

KARADENİZ TECHNICAL UNIVERSITY * INSTITUTE OF SOCIAL SCIENCES

DEPARTMENT OF WESTERN LANGUAGES AND LITERATURE

MASTER'S PROGRAM IN APPLIED LINGUISTICS

**CORPUS-BASED INTERLANGUAGE ANALYSIS OF MULTI-WORD COMBINATIONS
IN TERMS OF OVERUSE AND UNDERUSE AS EVIDENCED IN NATIVE AND NON-
NATIVE CORPORA**

MASTER'S THESIS

Handan İLYAS KARATAŞ

MAY-2019

TRABZON

KARADENİZ TECHNICAL UNIVERSITY * INSTITUTE OF SOCIAL SCIENCES

DEPARTMENT OF WESTERN LANGUAGES AND LITERATURE

MASTER'S PROGRAM IN APPLIED LINGUISTICS

**CORPUS-BASED INTERLANGUAGE ANALYSIS OF MULTI-WORD COMBINATIONS
IN TERMS OF OVERUSE AND UNDERUSE AS EVIDENCED IN NATIVE AND NON-
NATIVE CORPORA**

MASTER'S THESIS

Handan İLYAS KARATAŞ




Thesis Advisor: Asst. Prof. Ali Şükrü ÖZBAY

MAY-2019

TRABZON

APPROVAL

Upon the submission of the dissertation, **Handan İLYAS KARATAŞ** has defended the study “**Corpus-Based Interlanguage Analysis of Multi-Word Combinations in Terms of Overuse and Underuse as Evidenced in Native and Non-Native Corpora**” in partial fulfilment of the requirements for the degree of Master of Arts in English Language and Literature at Karadeniz Technical University, and the study has been found fully adequate in scope and quality as a thesis by **unanimous / majority** vote on **24.06.2019**

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

Approval of the Graduate School of Social Sciences

Prof. Dr. Yusuf SÜRMEŒ
Director

DECLARATION OF ORIGINALITY

I, **Handan İLYAS KARATAŞ**, hereby confirm and certify that;

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

17.05.2019

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my thesis advisor Ali Şükrü ÖZBAY who paved the way for the completion of this thesis with his never-ending support and incredible patience, whom I could get in touch whenever I needed, and who always was a perfect supporter of me. I heartily believe that this thesis would not have been possible without his support, patience and guidance, the three qualities he never hesitated to provide me anytime

My professors Mustafa Naci KAYAOĞLU and Mustafa Zeki ÇIRAKLI whose guidance and invaluable insights during my master's study are appreciated a lot. My special thanks go to İbrahim YEREBAKAN with whom I could have the chance of working together for only one month; however, who has always been supportive of and kind to me whenever I got in touch with him.

I am heartily thankful to my THESIS'TERS who became tight-knit during the research and who helped and supported each other more than a friend can do; that is why we call each other sisters. I owe many thanks to Zeynep, Seyhan, Neslihan, Binnur, Mevlüde, Elif, and Nilgün who were always in cooperation when we had questions to ask about research. I do not know how to thank you enough my dear thesis'ters. My dearest friend and former office mate Gökçenaz, you are loved so much! Tuncer, who helped me learn how to use corpus tools, you are, also, greatly appreciated.

I think I am the oldest master's student in the department; and after very many years of being apart from academic research, this was a golden opportunity to go into the depth of academic life and being a student again. This thesis study would not have been complete without the help and support of my life-long coach and my cheer-leader, my mother, who always was always there for me in proceeding and who took care of my kids while I was in full trance during the preparation of this research study. She did more than one can imagine in order for me to earn my master's degree. I cannot thank her enough! I have had the support and encouragement of my husband who was more eager than me on completing this research. My special thanks go to my sister and brother who were always supportive of me while studying. Last but not least, I received generous support from my sister-in-law Bahar who always gave me warm encouragement during my research. And my little babies, you are my inspiration! My apologies if I have inadvertently omitted anyone to whom acknowledgement is due.

May, 2019

Handan İLYAS KARATAŞ

CONTENTS

| | |
|----------------------------|------|
| ACKNOWLEDGEMENT | IV |
| CONTENTS | V |
| ABSTRACT | XII |
| ÖZET | XIII |
| LIST OF TABLES | IX |
| LIST OF FIGURES | XI |
| LIST OF ABBREVIATIONS..... | XII |
| INTRODUCTION..... | 1-9 |

CHAPTER ONE

| | |
|--|--------------|
| 1. LITERATURE REVIEW..... | 10-38 |
| 1.1. Introduction..... | 10 |
| 1.2. Theoretical Framework | 10 |
| 1.2.1. Phraseology Dimension in the Study of Language | 10 |
| 1.2.2. Multi-Word Combinations and their Significance for EFL Learners | 12 |
| 1.2.3. Collocations and their Significance for EFL learners..... | 15 |
| 1.2.4. Collocations in Vocabulary Learning | 20 |
| 1.2.5. Interlanguage..... | 21 |
| 1.2.6. Contrastive Interlanguage Analysis | 25 |
| 1.2.7. Foreignness | 30 |
| 1.2.8. Rationale for Researching Writing Skill..... | 31 |
| 1.2.9. Argumentative Writing and EFL Learners | 32 |
| 1.2.10. Verb+ noun Combinations and Turkish EFL learners | 33 |
| 1.2.11. The Use of Corpus as a Tool for Linguistic Investigation | 33 |
| 1.2.12. Corpus Investigation and ELT | 35 |
| 1.2.13. The Importance of Corpus in ELT Research | 36 |

CHAPTER TWO

| | |
|----------------------------|--------------|
| 2. METHODOLOGY..... | 39-49 |
| 2.1. Introduction..... | 39 |

| | |
|--|----|
| 2.2. Methodology | 39 |
| 2.3. Setting and Sample Selection: Participants..... | 41 |
| 2.4. Compilation of TUCLE..... | 42 |
| 2.5. Reference Corpus: LOCNESS | 47 |
| 2.6. Analysis of the Comparisons Among the Three Corpora | 47 |
| 2.7. Criteria for the Selection of Verb+noun Collocations | 48 |
| 2.8. Data Collection | 48 |
| 2.8.1. Demographic Information Questionnaire and Consent Form..... | 48 |
| 2.9. Corpus Tool..... | 48 |

CHAPTER THREE

| | |
|---|--------------|
| 3. FINDINGS AND DISCUSSION | 50-88 |
| 3.1. Overview of the Study | 50 |
| 3.2. Analysis of the Comparisons across TUCLE, KTUCLE and LOCNESS | 50 |
| 3.3. Discussion of the Research Questions | 81 |
| 3.3.1. Discussion of the First Research Question | 82 |
| 3.3.2. Discussion of the Second Research Question..... | 84 |
| 3.3.2.1. Overuse Evidences of Verb+noun Collocations in TUCLE and KTUCLE ... | 84 |
| 3.3.2.1.1. Overuse Evidences of Verb+noun Collocations in TUCLE | 84 |
| 3.3.2.1.2. Overuse Evidences of Verb+noun Collocations in KTUCLE | 85 |
| 3.3.2.2. Underuse Evidences of Verb+noun Collocations in TUCLE and KTUCLE . | 86 |
| 3.3.2.2.1 Underuse Evidences of Verb+noun Collocations in TUCLE..... | 86 |
| 3.3.2.2.2. Underuse Evidences of Verb+noun Collocations in KTUCLE | 87 |
| 3.3.2.3. Misuse Evidences of Verb+noun Collocations in TUCLE and KTUCLE | 87 |
| 3.3.3. Discussion of the Third Research Question..... | 88 |
| CONCLUSION AND SUGGESTIONS | 89-91 |
| REFERENCES | 92 |
| APPENDICES | 103 |
| CURRICULUM VITAE..... | 112 |

