, *47*(1), 56–63.

Balcı, Ö. (2006). *Teaching vocabulary through collocations and cliches in EFL classes*. Unpublished MA Thesis. Konya: Selçuk Üniversitesi, Sosyal Bilimler Enstitüsü.

Bartan, Ö. Ş. (2019). Lexical collocation errors in literary translation. *Dil Dergisi*, *Ocak*, 73–88. https://doi.org/10.33690/dilder.528981

Benhamou, E., & Melot, V. (2018). Seven proofs of the Pearson Chi-squared independence test and its graphical interpretation. *ArXiv Preprint ArXiv:1808.09171.*, *3*.

Benson, M., Benson, E., & Ilson, R. (1990). *The BBI dictionary of English word combinations*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing : An automated approach. *Journal of Second Language Writing*, *26*, 28–41. https://doi.org/10.1016/j.jslw.2014.09.004

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, *8*(4), 243–257. https://doi.org/10.1093/llc/8.4.243

Biber, D. (2010). Corpus-based and corpus-driven analyses of language variation and use. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (pp. 159–192). Oxford: Oxford University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. New York: Cambridge University Press.

Bıçkı, A. (2012). *Acquisition of English collocations by adult Turkish L2 learners*. Unpublished Doctoral Thesis. Adana: Çukurova University, Institute of Social Sciences.

Blum, S., & Levenston, E. A. (1978). Universals of lexical simplification. *Language Learning*, *28*(2), 399–415.

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research*, *18*(1), 54–74. https://doi.org/10.1177/1362168813505389

Borgatti, E. (2006). *The use of the verbs "make"and "do" by French and Dutch-speaking EFL learners*. Unpublished MA Thesis. Louven: Université Catholique de Louvain, Louvain-la-Neuve.

Bozoğlu, U. (2020). *A study on teaching collocations (verb + noun) in a secondary school EFL context*. Unpublished MA Thesis. Muğla: Muğla Sıtkı Koçman Üniversitesi, Eğitim Bilimleri Enstitüsü.

Brown, D. F. (1974). Advanced vocabulary teaching: The problem of collocation. *RELC Journal*, *5*(2), 1–11. https://doi.org/10.1177/003368827400500201

Campoy, M. C., Belles, B., & Gea, M. L. (2010). Introduction to corpus linguistics and ELT. In M. C. Campoy, B. Belles, & M. L. Gea (Eds.), *Corpus-based approaches to English language teaching*. Continuum International Publishing Group.

Chan, T. P., & Liou, H. C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning*, *18*(3), 231–250. https://doi.org/10.1080/09588220500185769

Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Québec replications of three European studies. *Canadian Modern Language Review*, *59*(3), 393–423. https://doi.org/10.3138/cmlr.59.3.393

Conrad, S. M. (1999). The importance of corpus-based research for language teachers. *System*, *27*, 1–18.

Cowie, A. P. (1981). The treatment of collocations and idioms in learners ' dictionaries. *Applied Linguistics*, *2*(3), 223–235.

Cowie, A. P. (1998). Past achievements and current trends. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications.* (pp. 1–20). Oxford: Clarendon Press.

Cowie, A. P. (2001). Speech formulae in English: problems of analysis and dictionary treatment. *GAGL: Groninger Arbeiten Zur Germanistischen Linguistik, 44*, 1–12.

Cowie, A. P., Mackin, R., & McCaig, I. R. (1984). *Oxford dictionary of current idiomatic English*. Oxford University Press.

Cowie, A. P., & Ronald, M. (1975). *Oxford dictionary of current idiomatic English*. Oxford: Oxford University Press.

Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, *28*(2), 130–144. https://doi.org/10.1080/09588221.2013.803982

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 600 million words, 1990-present.* https://www.english-corpora.org/coca/

Davies, M. (2015). Corpora. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics* (pp. 11–31). Cambridge: Cambridge University Press. https://doi.org/10.1007/9781139764377.001

Demir, C. (2016). *Lexical proficiency of collocation , boosting , and hedging in academic discourse : A comparative study* (Vol. 2016). Unpublished Doctoral Thesis. Erzurum: The University of Atatürk, The Institute of Educational Sciences.

Deveci, T. (2018). Analysis of collocations in a book chapter and learners' corpus and teaching implications. *Journal of Language Education and Research*, *4*(1), 24–40.

Durmuş, B. Y. (2019). *The differential effect of explicit and implicit instruction on secondary school learners' collocational knowledge*. Unpublished MA Thesis.Adana: Çukurova University, Instiiute of Social Sciences.

Durrant, P. (2008). *High frequency collocations and second language learning* [Unpublished Doctoral Dissertation. Nottingham: University of Nottingham]. http://etheses.nottingham.ac.uk/622/

Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. *International Journal of Corpus Linguistics*, *19*(4), 443–477. https://doi.org/10.1075/ijcl.19.4.01dur

Ergül, Y. (2014). *The effectiveness of using corpus-based the effectiveness of using corpus-based*. Unpublished MA Thesis. Denizli: Pamukkale University, The Institute of Educational Sciences.

Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *IRAL-International Review of Applied Linguistics in Language Teaching*, *33*(4), 315–332.

Farooqui, A. S. (2016). *A corpus-based study of academic- collocatio use and patterns in postgraduate Computer Science students' writing*. Unpublished Doctoral Dissertation. Essex: University of Essex, Department of Language and Linguistics.

Field, A. (2005). *Discovering Statistics Using SPSS* (2nd ed.). London: SAGE Publications.

Firth, J. R. (1968). A synopsis of linguistic theory, 1930–1955. In F. R. Palmer (Ed.), *Selected papers of J.R. Firth 1952–1959* (pp. 168–205). London: Longman.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, *67*(S1), 155–179. https://doi.org/10.1111/lang.12225

Gal'perin, P. Y. (1969). Stages in the development of mental acts. In M. Cole & I. Maltzman (Eds.), *A Handbook of Contemporary Soviet Psychology* (pp. 249–273). New York: Basic Books, Inc.

Gençer, B. B. (2004). *Raising EFL learners' awareness of verb-noun collocations through chunking to extend their collocational knowledge of familiar nouns*. Unpublished MA Thesis. Eskişehir: Anadolu University, Instititute of Educational Sciences.

Gilquin, G. (2007). To err is not all: what corpus and elicitation can reveal about the use of collocations by learners. *Zeitschrift Für Anglistik Und Amerikanistik*, *55*(3), 273–291.

Gitsaki, C. (1996). *The development of ESL collocational knowledge*. Unpublished Doctoral Dissertation. Queensland: The University of Queensland, Centre for Language Teaching and Research.

Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3–18). New York: Routledge.

Greenbaum, S. (1974). Some verb-intensifier collocations in American and British English. *American Speech*, *49*(1), 79–89.

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model : Does semantic transparency matter ? *Language Learning*, *66*(2), 296–323. https://doi.org/10.1111/lang.12143

Halliday, M., & Hassan, R. (1976). *Cohesion in English*. London: Longman.

Hama, H. Q. (2010). *Major sources of collocational errors made by EFL learners at Koya University*. Unpublished MA Thesis. Ankara: Bilkent University, The Graduate School of Education.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, *4*(2), 237–258. https://doi.org/10.1111/j.1473-4192.1994.tb00065.x

Henriksen, B. (2013). Research on L2 learners ' collocational competence and development – a progress report. In B. L. C. Bardel, C. Lindqvist (Ed.), *L2* (pp. 29–56).

Hiltunen, R. (1999). Verbal phrases and phrasal verbs in Early Modern English. In L. J. Brinton & M. Akimoto (Eds.), *Collocational and idiomatic aspects of composite predicates in the history of English* (pp. 133–165). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161–186). Oxford: Oxford University Press.

Hsu, J. (2007). Lexical collocations and their impact on the online writing of Taiwanese college English majors and non-English majors. *Online Submission*.

Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL*, *26*(2), 163–183. https://doi.org/10.1017/S0958344014000020

Hugon, C. (2008). High-frequency verbs : Starting block or stumbling block for advanced L2 communication? Insights from native and learner corpora. In G. Rawoens (Ed.), *Taal aan den lijve. Het gebruik van corpora in taalkundig onderzoek en taalonderwijs* (pp. 69–98). Gent: Academia Press.

Hundt, M., Nesselhauf, N., & Biewer, C. (2007). *Corpus linguistics and the web*. Amsterdam: Brill Rodopi.

Juknevičienė, R. (2008). Collocations with high-frequency verbs in learner English : Lithuanian learners vs native speakers. *Kalbotyra*, *59*(3), 119–127.

Kaszubski, P. (2000). *Selected aspects of lexicon, phraseology and style in the writing of Polish advanced learners of English: A contrastive, corpus-based approach*. Unpublished Doctoral Dissertation. Poznan: Adam Mickiewicz University.

Kayıran, M. T. (2012). *Vocabulary retention: Collocational study*. Unpublished MA Thesis. Mersin: Çağ University, Institute of Social Sciences.

Kennedy, G. (2010). Amplifier collocations in the British National Corpus: Implications for English language teaching. *TESOL Quarterly*, *37*(3), 467. https://doi.org/10.2307/3588400

Kennedy, G. (2014). *An introduction to corpus linguistics*. London: Routledge.