ABSTRACT

Argumentative essay writing is one of the areas in which language learners should be competent enough to be able to make their arguments and defences be visible by the readers. In this context, essays of students of English as a Foreign Language (EFL) are analysed since they may seem to present problems of writing which make their written productions seem rather inadequate in some respects when compared to the essays of native speakers in terms of collocational content, which is the main rationale for this study. Therefore, the purpose of this research study is to examine the verb+noun collocations with seven base verbs in order to reveal similarities and differences, the overuse and the underuse patterns applied by the native and two non-native corpora. With this objective in mind, the researcher compiled her own corpus named TUCLE which stands for Trabzon University Corpus of Learner English with 21 preparatory students during the time she was teaching reading and writing skills. The second non-native corpus is KTUCLE which stands for Karadeniz Technical University Corpus of Learner English. Lexical comparisons of TUCLE and KTUCLE were carried out to display the collocational competencies of non-native tertiary level EFL students; thus, the research required a reference corpus for comparison. As a result, LOCNESS, which stands for Louvain Corpus of Native English Essays, was used. All three corpora comprise of argumentative essays, which made it possible to make contrastive interlanguage analysis among them. The data for the study was quantitative in nature; the research was carried out with the help of AntFile Converter and Sketch Engine which are freely available online corpus tools. The contrastive interlanguage analysis of the corpora revealed that non-natives are not as competent as natives in terms of their collocational usages. Based on the analysis, it was observed that native students employed various types of combinations quite unlike their non-native counterparts who seemingly employed limited amount and number of these collocations with limited variety and scope. Additionally, non-native learners as evidenced in the two corpora (TUCLE and KTUCLE) seemed to show various overuse and underuse patterns in terms of the use of collocations, which may be due to the several factors such as inadequate competency and proficiency in vocabulary and collocations.

Keywords: Argumentative writing, learner corpora, collocations, tertiary level students.

ÖZET

Tartışmacı yazı, dil öğrenenlerin yazı yazarken savundukları fikirleri ve iddiaları okuyuculara başarılı bir şekilde aktarmalarını gerektiren bir yazın türüdür. Bu bağlamda, İngilizce'yi Yabancı Dil (EFL) olarak öğrenen katılımcılar, yazılarının eşdizim özelliklerinin zayıf olmasına sebep olan problemlerle karşılaştığı için incelenmiştir; eşdizim incelemesi bu çalışmanın ana konusudur. Bu hedefle, bu çalışmanın amacı kullanım sıklıklarını göstermek için anadili İngilizce olan ve anadili İngilizce olmayan derlemlerde yedi temel fiille birlikte oluşturulan fiil+isim eşdizimlerini incelemektir. Bu niyetle, araştırmacı TUCLE adını verdiği kendi derlemini oluşturmuştur. Bu derlemin katılımcıları Trabzon Üniversitesi İngilizce Öğretmenliği bölümünde hazırlık okuyan ve araştırmacının iki dönem boyunca okuma ve yazma dersi verdiği 21 öğrencidir. Bir diğer anadili İngilizce olmayan derlem Karadeniz Teknik Üniversitesi hazırlık öğrencilerinin oluşturduğu 480 katılımcılı KTUCLE'dir. TUCLE ve KTUCLE'nin sözcüksel incelemesi üniversite seviyesindeki EFL öğrencilerinin eşdizim yeterliliklerini göstermek amacıyla yapılmıştır. Bu karşılaştırma için bir referans gerekli olduğundan dolayı referans derlem olarak LOCNESS kullanıldı. Üç derlem de içerdiği tartışmacı yazılar üzerinden incelendi. Çalışmanın verileri nicel olarak elde edilmiştir; bunun için Sketch Engine adı verilen derlem aracı kullanılmıştır. Üç derlemin karşılaştırmalı ara dil analizi, eş dizim kullanımının anadili İngilizce olmayan katılımcıların anadili İngilizce olanlardan daha zayıf olduğunu göstermiştir. Anadili İngilizce olan katılımcılar çok sayıda eşdizim kullanırken anadili İngilizce olmayan katılımcılar bazı belirli eşdizimlerle sınırlı kalmıştır. Bunun yanı sıra, TUCLE ve KTUCLE katılımcıları bazı eşdizimleri fazla bazılarını da az kullanmıştır; bu da kelime ve eşdizim haznelerinin yeterli derecede güçlü olmadıklarının sonucu olarak görünmektedir.

Anahtar Kelimeler: Tartışmacı yazı, öğrenen derlemi, eşdizimler, üniversite öğrencileri.

LIST OF TABLES

| Table No. | Name of Table | Page No. |
|-----------|--|----------|
| 1. | Collocation Patterns..... | 20 |
| 2. | The Learner Corpora Used in the Study | 40 |
| 3. | The Study Design Steps Followed in the Methodology Section | 41 |
| 4. | Demographic Information of TUCLE Participants (Subjects) | 42 |
| 5. | The Argumentative Essay Topics of TUCLE..... | 45 |
| 6. | The Corpus Design Criteria of TUCLE..... | 46 |
| 7. | Design Criteria for Language Variable Distribution of TUCLE | 46 |
| 8. | Distribution of “do” in TUCLE, KTUCLE and LOCNESS | 51 |
| 9. | Log-likelihood Ratio and Significance Level with “do” in TUCLE and LOCNESS | 53 |
| 10. | Log-likelihood Ratio and Significance Level with “do” in KTUCLE and LOCNESS .. | 54 |
| 11. | Distribution of “make” in TUCLE, KTUCLE and LOCNESS | 56 |
| 12. | Log-likelihood Ratio and Significance Level with “Make” in TUCLE and LOCNESS | 58 |
| 13. | Log-likelihood Ratio and Significance Level with “Make” in KTUCLE and LOCNESS | 59 |
| 14. | Distribution of “have” in TUCLE, KTUCLE and LOCNESS | 60 |
| 15. | Log-likelihood Ratio and Significance Level with “have” in TUCLE and LOCNESS | 64 |
| 16. | Log-likelihood Ratio and Significance Level with “have” in KTUCLE and LOCNESS | 67 |
| 17. | Distribution of “give” in TUCLE, KTUCLE and LOCNESS | 69 |
| 18. | Log-likelihood Ratio and Significance Level with “give” in TUCLE and LOCNESS .. | 71 |
| 19. | Log-likelihood Ratio and Significance Level with “give” in KTUCLE and LOCNESS | 72 |
| 20. | Distribution of “take” in TUCLE, KTUCLE and LOCNESS | 73 |
| 21. | Log-likelihood Ratio and Significance Level with “take” in TUCLE and LOCNES .. | 75 |
| 22. | Log-likelihood Ratio and Significance Level with “take” in KTUCLE and LOCNESS | 76 |
| 23. | Distribution of “get” in TUCLE, KTUCLE and LOCNESS | 77 |
| 24. | Log-likelihood Ratio and Significance Level with “get” in TUCLE and LOCNESS .. | 78 |
| 25. | Log-likelihood Ratio and Significance Level with “get” in KTUCLE and LOCNESS .. | 79 |

| | | |
|-----|--|----|
| 26. | Distribution of “go” in TUCLE, KTUCLE and LOCNESS | 79 |
| 27. | Log-likelihood Ratio and Significance Level with “go” in TUCLE and LOCNESS | 80 |
| 28. | Log-likelihood Ratio and Significance Level with “go” in KTUCLE and LOCNESS .. | 81 |
| 29. | The Total Number of Verb+noun Collocations in the Three Corpora..... | 82 |
| 30. | The Overused Verb+noun Collocations in TUCLE | 85 |
| 31. | The Overused Verb+noun Collocations in KTUCLE..... | 86 |
| 32. | The Underused Verb+noun Collocations in TUCLE | 86 |
| 33. | The Underused Verb+noun Collocations in KTUCLE..... | 87 |
| 34. | The Misused Verb+noun Collocations in the Non-native Corpora | 88 |



LIST OF FIGURES

| Figure No. | Name of Figure | Page No. |
|------------|--|----------|
| 1. | Phraseology..... | 11 |
| 2. | Contrastive Interlanguage Analysis | 27 |
| 3. | Core Components of Learner Corpus Research | 28 |



LIST OF ABBREVIATIONS

| | |
|---------|--|
| BNC | : British National Corpus |
| CIA | : Contrastive Interlanguage Analysis |
| EFL | : English as a Foreign Language |
| ELT | : English Language Teaching |
| IL | : Interlanguage |
| KTUCLE | : Karadeniz Technical University Corpus of Learner English |
| L1 | : First Language |
| LL | : Log-likelihood Measure |
| LOCNESS | : Louvain Corpus of Native English Essays |
| MWC | : Multi-word Combination |
| NL | : Native Language |
| NNS | : Non-native Speakers |
| NS | : Native Speakers |
| SLA | : Second Language Acquisition |
| TL | : Target Language |
| TUCLE | : Trabzon University Corpus of Learner English |

INTRODUCTION

This corpus-based descriptive study presents a detailed examination of multi-word combinations in English used by natives and non-native tertiary level English as a Foreign Language (EFL) students. The motivation to employ a corpus-based study derived from my interest in teaching and in analyzing EFL learners' argumentations with a special focus on the ways they deal with multi-word combinations. The place of corpus-linguistics in foreign language teaching environment encouraged me to employ my study with the help of corpus data.