Kim, S. J. (2015). The effect of learner proficiency and L1 transfer on the use ofmake by Korean EFL learners of English. *Language Research*, *51*(1), 139–166.

Koç, G. (2006). *Devoloping collocational awareness*. Unpublished MA Thesis. Ankara: Bilkent University, The Graduate School of Education.

Koskenniemi, I. (1977). On the use of verbal phrases of the type ‘to take revenge’ in English Renaissance drama. *Poetica*, *7*, 80–90.

Koya, T. (2005). *The acquisition of basic collocations by Japanese learners of English*. Unpublished Doctoral Dissertation. Tokyo: Waseda University, Graduate School of Waseda University.

Koya, T. (2006). What is the reality of collocation use by native speakers of English ? *Dialogue*, *5*, 1–18.

Kozłowska, C.-D., & Dzierżanowska, H. (1982). *Dictionary of selected collocations* (J. Hill & M. Lewis (eds.)). London: Language Teaching Publications.

Kurtböke, P. (1998). Non-Equivalence of Delexicalised Verbs in Bilingual Dictionaries. *EURALEX*, 397–404.

Lantolf, J. P., & Tsai, M.-H. (2018). L2 developmental education and systemic theoretical instruction: The case of English verb+noun collocations. In A. E. Tyler, O. Lourdes, M. Uno, & H. I. Park (Eds.), *Usage-inspired L2 Instruction. Researched pedagogy* (49th ed., pp. 29–54). Amsterdam / Philadelphia: John Benjamins Publishing Company. https://doi.org/doi 10.1075/lllt.49

Laporte, S. (2012). Mind the gap! Bridge between world Englishes and learner Englishes in the making. *English Text Construction*, *5*(2), 265–292.

Lareo, I. (2009). Make- collocations in nineteenth-century scientific English. *Studia Neophilologica*, *81*(1), 1–16. https://doi.org/10.1080/00393270802083067

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

Lay, K. J., & Yavuz, M. A. (2020). Targeting Turkish-to-English interlingual interference through context-heavy data-driven learning. *SAGE Open*, *10*(2), 1–12. https://doi.org/10.1177/2158244020920596

Lennon, P. (1996). Getting easy verbs wrong at the adbanced level. *IRAL - International Review of Applied Linguistics in Language Teaching*, *34*(1), 23–36. https://doi.org/doi:10.1515/iral.1996.34.1.23

Leśniewska, J., & Witalisz, E. (2007). Cross-linguistic influence and acceptability judgments of L2 and L1 collocations: A study of advanced Polish learners of English. *EUROSLA Yearbook*, *7*, 27–48. https://doi.org/10.1075/eurosla.7.04les

Lewis, M. (1993). *The lexical approach* (Vol. 1). Hove: Language Teaching Publications.

Lin, C. (2019). Grammatical and lexical patterning of make in Asian learner writing : A corpus-based study of ICNALE. *The Southeast Asian Journal of English Language Studies*, *25*(3), 1–15.

Liu, D., & Lei, L. E. I. (2009). Teaching idiomatic expressions of make, do, have, and other common verbs. *The TESOL Encyclopedia of English Language Teaching*. https://doi.org/10.1002/9781118784235.eelt0786

Macis, M., & Schmitt, N. (2016). The figurative and polysemous nature of collocations and their place in ELT. *ELT Journal*, *71*(1), 1–10. https://doi.org/10.1093/elt/ccw044

McIntosh, C. (2009). *Oxford collocations dictionary for student of English*. Oxford: Oxford University Press.

Mel'čuk, I. (1998). Collocations and lexical functions. In A. P. Cowie (Ed.), *Phraseology, theory, analysis, and applications* (pp. 23–53). Oxford: Oxford University Press.

Men, H. (2020). Data-driven learning in enhancing learners' language idiomaticity. *International Journal of Emerging Technologies in Learning*, *15*(23), 27–41. https://doi.org/10.3991/ijet.v15i23.19023

Millar, N. (2010). The processing of malformed formulaic language. *Applied Linguistics*, *1*(21), 1–21. https://doi.org/10.1093/applin/amq035

Molavi, A., Koosha, M., & Hosseini, H. (2014). A comparative corpus–based analysis of lexical collocations used in EFL textbooks. *Latin American Journal of Content & Language Integrated Learning*, *7*(1), 66–81.

Murao, R. (2004). Ll Influence use on learners ' high-frequency verb + noun collocations. *ARELE: Annual Review of English Language Education in Japan*, *15*, 1–10.

Mutlu, G. (2015). *Knowledge and perceptions of collocations: The cases of Turkish EFL students and teachers*. Unpublished MA Thesis. İstanbul: Marmara University, Institute of Educational Sciences.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston: Heinle Cengage Learning.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*(2), 223–242.