Language is organized in itself, and according to Cowie (1998), Schmitt (2004), and Wray (2002), a language user should be competent in that systematic nature of the language. One systematic feature of a language is multi-word combinations (MWCs), and collocations are multi-word structures as well as idioms and fixed expressions. However, collocations are the basically analysed phraseological structures in this thesis. According to Palmer (1933: 5), "a collocation is a succession of two or more words that must be learned as an integral whole and not pieced together from its component parts" (as cited in Cowie, 1999: 54). They are of utmost importance in a language since a learner is not likely to be competent enough or native-like without being able to use them properly. Granger (2003; 1998) argues that non-native users of a target language are less likely to use phraseological items, specifically collocations, than the native speakers who are inherently capable of using collocations.

There is no doubt that collocations are significant parts of a language vocabulary (Nation, 2000). Hence, it would be possible to say that language proficiency comes with sturdy use of collocations. However, non-native users of a language may not employ fully correct collocational usage and therefore, it creates foreignness in using the target language. Furthermore, as Hill (2000), Wray (2002) and Sadeghi (2009) argue, when a learner has a substantial amount of collocations in mind, she can be a proficient and a fluent language user. Collocations are vital since, according to Gledhill (2000) they "appear to be fundamental units in the stylistic description of text" (2000: 234).

After presenting the importance of collocations, it would be wise to give the definition of corpus linguistics and how its significance is related to this thesis, since this present study is a corpus-based analysis of multi-word combinations. Sinclair (2005) explains what corpus is as "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic

research” (2005: 16) (as cited in Özbay and Kayaoğlu, 2016: 343). Additionally, Koo (2006) defines corpus as “a collection of authentic language production” (2006: 1). Granger et al. (2002) defines corpus linguistics “as a linguistic methodology which is founded on the use of electronic collections of naturally occurring texts, viz. corpora” (2002: 4). The researchers also defend that “the very nature of the evidence it uses makes it a particularly powerful methodology, one which has the potential to change perspectives on language” (2002: 4). This argument reveals the importance of corpus linguistics since as Kennedy (1998) argue that corpora (plural form of corpus) promises with “the basis for more accurate and reliable descriptions of how languages are structured and used” (1998: 88).

In the light of these, this study aims to reveal the English language learners’ multi-word combination usage -specifically verb+noun collocations- differences by comparing three different corpora; namely, LOCNESS (Louvain Corpus of Native English Essays) which includes 361,054 words; KTUCLE (Karadeniz Technical University Corpus of Learner English) including 709,748 words; and thirdly TUCLE (Trabzon University Corpus of Learner English) that covers almost 195,684 words. TUCLE is the corpus which consists of the argumentative essays compiled by the researcher herself according to some strict design criteria.

The introduction chapter of this study takes in six subheadings named background of the study, purpose of the study, significance of the study, statement of the problem, research questions, and finally organization of the thesis.

English language has been widely used in very many countries both as a foreign language and a second language; however, learners of the target language face problems with using collocations, multi-word combinations, and lexical bundles. So, this incompetency in the target language paves the way for incorrect and imperfect English usages (Hinkel, 2004; Reynolds, 2005). This present thesis deals with collocations which are considered as one of the phraseological terms.

Phraseology studies the use of word combinations that are of utmost importance and therefore it can provide us with a perception about the intended meanings in the contexts used. Phraseology can be described as “the linguistic sub-discipline which deals with MWEs” (multi-word expressions) (Müller, 2011: 1). In the abovementioned definitions, multi-word expressions refer to multi-word combinations since the term multi-word carries various names such as combinations, structures, expressions, and constructions. Additionally, phraseology can be defined as “the preferred way of saying things within a particular discourse” (Gledhill, 2000: 235). The researcher provides another definition of phraseology as “a dimension of language use in which patterns of wording (lexico-grammatical patterns) encode semantic views of the world, and at a higher level idioms and lexical phrases have rhetorical and textual roles within a specific discourse” (Gledhill, 2000: 235).

Idioms, fixed expressions and collocations are listed under the umbrella term ‘phraseology’ as well (Cowie, 1998). Idioms are “the single words that make up the phrase and the semantic and syntactic information they entail do not play a role for the idiom as a unit” (Sprenger, 2003: 42). This results in the inference that idioms are multi-word combination types which do not have the dictionary meanings of its components; that is, in idioms every single unit has a dictionary meaning when examined separately, yet they create a totally different meaning when they are lined together. ‘Catch one’s eye’ can be given as an example of idioms as catch and eye do not carry their literal meanings when they are used together.

As for fixed expressions, Sprenger (2003) defines them as “specific combinations of two or more words that are typically used to express a specific concept” (2003: 4). Some examples of fixed expressions can be ‘all of a sudden, come to no good’. These meanings of these expressions can be inferred from the literal meanings of every single word; the important thing is they are generally used together and this feature of them makes those units fixed expressions. Finally, the third part of phraseological units is collocations and they can be defined as two or more words that generally go together and form a multi-word combination (Palmer, 1933; Hill, 2000; Wray, 2002). With all these being said, it would be wise to define phraseology as a linguistic division which deals with each form of word combinations that are collocations, fixed expressions and idioms (Granger and Paquot, 2008). Besides, those phraseological structures in a language are of importance since as Gledhill (2000: 1) argues, “it is impossible for a writer to be fluent without a thorough knowledge of the phraseology of the particular field he or she is writing in”.

One should not ignore the role of phraseology in a language study because as Müller (2011) states, corpus linguistics and phraseology are strongly connected to each other. Corpus plays a significant role in language studies with the help of advancements in corpus tools and computerized corpora (Granger, 1998a). In addition, multi-word combinations are better studied with corpus based research (Altenberg & Granger, 2001; Altenberg & Granger, 2002; Nesselhauf; 2005). Research carried out by different researchers (Moon, 1998a; De Groot, 1999; Cox, 2000) can reveal the importance of corpus-based research in phraseology (as cited in Sprenger, 2003) since they could list thousands of items including idioms, proverbs, and sayings; and they could study the usage frequencies subsequently.

This present study works on multi-word combinations in terms of their positions in interlanguage. Interlanguage can be called as the learner language in which the non-native users of the target language attempts to produce language; in other words, it is as Selinker (1972: 214) states “a learner’s attempted production”. This production attempt can refer to the imperfect language behavior because it is merely a try to be fluent and competent in the target language. The effect of interlanguage process can be defined by Granger (1996: 48) as it is “situated somewhere between L1 and L2 and are likely to contain examples of transfer” (as cited in Xiao, 2007: 5) which

can refer to the supposition that interlanguage is not perfect and it has the potential to include samples of L1 transfer which leads to the imperfect usage of the target language.

This current research focuses on verb+noun collocations in interlanguage proficiencies of tertiary level EFL students. Fontenelle (1998) argues that verb+noun combinations hold an active role among the collocation types; and one unit of those phraseological patterns, verb+noun collocations are examined in this present study as they are the most negatively affected constructions resulting from native language transfer (Farghal and Obiedat, 1995; Martelli, 2006). This corpus-based research examines tertiary level students' argumentative essays with the purpose of seeing the frequencies of the most frequently used verb+noun collocations in their writings.

With all the information provided above, the background of the study deals with related research about multi-word combination studies conducted with EFL (English as a Foreign Language) students both in Turkey and around the world. The purpose of the study part deals with the aim why this research was carried out; that is, does it fill any gap in EFL teaching environment? The significance of the study part is related to how this thesis can contribute multi-word combination usage in EFL environments. The problem statement part is interested in clarifying the focus of this study. Three main research questions of this study are listed with the purpose of being answered in the following sections of this study. In the limitations of the study part, encountered drawbacks are revealed to let the next researchers to be pre-informed about the studies they intend to employ.

Background of the Study

Language composes of four main skills named reading, speaking, listening, and finally and most importantly for my study; writing. These productive skills require a proficient knowledge of the target language in an attempt to write scholarly in general; and in this context, writing academically in particular. Writing is thought to form a challenging part of those abovementioned skills. It is considered to be the most difficult skill since as Asikin (2017) argues, one cannot use any gestures, non-verbal or facial expressions to clearly deliver the message as it is in speaking. It is of utmost importance since readers should understand the implied meaning only with the transcribed words; that is, the writer is left alone with her words and sentences. A target language learner's writing goes through tough processes since they are not likely to develop writing competency when even a native cannot write perfectly. In brief; an EFL learner, tertiary level EFL learners in particular, should be able to write clearly and grammatically correct sentences to achieve her aim to be able to write scholarly.

Some of the studies conducted on writing are listed in this section of the background of the study part. Narita, Sato, and Sugiura (2004) carried out a corpus-based study whose participants

were Japanese learners of advanced English. The study examined the students' essays in terms of connectors and found that Japanese students overused some transitions such as 'first', 'of course', 'in addition' and 'moreover'. However, they underused transitions such as 'yet', 'then' and 'instead'. This underuse and overuse can be a result of L1 transfer despite its being not certain. Although my study focuses on collocations, the reason why I attached this study related to connectors into my research is that non-native learners of a language tend to make mistakes or make L1 transfer in their writings.

This present research concentrates on collocations because they seem to have a noteworthy place in proficiency and competency in a target language. Sadeghi (2009) puts forward that a competent user of a target language should not be negatively affected by her native language because of her inadequate proficiency. Thus, this can be avoided with the help of acquiring collocational usage because the more a language learner can employ collocations in her productive skills, the more competent she can be. In the light of this, Nizonkiza (2011) implemented a research with 104 students and reached to the conclusion that competent use of collocations lead the language user to be fluent and proficient in the target language.

In the same way, the study carried out by Zareva, Schwanenflugel and Nikolova (2005) presented that the proficiency of a language learner is positively correlated with her lexical and collocational knowledge. The study also introduced that the more collocations the learner knows and uses, the more competent and proficient she will be in the target language.

In addition, Laufer and Waldman's (2011) research supports the research above with their conclusion "that the number of collocations increased only at the advanced level, and that errors continued to persist even at advanced levels of proficiency" (2011: 647). This simply backs the argument that a learner can be proficient when she can employ collocational usage.

This study focusses on multi-word combinations, specifically verb+noun collocations, and some related research was found to compare writings of Turkish and foreign students in terms of their collocational usages. Before moving into the next title, I would like to place Barfield's (2006) and Laufer and Waldman's (2011) studies into my background of the study. The researchers analysed collocational usages of non-native learners; however, Barfield (2006) studied with Japanese learners of English whereas Laufer and Waldman (2011) studied Israeli Learner Corpus of Written English in which native Hebrew and native Arabic learners of English had essays. Japanese learners of English were not as well in using multi-word combinations as they were in using single nouns and verbs, although the multi-word combinations included same nouns and verbs. Additionally, Laufer and Waldman's (2011) study revealed that non-native learners of English in their study utilized verb+noun collocations fewer than natives. They also asserted that the ones who used verb+noun collocations were only advanced level learners of English.

The purpose of the study

The rationale for the present study is given in this section. The purpose of this study is to state why collocations are chosen whereas there are other multi-word combinations such as idioms (e.g. spill the beans) and fixed expressions (e.g. a pain in the neck). One reason is that, idioms are rarely seen in students' papers; they are not frequently used even in native speakers' productive skills. Cowie (1998: 10) states that "idioms in the strict sense (bite the dust, spill the beans) are exceedingly rare". As Seidlhofer (2009) argues idioms are generally faced in native-speaker usages whereas non-natives do not seem to produce them as much as natives. The reason for this may be the need for communication which can be provided without idioms. Native language speakers can use idioms without contemplating much since they already have those idioms instinctively whereas non-native speakers need "long exposure or experience within a particular community" (Seidlhofer, 2009: 199). This statement may support the idea that when a person can use idioms in productive skills, it actually means she has been exposed to target language environment with a sufficient time and effort. As for fixed expressions, it can be said that they are not possible to examine in students' essays since there are too many expressions to study.

Leaving idioms and fixed expressions above, this research studies verb+noun collocations which are one type of collocations. The others can be listed as adjective+noun collocations (weak tea); noun+verb collocations (cats purr); adverb+adjective collocations (fully aware); noun+noun collocations (milk chocolate); verb+adverb collocations (rain heavily); adverb+verb collocations (finely chopped); and for this study verb+noun collocations (do homework). The reason of choosing verb+noun collocations is as Martelli (2006) states, verb+noun collocations are the most negatively affected combinations as a result of L1 transfer. Similarly, the research carried out by Farghal and Obiedat (1995) included collocation analysis and concluded that mistakes and problems in collocational usage resulted from native language transfer.

Bahns and Eldaw (1993) carried out a translation study with advanced speakers of German language about verb+noun collocations, and they found that the base verbs in verb+noun collocations created problems in translation whereas translation of other lexical items did not lead to as many problems as verb+noun collocations. Their study also concluded that collocational errors occurred both in beginner and advanced learners' translations.

The verb+noun collocations in this present thesis are formed through seven base verbs, which are *do, make, have, give, take, get* and *go*. The reason I decided on these verbs is that these are the mostly known and used words in students' papers which I had analyzed during my 11 years of teaching English as a foreign language. Sprenger (2003) states in his doctoral dissertation that if a multi-word combination occurs frequently, it means it could be examined in more detail. That's why I determined these base verbs to study the verb+noun collocations.

This study excluded academic corpora; it studied learner corpora. The study was conducted through expository argumentation as opposed to academic argumentation. The reason why academic corpora were not chosen can be explained by Flowerdew (2004: 11) who state “that as general corpora have proved to be extremely useful for understanding how naturally-occurring language operates, then by the same token, specialized corpora can also prove to be of value in understanding academic and professional language”. Any verb+noun collocation was analysed to see if it has a combination feature by looking at its frequency. Verb+noun combinations were studied on a wide basis and they were included in the study. The verbs to be used in verb+noun collocations are *do, make, have, give, take, get* and *go*.

The verb+noun collocations in this present study will be based on the verbs *do, make, have, give, take, get* and *go*; for example, in “have a look” the base is ‘have’ and the ‘noun’ is the element which carries the majority of the meaning. In the abovementioned verb+noun construction, the base verb works as a complementary element which helps the noun gain a clear meaning. The rationale behind studying collocations is that non-native speakers of English face problems during the production of collocations in the target language (Lewis, 2000). Their proficiency levels are not directly correlated with the competent use of collocations. However, the more proficient a language user is the more collocations she can employ in her skills (Zareva, Schwanenflugel & Nikolova, 2005; Laufer & Waldman, 2011; Nizonkiza, 2011). This thesis study intends to analyse the use of verb+noun collocations in intermediate level speakers’ argumentative essays. The reason why argumentative writing was chosen is that argumentative essay writers should be able to create a claim and support it with her own ideas and evidences (Coşkun and Tiryaki, 2011). This procedure of producing a claim, supporting it and trying to persuade the reader needs logical argumentative thinking; therefore, this study investigates how much and how efficient the learner can employ collocations during the procedure of persuading the readers.

Significance of the Study

Multi-word combinations have been given different names so far, some called them pre-fabricated (pre-fab) items, formulaic sequences, chunks, and ready-made units (Howarth, 1998; Wray, 1999). No matter what their names are, they are considered as essential parts of a language without which a learner cannot gain native-like fluency and scholarly writing proficiency (Hill, 2000).

In the light of these, this study aimed at analysing verb+noun collocations to see how non-native users of English utilize collocations. The findings of the research are likely to reveal some truth regarding the nature of students that they cannot use verb+noun collocations properly and proficiently, or they are proficient enough. However, it may not be possible to infer that since non-native learners cannot always be as competent as native users of a target language. Lack of

problems in usages of verb+noun collocations may lead teachers and instructors to design a path to proceed. Therefore, it can be said that new vocabulary or multi-word combination teaching methods can be created following the results of this study. New practical pedagogical methods or tools can be devised to prevent the learners from, as Granger (1998b: 158) argues “foreignsounding”.