Nesselhauf, N. (2004). What are collocations? In D. Allerton, N. Nesselhauf, & P. Skandera (Eds.), *Phraseological units: Basic concepts and their application* (pp. 1–21). Schwabe.

Nesselhauf, N. (2005). *Collocations in a learner corpus* (N. Nesselhauf (ed.)). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Nickel, G. (1968). Complex verbal structures in English. *IRAL: International Review of Applied Linguistics in Language Teaching, 6*(1), 1.

Nişancı, S. (2014). *Factors influencing Turkish EFL learners' collocation production*. Unpublished MA Thesis. İstanbul: Yeditepe University, The Instıtute of Educational Sciences.

Ördem, E. (2013). *Lexical collocations (verb-noun) across written academic genres in English*. Unpublished Doctoral Dissertation. Adana: Çukurova University, The Institute of Social Sciences.

Otto, P. (2021). Choosing specialized vocabulary to teach with data-driven learning: An example from civil engineering. *English for Specific Purposes, 61,* 32–46. https://doi.org/10.1016/j.esp.2020.08.003

Öztuna, S. (2009). *Effects of input flood and negative evidence on learning of make/do collocations: A study with seventh grade Turkish EFL students*. Unpublished MA Thesis. Eskişehir: Anadolu University, Institute of Educational Sciences.

Öztuna, S. (2014). *Katılım yükü varsayımının yabancı dilde eşdizimli sözcük öğretimindeki rolü: İngilizceyi yabancı dil olarak öğrenen üniversite hazırlık öğrencileri üzerine bir çalışma*. Unpublished Doctoral Dissertation. Ankara: Ankara Üniversitesi, Sosyal Bilimler Enstitüsü.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory : Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication*. London: Longman.

Peksoy, E. (2013). *The corpus-based analysis of authenticity of ELT course books used in high schools in Turkey*. Unpublished MA Thesis. Diyarbakır: Dicle University, The Institute of Educational Sciences.

Qader, D. S. (2018). *The role of teaching lexical collocations in raising learners' speaking fluency*. Unpublished MA Thesis. İstanbul: İstanbul Aydın University, Institute of Social Sciences.

Richards, J. (1976). The Role of Vocabulary Teaching. *TESOL Quarterly, 10*(1), 77–89. https://doi.org/10.2307/3585941

Ringbom, H. (1998). High frequency verbs in ICLE corpus. In *Explorations in Corpus Linguistics* (pp. 191–200).

Salihoğlu, U. M. (2019). *A corpus informed study on learning technical collocations by environmental engineering students*. Unpublished MA Thesis. İstanbul: Yeditepe University, Institiute of Educational Sciences.

Shaw, P. (2001). Investigating learner vocabulary : A possible approach to looking at EFL / ESL learners qualitative knowledge of the word. *IRAL - International Review of Applied Linguistics in Language Teaching*, *39*(January 2001), 171–194. https://doi.org/10.1515/iral.2001.001

Shehata, A. K. (2008). *L1 Influence on the reception and production of collocations by advanced ESL/EFL Arabic learners of English*. Unpublished MA Thesis. Ohio: Ohio University, The College of Arts and Sciences.

Shibliyev, J. (1993). *Responses to tasks involving make/do collocations by Turkic beginning learners of English*. Unpublished MA Thesis. Ankara: Bilkent University, The Institute of Economics and Social Sciences.

Şimşek, G. (2008). *An effective way of teaching vocabulary: Collocations*. Unpublished MA Thesis. Konya: Selçuk Üniversitesi, Sosyal Bilimler Enstitüsü.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sobucalı, G. (2019). *The comparative effects of meaning-focused and form-focused instruction on L2 collocation learning*. Unpublished MA Thesis. İstanbul: Boğaziçi University, Institute for Graduate Studies in Social Sciences.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, *2*(1), 23–55. https://doi.org/10.1075/fol.2.1.03stu

Sun, X., & Hu, G. (2020). Direct and indirect data-driven learning: An experimental study of hedging in an EFL writing class. *Language Teaching Research*, 1–29. https://doi.org/10.1177/1362168820954459

Thornbury, S. (2002). *How to Teach Vocabulary*. London: Pearson Longman.

Üstünalp, İ. (2013). *An analysis of verb-noun collocational error types and error sources in written narrative texts of students majoring in English*. Unpublished MA Thesis. Mersin: Mersin Üniversitesi, Eğitim Bilimleri Enstitüsü.

Vassiljev, L., Skopinskaja, L., & Liiv, S. (2015). The treatment of lexical collocations in EFL coursebooks in the Estonian secondary school context. *Eesti Rakenduslingvistika Ühingu Aastaraamat*, *11*, 297–311.