Statement of the Problem

Collocations are a language’s vital parts since they are the signals of proficient and competent usages (Brown, 1974; McArthur, 1992; Fontenelle, 1998; Howarth, 1998; Hill, 2000; Lewis, 2000; Nation, 2000; Martynska, 2004; Wei and Lei, 2011;). When a learner can place collocations into her writing, it shows she is a proficient user of the target language. The problem in a non-native language learner is that she either does not know any collocations or she avoids using any in order to prevent mistakes. Biskup (1992) carried out a study with advanced level 34 Polish and 28 German learners. The study required the participants to translate collocations into English, and there occurred differences which revealed that Polish students were able to use more collocations than Germans; however, Germans could provide the intended meanings more clearly although they were not as proficient in collocational usage as Polish were. This can lead to the argument that, these European countries differ in foreign language teaching as the students are possibly taught via different teaching methods because Polish students were more likely to employ L1 transfer than German students did.

Similar to this present thesis study, Kaszubski (2000) carried out a research with a native corpus and different level non-native learners’ corpora including verb+noun collocations (as cited in Nesselhauf, 2005). The study included intermediate learners of English who were Polish and Spanish, and advanced level learners of English who were Polish and French. This current study of mine includes verb+noun collocations whose base verbs are *do*, *make*, *have*, *give*, *take*, *get* and *go* as well. Kaszubski’s (2000) study deals with additionally the verb ‘be’. Kaszubski (2000) reveals that intermediate and advanced level learners of English use few collocations whereas they repeatedly use those few collocations.

In a similar way, Howarth’s (1996) study examined verb+noun collocations and found that non-native users of a language employ fewer collocations than natives, and the proficiency level of a non-native learner is not effective in collocational usage. In the light of these, it could be inferred that non-native speakers of a language use fewer collocations than the native speakers of that language (Biskup, 1992; Howarth, 1996; Kaszubski, 2000).

With all these being said, this study will research the non-native learners of English in terms of their verb+noun collocation usages in their argumentative essays. The base verbs of the

verb+noun collocations to be studied are do, make, have, give, take, get and go. In addition, the research did not include all types of collocations since it would be impossible to focus on all combinations to infer a meaningful result. It consisted of only one type to be able to get more detailed information and therefore let the research create a meaningful finding.

Research Questions

This present study centrally puts emphasis on verb+noun collocations and their usage differences by native and non-native users of English. Although it may answer some specific questions, main research questions are primarily focused on.

1. How do native and non-native users differ in verb+noun collocation usages? Is there any statistically significant difference between natives and non-natives in terms of interlanguage developmental levels?
2. What are some of the most common overused, underused and misused patterns of verb+noun collocations in non-native corpora?
3. What are some of the most common “base verbs” that are used as complementary to in the verb+noun collocations?

Organization of the Thesis

This research aims at studying verb+noun collocations which are a part of multi-word combinations. The study is carried out by means of three different corpora two of which are composed by two different non-native learner groups of English in tertiary EFL setting in Turkey, and last one is called LOCNESS which is a non-academic native corpus. Two learner corpora consist of TUCLE (Trabzon University Corpus of Learner English) which is formed by the English Language Teaching students' argumentative essays while the other corpus is KTUCLE (Karadeniz Technical University Corpus of Learner English) which consists of argumentative papers of Karadeniz Technical University English Language and Literature Department students. This thesis includes six chapters.

CHAPTER ONE

1. LITERATURE REVIEW

1.1. Introduction

This chapter revolves around the issues which are directly related to the research questions listed in the introduction part. The major titles to be examined are as follows; phraseology and its main components such as multi-word combinations and collocations and their importance in language. Additionally; interlanguage, Contrastive Interlanguage Analysis (CIA) and foreignness are examined. Reasons of utilizing multi-word combinations and the verbs ‘*do, make, have, give, take, get* and *go*’ are listed in the related section. This study is researching argumentative writing, and the cause of choosing the writing skill and the argumentative type is explained in the related section. Besides, this study is a corpus-based work, and the reason of utilizing corpus is revealed as well. This is a linguistic study which focuses on English Language Teaching (ELT); therefore, the relationship between corpus and ELT and the ELT studies implemented with the help of corpus are presented in the related sections.

1.2. Theoretical Framework

This study is associated with ELT, corpus and multi-word combinations; specifically, verb+noun collocations and includes major concepts such as phraseology, contrastive interlanguage analysis (CIA) and corpus investigation. In other words, theoretical framework includes the study of phraseology, contrastive interlanguage analysis, learner corpora through argumentative essays and corpus-based ELT. These titles form the theoretical framework of the present study and is dealt in this literature review section.

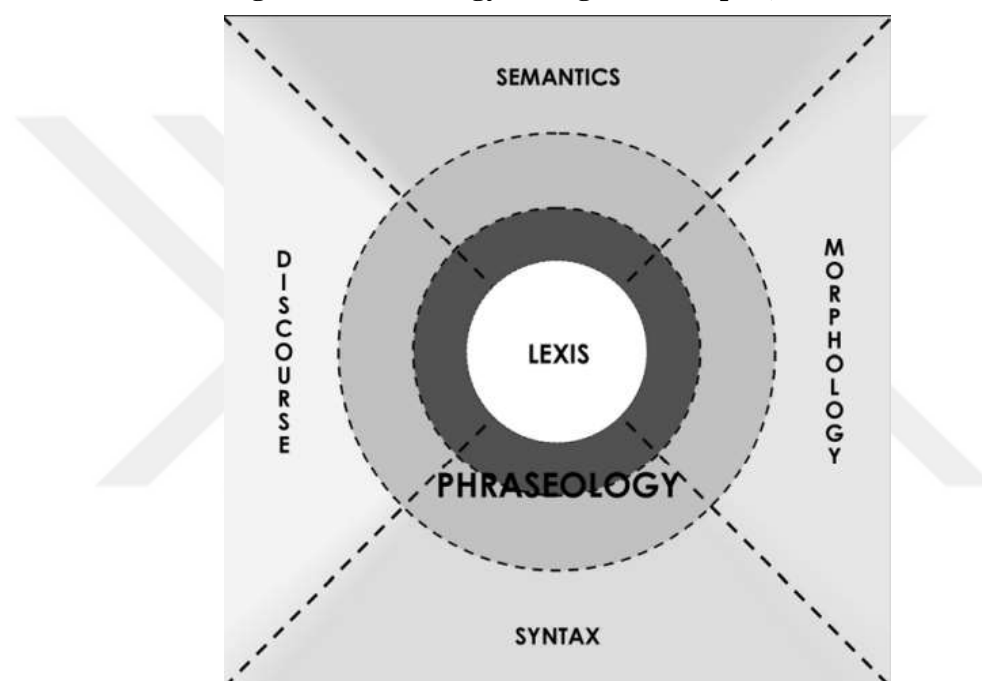
1.2.1. Phraseology Dimension in the Study of Language

This thesis study investigates collocations which are one significant fragment of phraseology. Phraseology is the study of word combinations which hold a noteworthy place in a language (Fontenelle, 1998; Granger, 1998b). It can also be defined “as the linguistic sub-discipline which deals with MWEs” (multi-word expressions) (Müller, 2011: 1). Gledhill (2000: 235) provides another definition for phraseology which considers it as “the preferred way of saying things within

a particular discourse”. The same researcher gives additional explanation for phraseology as “a dimension of language use in which patterns of wording (lexico-grammatical patterns) encode semantic views of the world, and at a higher level idioms and lexical phrases have rhetorical and textual roles within a specific discourse” (2000: 235).

The figure by Granger and Paquot (2008) reveals that phraseology is directly related to linguistic fields which can lead to the inference that phraseology carries linguistic studies under itself.

Figure 1: Phraseology-Granger and Paquot, 2008: 14



Phraseology holds three basic terms under itself which can be listed as idioms, fixed expressions and specifically significant for this research collocations (Cowie, 1998). Sprenger (2003) defines idioms as “the single words that make up the phrase and the semantic and syntactic information they entail do not play a role for the idiom as a unit” (2003: 42). This results in the inference that idioms are multi-word combination types which do not carry the dictionary meanings of its components literally; that is, in idioms every single unit has a dictionary meaning when examined separately, yet they create a totally different meaning when they are lined together. ‘Catch one’s eye’ can be given as an example of idioms as ‘catch’ and ‘eye’ do not carry their literal meanings when they are used together.

Another phraseological item, fixed expressions, is defined by Sprenger (2003: 4) as “specific combinations of two or more words that are typically used to express a specific concept”. Some examples of fixed expressions can be ‘all of a sudden, come to no good’. The literal meanings of

every single word in these expressions can help the language user infer the intended meaning. The important thing about the fixed expressions is that they are generally used together and this feature of them makes those units have fixed expressions.

Finally, the third part of phraseological units is collocations and they can be defined as two or more words that generally go together and form a multi-word unit (Palmer, 1933; Hill, 2000; Wray, 2002). With all these being said, phraseology can be defined as a linguistic division which deals with each form of word combinations that are collocations, fixed expressions and idioms (Granger & Paquot, 2008). Besides, those phraseological structures in a language are of importance since as Gledhill (2000: 1) argues, “it is impossible for a writer to be fluent without a thorough knowledge of the phraseology of the particular field he or she is writing in”. This statement can lead to the interpretation that phraseological competence assists the learner be more fluent and competent in the target language. Chomskyan linguists (Chomsky, 1965) assert that ‘competence’ comes with grammatically correct usages of the target language products, and these products can be gained best with the help of the intuitions of native users.

Phraseology should not be disregarded in a language study since it is strongly connected to corpus linguistics (Granger, 1998b) which is the base this study is built on. Müller (2011: 2) argues that “there is a close relation between corpus linguistics and phraseological research in usage-based frameworks, such as constructionist approaches to grammar, as these theories attach great importance to authentic corpus data”. In addition, the importance of corpus for the linguistic studies is argued by Altenberg and Granger (2001), Altenberg and Granger (2002) and Nesselhauf (2005) as multi-word combinations are better studied with corpus based research. Research carried out by different researchers (Moon, 1998a; De Groot, 1999; Cox, 2000) can reveal the importance of corpus-based research in phraseology (as cited in Sprenger, 2003) because they had the potential to itemize thousands of items including idioms, proverbs and sayings, and their usage frequencies.

1.2.2. Multi-Word Combinations and their Significance for EFL Learners

Before giving the definition of multi-word combinations it would be a better step to start with the ‘word’ which can be utilized in many ways. However, for this study, it will be elaborated only with its primary features. A word can represent an agent such as a ‘pencil’, a ‘wardrobe’, or a ‘mug’. In addition, it can be a prepositional phrase when it is used with prepositions; e.g. in the morning. Morning is the basic root and ‘in’ is the preposition which makes it a prepositional phrase. Besides, and most importantly for this study, a word can create a ‘combination’ when it is adhered to another word. To give an example, ‘living room’ is formed with the two words ‘living’ and ‘room’ which make them a combination.

Multi-word combinations are the phrasal patterns consisting of at least two words (Biber et al., 2004; Sinclair, 2004). In this present thesis, verb+noun combinations are examined. This kind of combination is the formation with a verb and at least one noun which go together to form a unit. They are considered as complex units since they create a compound group/syntactic entity when they are together (Sinclair, 1991). Having presented the definition of multi-word combinations, it is possible to revise the explanation of word combinations, namely combined words, as the phrases which co-occur together with other words.

Multi-word combinations; in other sayings, multi-word constructions, multi-word expressions or multi-word units have been of interest in language learning and language teaching as they are unsurprisingly significant parts of a language. Researchers named multi-word combinations with different labels such “as lexicalized sentence stem (Pawley and Syder, 1983), lexical phrase (Nattinger and DeCarrico, 1992), formulaic sequence and chunks (Wray, 2002; 1999), phraseology (Cowie, 1998), chunks (De Cock, 1998), pre-fabricated items and ready-made units (Howarth, 1998) or n-grams in computational linguistics and computer science (Manning & Schütze, 2008). No matter what their names are, they are considered as essential parts of a language without which a learner cannot gain native-like fluency and scholarly writing proficiency. Multi-word combinations have recently been of interest throughout the language teaching world. Besides, they are examined as a phraseological unit since they are parts of phraseology. It has been valued so much that there even are workshops about them at Stanford University. The educational institution has a significantly wide project called Multiword Expression Project. It even has a page including detailed information about the workers, publications, related research, and extensive materials which can be of great use for the researchers who are interested in this field of study.

Müller (2011) has a definition of multi-word combinations as they are “consisting of a minimum of two words, they cut across word boundaries” (Müller, 2011: 3). This description of multi-word combinations (MWCs) can be considered as the touchstone while defining the term. To give a simpler representation, they are formed with at least two words and when they are together they can change meanings infinitely. Erman and Warren (2000) present examples of multi-word combinations as idioms, compound nouns, collocations (which will be the mainly investigated units), prepositional and phrasal verbs.

Multi-word combinations can be defined by various statements; one other explanation of MWCs can be given by (Masini, 2005: 145) as “lexical units larger than a word that can bear both idiomatic and compositional meanings”. This is a clarified description which argues that when two words are together, they can form an idiom whose connotation cannot be inferred with the meanings of its elements, e.g. ‘to have other fish to fry’ which means to have a more important thing to do than the one occupied. McDevitt (2006) argues that MWCs have an assortment of types which include idioms; ‘a can of worms’, proverbs; ‘actions speak louder than words’, collocated

terms which generally go together; ‘to make a bed’, ‘to give advice’, and proper names such as ‘The Pacific Ocean’. These types display the common examples of multi-word combinations.

Above and beyond; when two words come along, the intended meaning can be provided by the lexical meanings of words and that means those MWCs have compositional meanings. Sprenger (2003), in his doctoral dissertation, comes to a conclusion that a compositional unit is formed with words whose dictionary meanings are employed so every single unit carries the lexical meanings of the words. These features of MWCs are of importance since, according to Masini, they are a “privileged domain for the study of the relationship between syntax proper (intended as free combination of words) and the lexicon” (2005: 145). To put it another way, the meaning obtained from the units can differ in the way they are used because the language users can use them in an idiomatic meaning or as a compositional figure.

For further definition, Lamidi’s (2017) argument can be consulted which defends the view that multi-word combinations are located in chunks everywhere they are used. Additionally, Sinclair and Renouf (1988) argue that MWCs have a significant place in English Language Teaching since they are of critical value (as cited in Özbay & Kayaoğlu, 2016: 342).

Multi-word combinations are defined by Biber, Conrad and Cortes (2004: 371) as “the most frequent sequence of words in a register” which yield to the meaning that a word is generally used with another word and that results in the formation of a multi-word combination. Furthermore, Lopez Rodriguez (2005) defends that “vocabulary has been conceptualized as individual words, but now it has become clear that lexis consists of sequences of words that function as a unit” (Lopez Rodriguez, 2005: 27), this view simply paves the way for the inference that combinations are made of chains of words which are utilized as units. This argument is supported by Müller (2011: 6) who asserts that MWCs are potential lexical units too rather than as being lexical by definition.

Having provided a detailed definition of MWCs, this study is concerned with multi-word combinations since some certain words mostly co-occur in definite structures. As Wray (2002) argues, these structures are named as MWEs, in this study MWCs, and this name covers linguistic units because they are more than single words; they are lexical units. As stated by Müller (2011), these lexical units are pre-formed so the language user does not have to create a new one, she retrieves the already present phrases from her mental lexicon.

A significant part of a language is formed by multi-word combinations which can be listed as collocations, idioms, prepositional verbs, phrasal verbs and other groups of words and units. Therefore, studying and examining multi-word constructions would be a noteworthy section in a language study. This study of tertiary level EFL learners’ usage of these combinations would be of

great interest as students from different departments of English language are observed in terms of their usage of word constructions and specifically collocations.

Multi-word combinations are essential in language since they help the language user be confident and have extensive information while employing them. To put it differently, fluency can be formed with the broad employment of multi-word combinations (Lopez Rodriguez, 2005). When MWCs are pre-formed as they are recognizable units, they trigger a secure conversation because they make the language clear and understandable. It makes the language usage understandable because the already formed chunks are easy to remember and ready to use (Wray, 2002). As Lopez Rodriguez (2005) claims, MWCs are important for non-native speakers because they are characteristically related to practical usage, so they enable the speakers to be more proficient and to have clearer understanding by others.