Visser, F. T. (1963). *An historical syntax of the English language: Part one, syntactical units with one verb*. Leiden: EJ Brill.

Vural, E. (2010). *Explicit and incidental teaching of English verb-noun collocations in an EFL context*. Unpublished Doctoral Dissertation. Eskişehir: Anadolu University, Graduate School of Educational Sciences.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wang, Y., & Shaw, P. (2008). Transfer and universality : Collocation use in advanced Chinese and Swedish learner English. *ICAME Journal*, *32*, 201–232.

Woolard, G. (2000). Collocations - Encouraging learner independence. In M. Lewis (Ed.), *Teaching collocations* (pp. 28–46). Hove: Language Teaching Publications.

Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching*, *32*(4), 213–231. https://doi.org/10.1017/S0261444800014154

Yamashita, J., & Jiang, N. (2010). L1 Influence on the acquisition of L2 collocations : Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*(4), 647–668. https://doi.org/10.5054/tq.2010.235998

Yan, H. (2010). Study on the causes and countermeasures of the lexical collocation mistakes in college English. *English Language Teaching*, *3*(1), 162–165. https://doi.org/10.5539/elt.v3n1p162

Yılmaz, F. (2004). *A study on error analysis in the use of collocation and ididomatic expressions in sentence translation from Turkish to English*. Unpublished MA Thesis. Konya: Selçuk Üniversitesi. Sosyal Bilimler Enstitüsü.

Zhang, X. (1993). *English collocations and their effect on the writing of native and non-native college freshmen.* Doctoral Dissertation. Indiana: Indiana University of Pennsylvania, Department of English.

Zhou, X. (2016). A corpus-based study on high frequency verb collocations in the case of "have." *International Forum of Teaching & Studies*, *12*(1), 42–50.

APPENDICES

APPENDIX – 1a: A Sample Essay from LOCNESS

APPENDIX – 1b: A Sample Essay from the Learner Corpus

APPENDIX – 2: Illustration of the Work with the Concordance Lines

APPENDIX – 3: The Research Ethics Approval

APPENDIX- 1a: A Sample Essay from LOCNESS

Medicine has made many advancements in the past several years. Drugs for many diseases, like Dilantin for the treatment of seizures caused by epilepsy have made the quality of life for several people increase. Drugs promoted the longevity of many patients who otherwise wouldn't have lived otherwise. People victimized by cancer and AIDS have long depended upon Depo-Provera and AZT for their therapeutic value. The miracles of modern biochemical research have dramatically changed the way we treat disease. However, drug companies have escalated their prices over 82% over the decade of the 80's, making a crunch in the pocketbooks of many Americans including the pharmacies and drug buyers, who finally sued to get their fair share of discounts. Blatant overpricing of many important drugs has drastically declined the quality of health care among many Americans, especially the elderly and people with fixed incomes. This is why federal intervention is necessary with the establishment of a price regulation review board to stop the skyrocketing of already outrageous prices.

Health care statistics on expenditures alone are mind boggling. The pharmaceutical industry argues that the cost of drugs used outside hospitals was $36.4 billion in 1991, only five percent of the total national health costs--a whopping $751.8 billion. That sounds like a nice picture. However, these statistics are flawed because 55 percent of what American consumers paid for health care was directly out of pocket, mostly because Americans were getting prescriptions outside of hospitals. This price-gouging shows in high hospital bills, which <*>.

Even though the pharmaceutical industry argues that medical pricing boards would raise prices and eliminate competition between companies, actually the opposite seems to be true. As a result of having medical price boards in other countries, for example, drugs are sold at rates substantially cheaper than the United States. In Canada, where the Patented Medicine Prices Review board was established, prices of drugs by Canadian subsidiaries of American companies are substantially lower, averaging 32%. Similar lower rates also happen in European subsidiaries, where rates are anywhere from 74-45 percent lower than the United States. Most European countries under the European Community tend to have drug price regulation boards as well, hence the lower prices.

In 1991, the Senate Special Committee on Aging reported in 1991 that during the 80's inflation rose only 58 percent while drug prices rose 152%. This happens partly because <*>. They do so in several ways. As companies change packaging and find better ways to use their drugs, they tend to find ways to make more of a profit, sometimes marking up their prices outrageously. An Example of this type of gouging is Depo-Provera. Because the FDA approved the use of this former cancer drug for use as an oral contraceptive, the price jumped from 14$ to 34$ per dose. Even though Upjohn argued that it was trying to recap the cost of RandD by raising the price of a "luxury drug", they were mistaken in doing so. An oral contraceptive is an important drug, too because it