When the question “why are MWCs important?” is highlighted, it can be answered with various explanations. The language user knows that vocabulary is not just formed with single words; they can be in a multi word form. Besides, the most common chunks are important in spoken language since they make the speech more natural, fluent and successful as they are already formed. This makes the language user confident since, as Wray (2000) argues, she does not have to create the chunks whenever she needs them. To conclude, ready-made combinations make the speaker more fluent and confident because the language user does not always have to think of new units while speaking, so she may employ the prefabricated constructions to be a successful learner of the target language.

Multi-word combinations, as mentioned above, have a significant place in language usage because when the user has ready-made chunks in her mind it makes her recall the units quickly and therefore be more fluent. These chunks are to become predictable to the language user if she is taught multi-word constructions; that is to say, as Nattinger notes, we can realize “the importance of prefabricated speech routines in language behaviour” (1980: 337) when the user is confident and proficient in using the language.

1.2.3. Collocations and their Significance for EFL learners

Having given the definitions of MWCs above, collocations are to be explained. MWCs are formed with collocations, idioms, phrasal verbs, etc. In this study, collocations are examined. It would be wise to start with the definition of collocation and its importance in language usage. According to the definition of Collins Cobuild English Dictionary (1995: 309), “collocation is the way that some words occur regularly whenever another word is used”. Additionally, Granger

(1998b) defines a collocation as “the linguistic phenomenon whereby a given vocabulary item prefers the company of another” (1998b: 146).

Compositional features of collocations are of need following their explanation. Collocations are made of two components named bases and collocators. Hausmann (1989) defines the terms in this way; while the base is the semantic nucleus, the collocator is the modifier of the base. For example, in a noun+verb collocation a bird flies, here what is talked about is bird, which makes it the base. The verb ‘flies’ is the collocator for providing information on what the bird does. In brief, collocations are two or more words with a high co-occurrence possibility (as cited in Nesselhauf, 2005). In this thesis, verb+noun collocations are studied; therefore, the bases are the verbs and the collocators are the nouns.

In addition to the definitions above, collocations can be shaped with the “recurrence of two or more words in a way more than arbitrary” (Demir, 2018: 293) and they are employed frequently in scholar usages. Müller (2011:3) provides a very simple and brief definition which makes it easier to understand; collocations are “fully compositional, non-idiomatic MWEs (e.g. fish and chips)” This explanation offers the understanding that collocations cannot be idiomatic. As mentioned above in the phraseology section, idioms are one of MWC types whose every single unit has a dictionary meaning when examined separately but when they are together they have a completely different meaning. As an example, the idiom ‘Sisyphian’s task’ can be examined to provide a better understanding. Sisyphus is a mythological character who is punished by gods with the sentence of pulling a huge rock to the top of a mountain; however, the rock falls down the mountain every time it is rolled up to the summit. That means, Sisyphus’s efforts are in vain although he is in great struggle. To sum up, ‘Sisyphian’ term relates to Sisyphus who is a mythological character and the word ‘task’ means a mission or a duty. Yet, when these words are bundled together they create a diverse meaning which refers to a work that needs an endless and a worthless effort. This idiom is also used to refer to the absurdity of life, and that is the basic of absurd theatre. Absurd theatre is based on Sisyphus myth and he is an absurd hero. It has common features with Samuel Beckett’s *Waiting for Godot* which presents the senselessness of life by referring to the idea that in spite of all the efforts and works, nothing is going to be gained. In short, the idiom Sisyphian’s task stands for the irrationality and insignificance of life although both units reveal different meanings when they are examined compositionally. Collocations include criteria as non-compositionality, non-substitutability, non-modifiability, non-translatable word for word (Yoon, 2016; Ackermann and Chen, 2013; Fawzi, 2006; Manning and Schütze, 1999).

When it comes to the description of non-compositionality of collocations, it would be possible to define compositional phrases as the units whose meaning can be inferred with the help of its parts; an example can be ‘a new laptop’ (Fawzi, 2006). Additionally, collocations hold

limited compositionality; besides, non-compositionality includes idioms as the strongest examples; “break a leg” which has nothing to do with the literal meanings of each single part.

The non-substitutability feature of collocations presents the inference that it is not possible to substitute synonyms to create a similar meaning (Ackermann and Chen, 2013). For example, strong and powerful are sometimes used interchangeably since they have similar meanings; however, a “strong tea” collocation cannot be reiterated as “powerful tea” since collocations do not let synonyms be in the charge of the real meaning of the words.

Collocations are non-modifiable which means that not all collocations can be modified easily by only adding some grammatical formations; for example, whereas “white wine” is acceptable “whiter wine” is not acceptable since it includes grammatical reformation (Manning and Schütze, 1999). In addition to the abovementioned features of collocations, non-translatability of collocations is another issue to be focused on. Collocations seem to be generally non-translatable; furthermore, it is agreed that if a word group is unable to be translated word by word, it shows that those words are the components of a collocation (Manning and Schütze, 1999).

Having mentioned the criteria of the collocations, it would be necessary to display their subclasses which are light verbs, verb particle constructions, proper names and terminological expressions (Manning and Schütze, 1999).

Light verbs are the ones “with little semantic content like make, take and do are called light verbs in collocations like make a decision or do a favor” (Manning and Schütze, 1999: 186). Verb particle constructions are the units in which some verbs go with some fixed prepositions such as “go down, tell off”. Proper nouns are another subclass of collocations an example for which is “The Pacific Ocean”. The last subclass is the terminological expressions; as the name suggests, they are “the concepts and objects in technical domains” (Manning and Schütze, 1999: 186), and an example is “clutch pressure plate”.

Language is systematic in itself, and to be a fluent and a proficient user, one needs to have competency in that systematic nature of the language (Cowie, 1998; Wray, 2002; Schmitt, 2004). In other words, a language has pre-fabricated collocations which are ready to use and which make the learner have native-like proficiency. However, non-native users of languages do not have full knowledge of collocations and this causes them to be timid and insufficient both in speaking and writing. The significance of collocations cannot be denied since they are tied closely to language competency. In addition, Nation (2001), states that a learner should have competency in collocational usage since they are “one important aspect of vocabulary knowledge” (2001: 529). Similarly, Zareva, Schwanenflugel and Nikolova (2005) carried out a research which revealed the result that the proficiency of a language learner is directly related to her lexical and collocational

knowledge. The study also introduced that the more collocations the learner knows and uses, the more competent and proficient she will be in the target language.

How collocations let language users be proficient can best be explained by the formulaic chunks which Brown (1974) believes are of utmost importance. Additionally, Wray (2002) argues that the richer the collocational knowledge, the more proficient the language user is. These collocational usages, in this study verb+noun collocations, lead the language learner to have a distinctive capability because she already has them in mind and can recall them when necessary. These pre-prepared chunks; that is to say, collocations let the user think like a native. Employment of collocations assists the learner to prevent L1 (mother tongue) transfer by giving them the chance of using pre-prepared combinations which are already in mind. This positive situation creates a more scholar writing and a more fluent speaking.

Multi-word constructions or combinations, as mentioned above, play a significant role in language competency. As Granger (1998b:145) states “since the mid- 1980s, the use of prefabricated language ('prefabs') has become a major focus of interest in EFL”. Collocations, as an important part of MWCs have a noteworthy place in proficiency as well. The more collocational usage is in, the more competent is the user. As Chomsky (1965) argues, native speakers have intuitions for not recalling what they learned, and this leads to competency. Therefore, being nativelike goes through the way of competency. After the research he carried out with 104 students, Nizonkiza (2011) found out that proficient use of collocations paves the way for a more competent and fluent use of target language. As Hill (2000) states, collocation is a vital factor for fluency which leads to the inference that we cannot have a native-like fluency without being able to use collocations.

Additionally, it was put forward by Sadeghi (2009) that a proficient language user should not be withdrawn by her native language as a result of her insufficient competence. Negative transfer or native language interference may create problems in target language usage, so acquisition of collocations are needed in order to prevent this.