decreases the risk of unwanted pregnancy. Another example of overpricing through packaging is in 1993, when buyers of drugs in hospitals were angered after Dupont-merck discontinued individually wrapped doses of products such as clot preventing agent Coumadin and pain killers Percocet and Percodan. Unit doses prevent mix-ups which could put a client in danger of overdose or receiving the wrong medication. This caused hospitals to absorb a 2400% increase in price because hospitals had to hire more staff to sort the drugs into unit dosages, purchasing drugs at a bulk bottle rate. Yet another example of price blackmail is Sandoz corp. Requiring healthcare providers to buy mandatory monitors for their drug Clozapril, a drug used to treat senile dementia. This drug is used primarily in VA Hospitals, where dementia is common. These monitors cost $9000 per patient and are often of less quality than the ones the VA hospitals can use at a much cheaper rate. Sandoz argues they are responsible is an outside monitor is used, and complications occur from the drug. However, if the monitors are of better quality, this should not be a problem. The FDA did not require this $9000 heart monitor in their agreement when Sandoz got the patent for clozapril. Meanwhile the hospitals are faced with negotiating with another manufacturer of the drug. How can drugs be given and healing happen with such prices over people's heads?

As a nursing student, the indications of this are ominous, patients I treat will not receive proper care because they are afraid of it costing too much for them to take their medicine. I don't want to be the one to tell them that they can't have a certain medicine because they cant afford it. Because of this, these Federal price regulatory board are necessary to keep fairness on the mind of the pharmaceutical industry. If companies like Merck and Upjohn don't look at who they are serving, the Mothers, Fathers and children of America, they will lose business. Yes, it is morally ethical for drug companies to make a profit, but where is the limit? When does drug overpricing stop? It is up to the American people to decide.

APPENDIX- 1b: A Sample Essay from the Learner Corpus

Should Vaccinations Be Mandatory For Children?

Anti-vaccination movements have been rising and gaining supporters all around the world day by day. In 21st century, apart from conventional and anti-modernization groups, educated people living a modern life started to have questions about vaccination as well. Even though vaccinations are one of the most important achivements of mankind, some people go as far as to prevent their child from getting shots knowing the danger they are causing to their children, environment and other people. Although anti-vaxxers think vaccines are harmful for their children because they believe that vaccination cause diseases rather than preventing them and they have side effects and cause autism, vaccines are safe and effective to use and they prevent deadly diseases and save lifes.

Immune system is the way body fights against germs and diseases. When confronted with a disease it produces antibodies that protect the body from illness. Vaccines are also responsible for protecting and giving immunity to the body without actually being sick. Vaccines contain dead or weakened antigens that cause diseases. After getting a vaccine shot, the body trains itself with these weakened antigens and recognizes the disease when actually exposed. As these antigens are either dead or weak they do not cause sickness ("Vaccines: Vac-Gen/Why Are Childhood Vaccines So Important?", 2018). Although new researches about the benefits of vaccines increase day by day, anti-vax movement is also gaining supporters. Currently 10% of U.S are against vaccination and the numbers are continuing to rise (Williams, 2019). They are against vaccines because they think vaccines cause autism and spread illnessess although there is no scientific prove for that.

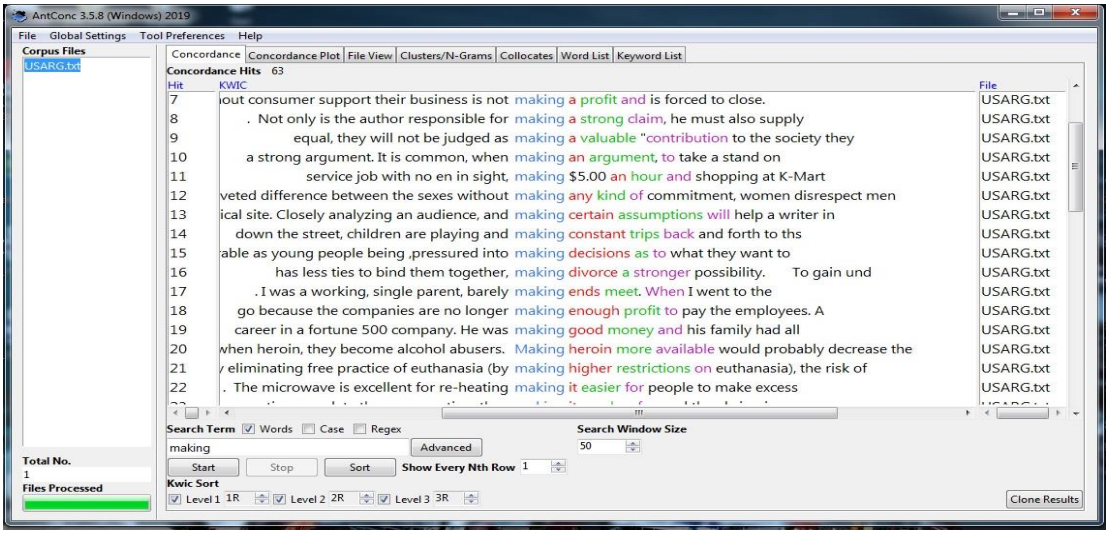
Firstly, parents who don’t want to vaccinate their child put forward the dispute of safety. Anti-vaxxer parents believe that vaccines cause diseases, not prevent them. They often say that chicken pox and measles are not diseases, they are infections and incfections come and go in a week to ten days. By vaccinating, we spread the disease while we have a lower chance getting it naturally they say. However, the reason getting the disease naturally is lower in chance because the diseases we are talking about have nearly been eradicated thanks to the vaccines. For example, your child do not get vaccinated for smallpox because this disease has been eradicated worlwide thanks to smallpox vacciantion. Also, some diseases like diphtheria and polio have become very rare in United States because we kept vaccinating against them ("Vaccines: Vac-Gen/Why Immunize?", 2018). In addition to chicken pox, measles, polio, diphtheria and smallpox, immunisation protects us from other deadly and serious childhood diseases as well, such as meningococcal C, tetanus, rotavirus, mumps and hepatitis which can cause serious health conditions like brain damage, cancer, deafness or even death ("Immunisation is important for children", 2018). As another example and common misconception, many people believe that they get sick after getting a flu shot. However, it is not possible to get sick from the shot because flu shots contain dead viruses. For these

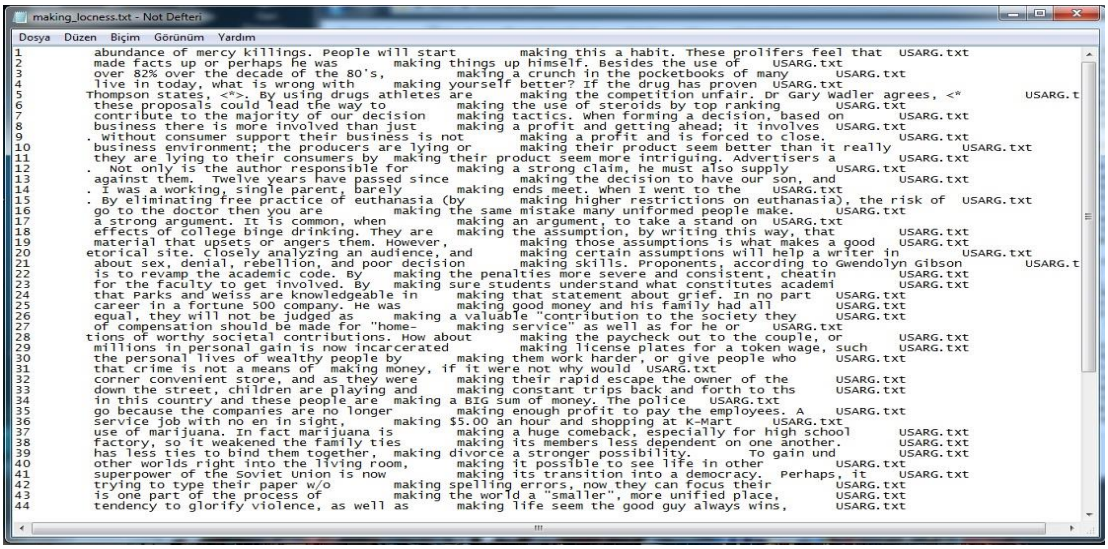
reasons, immunization through dead or weakened pathogens gained by vaccinating promises more protection and is better than risking your child's life with full scale life-threatening diseases and it also prevents outbreaks.

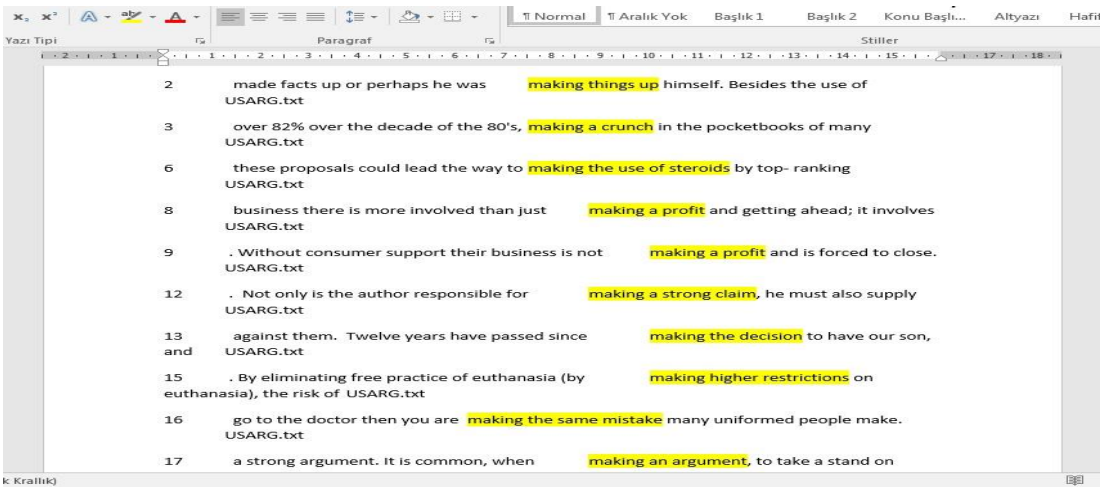
Moreover, another controversial issue about vaccines is that they cause autism and other serious side effects. First of all, no study have ever found an association between autism and vaccines. Many researches have shown that infants can be born with autism without any vaccinations given before. Also, vaccines do not contain harmful toxins that lead up to autism or such things. Some vaccines have so small doses of mercury, formaldehyde and aluminum that these vaccines are completely safe. For example, by the time a child is at the age of 2 he will have taken in only 4mg of aluminum from all mandatory vaccines, whereas a breast-fed baby takes in 10mg aluminum regularly in 6 months and even though not by breast feeding, a soy-based formula will give 120mg aluminum in 6 months again. In addition, the toxic form of mercury is not used in vaccines and infants have 10 times more formaldehyde build up in their bodies naturally than the amount contained in vaccines ("The Importance of Vaccinations", 2018). Although there can be side effects they are very mild. For example, in some cases children can develop mild fever or there can be redness on some parts of the skin like a harmless allergic reaction or swelling around the injection but all these symptoms usually go away in a day. More serious complications are very rare and as your child gets through a series of tests for allergies beforehand, he does not get the shot he is allergic to. Eventually, it has not been scientifically proved that vaccines cause autism and their side effects are so mild that their advantages leave them in shadows.

In conclusion, anti-vaxxers are wrong about thinking that vaccines are harmful, cause autism, contain toxins and spread the diseases instead of preventing them. No relation has been found between autism and vaccines and their ingredients are far from being harmful than the daily products our children eat. When parents decide not to vaccinate their children outbreaks of preventible disesases come into existence. Other kids who are not old enough to get vaccine shots or people who have weak immune systems like cancer patients and transplant recipients can get diseases or even result in death if you do not vaccinate your child ("Importance of Vaccines", 2018). Of course, no medicine can be fully safe, however, benefits of vaccines are far more greater than the risks of getting a disease that can cause serious health conditions and for these reasons vaccination should be mandatory for children.

APPENDIX- 2: Illustration of the Work with the Concordance Lines







APPENDIX – 3: The Research Ethics Approval

Evrak Kayıt Tarihi: 13.03.2020	Protokol No: 23653	Tarih: 03.06.2020

ANADOLU ÜNİVERSİTESİ
SOSYAL VE BEŞERÎ BİLİMLER BİLİMSEL ARAŞTIRMA VE YAYIN ETİĞİ KURULU
KARAR BELGESİ

ÇALIŞMANIN TÜRÜ:	Doktora Tez Çalışması
KONU:	Eğitim Bilimleri
BAŞLIK:	İngilizce Öğretmenliği Bölümü'nde Okuyan Türk Öğrencilerin Tartışmacı Yazılarında Yüksek Sıklıktaki Fiilerle Yapılan Fiil-İsim Kombinasyonlarının İncelenmesi: Make ve Do Örneği An Analysis of Verb-Noun Combinatfons in High Frequency Verbs in Argumentative Essays of Turkish ELT Students: The Case of Make and Do
~~PROJE~~/TEZ YÜRÜTÜCÜSÜ:	Dr. Öğr. Üyesi Gonca SUBAŞI
TEZ YAZARI:	Mehmet KAHRAMAN
ALT KOMİSYON GÖRÜŞÜ:	-
KARAR:	Olumlu

Prof.Dr. Emel ŞIKLAR *(Başkan-İkt. ve İdari Bil. Fak.)*	
Prof.Dr. T. Volkan YÜZER *(Başkan Yardımcısı-Açıköğretim Fak.)*	**Prof.Dr.Esra CEYHAN** *(Eğitim Fak.)*
Prof. Hayri ESMER *(Güzel Sanatlar Fak.)*	**Prof.Dr. M. Erkan ÜYÜMEZ** *(İkt. ve İdari Bil. Fak.)*
Prof.Dr. Handan DEVECİ *(Eğitim Fak.)*	**Prof.Dr. Oktay Cem ADIGÜZEL** *(Eğitim Fak .)*