Moreover, when collocations are not used properly they can cause miscommunication or misperceptions, and these end up with failure in language use. Vincze (2008) supports this view with her statement “in natural language processing, one of the most challenging task is the proper treatment of collocations, which term comprises multi verb constructions as well” (as cited in Özbay and Kayaoğlu, 2016: 342). This expression reveals the need for correct teaching and proficient use of target language; to put it in a different way, a learner can be a competent user of foreign language she is acquiring or learning if she is aware of the importance of collocations and if she acts accordingly. This is agreed by many researchers, and Howarth (1998) is one of them. The researcher argues that if a language learner wants to be native-like, she should know the value of

collocations in language and should be careful while using it to be a skilled user. Therefore, it is possible to argue that collocations are of great help for teaching and learning vocabulary, as Lewis (2000) drew attention.

As Hill (2000) summarizes fluency comes with the use of collocations. Native and fluent speakers of languages basically have ready-made collocations in mind. However, as Erman and Warren (2000) argue, non-native users can gradually broaden their knowledge when they are more into the language. They can acquire new pre-prepared combinations although the process may be “limited in dispersion and entrenchment” (Erman and Warren, 2000: 33). In other words, non-native user of a language can increasingly improve her proficiency by acquiring more and more collocations which are already existent in native languages. Woolard (2000) argues that teaching multi-word combinations is of help during the improvement of a learner’s proficiency. As a summary, collocations are appreciated for language learners, because they pave the way for the user to be a native-like speaker and a writer.

The innate collocational knowledge and proficiency in native speakers is crucial in the language as natives can use idioms, phraseological groups, and the language fluently without contemplating too much. This reveals the competency and proficiency they have. Therefore, it can be argued that natives and non-natives differ in their collocational usages. In order to prevent the lack of collocation use, language learners should seek out gaining proficiency and competency to be native-like (Stubbs, 2001; Nation, 2001; Coxhead, 2000; Wouden, 1997). This common view of the abovementioned researchers is supported by Martynska (2004: 11) with a very simple defence which says “consequently, if learners’ sensitivity to various relations between words is not heightened enough or words are not learned in chunks, learners are not bound to approach the native-like level of proficiency”. Martynska (2004) additionally concludes her discussion with the sentence, “the richer in collocations the learner’s lexicon is, the higher precision, accuracy, coherence and authenticity of his/her speech. This is a perfect way to fluency and proficiency in the language as well as to greater language competence” (Martynska, 2004: 11-12). In addition to the abovementioned importance of collocations, Nizonkiza (2011: 113) revealed after his research that “mastery of collocations is found to be related to frequency and to predict lexical competence”. In brief, collocations, like other word units in a language, are of importance and should be taken into more careful consideration since they constitute a part of a language without which it would be lacking fluency and native-like property.

Collocations can be studied under the names of lexical and grammatical collocations if more detail is required. These two titles differ in terms of the formation; that is, a grammatical collocation can consist of verb+noun, noun+verb, preposition+noun and bare verbs. On the other hand, a lexical collocation can be created by adjectives, nouns, prepositions and adverbs with their different formations (Benson, Benson, and Ilson, 1986). Lewis (2000) and Benson et al. (1986)

presented two groups for collocations as lexical and grammatical collocations. The difference standing between them is that grammatical collocations hold grammar words whereas lexical collocations do not have that feature.

Table 1: Collocation Patterns

| Type | Example |
|-------------------------|--------------------------|
| Adjective+ noun | bright color |
| Noun + verb | the economy boomed |
| Verb+noun | submit a report |
| Noun + noun | a sense of pride |
| Adverb +adjective | happily married |
| Verb + adverb | smiled proudly |
| Verb + preposition+noun | filled with horror |
| Verb+adjective+noun | revise the original plan |

Reference: McCarthy and O' Dell, 2005: 12; Lewis, 2000: 133 (As cited in Özbay's doctoral dissertation, 2015: 73)

1.2.4. Collocations in Vocabulary Learning

Why do collocations help learners in vocabulary acquisition? This seems like a critical question; critical because the aim of choosing collocations as the thesis topic is to see how Turkish learners of English can or cannot use collocations properly by comparing them with a native corpora called LOCNESS. To put it differently, being able to appropriately use multi-word combinations makes a learner more native like and more fluent which brings more competence and self-confidence while utilizing the language both in academic and daily life. As Wei and Lei (2011) argue, collocations take a significant part of scholarly writing, and they also state that a language user who is not skilled in collocational usage definitely lacks native-like fluency. A supporting result comes from Laufer and Waldman's (2011) research in which three proficiency levels of native speakers of Hebrew were examined. The study found "that the number of collocations increased only at the advanced level, and that errors continued to persist even at advanced levels of proficiency" (Laufer and Waldman, 2011: 647). This undoubtedly can reveal the importance of collocations in target language usage. These do not mean that collocations are only advantageous for writing, they are of prodigious support in other language skills such as listening, speaking and reading without which a learner is not fully competent.

There is an intense relation between collocation units and native-like fluency in a foreign language as collocations prove that the non-native user of the language has a brilliant mastery in the target language. The more collocations the learner uses; the better knowledge is considered to be possessed by him. Thereby, foreignness is left behind as the user has the control of the language

in terms of collocations, therefore the language itself. Being able to use collocations in a brilliant way paves the way for being read like a native speaker of the target language, and therefore being respected as a scholar when used in academic writings. Cowie (1998:1) argues “that native-like proficiency in a language depends crucially on a stock of prefabricated units -- or 'prefabs'”, which can result in the inference that one needs to have collocational competence in order to be a native-like user of the target language.

Granger (1998b) employed a research to figure out the differences in collocational usages by native and non-native speakers of English. The non-native participants were 56 French learners of English, and the other 56 participants were native speakers of English. They were compared in terms of their collocational usages of adjectives. The results showed that non-native users of English utilized much fewer collocations than the natives. The rate of non-native collocations was almost the half size of the native collocations. For instance, the collocation ‘readily available’ was used by 8 non-natives whereas 43 natives utilized it. Additionally, ‘bitterly cold’ was chosen by 7 non-natives while 40 natives preferred it. These results can lead to the conclusion that non-native speakers of English did not have a strong sense of salience; that is, the adverb+adjective combinations were not salient in non-native usages as much as the native ones.

Natives are dominant in utilizing the collocations intuitively whereas the non-natives make a lot of effort to acquire them. That is, when a non-native user has the capability of writing with collocations, it is ascertained that s/he paid invaluable attention and cared about the language units in order to sound as natural as possible.

1.2.5. Interlanguage

This present study researches interlanguage process in non-native students’ argumentative essays by utilizing KTUCLE and TUCLE which are non-native learner corpora. A language learner is considered to be native-like when she becomes indistinguishable from the natives of that language. The learners and native speakers of a language differ in many ways; one of them is sentence production which needs clarified understanding both in speaking and writing. When a learner has native-like proficiency, she is more confident and clearer in communication. However, before having that competency, a learner goes through a language in which she makes mistakes and corrects them as time passes and as she acquires the language. Yet, sometimes those mistakes can be fossilized no matter how much exposure to language occurs.

The abovementioned process can be named as the learner language. It was previously labelled as “idiosyncratic dialects” by Corder (1971: 148), next referred to as “approximative systems” by Nemser (1971: 115), and finally as interlanguage by Selinker (1969) whose term has been in more common use. This learner language; or as Selinker (1969) names ‘interlanguage’

hereafter IL. He calls this as “a learner's attempted production” (Selinker, 1972: 214) which briefly argues that this term refers to an imperfect language behavior since it is only an attempt in the target language. The term interlanguage by Selinker (1972) was devised with the purpose of implying that this learner language is not pure target or native; it is hybrid language. This hybrid term refers to the meaning that interlanguage is somewhere between the native language and the target one. For Selinker (1972) interlanguage stands between the mother tongue and the target language. Larsen-Freeman and Long (1991: 130) define this interlanguage terms as “the language system that the learner constructs out of the linguistic input to which he has been exposed”. As an addition to the abovementioned definitions, Granger (2003: 127) puts forward another description for interlanguage as a “variety in its own right, which can be studied as such without comparing it to any other variety. With this in mind, however, it may be useful to compare it to other language varieties in an attempt to reveal its characteristics for better understanding the internal structure of it”.

Interlanguage can be expressed as “the systematic language knowledge independent of both learner’s native language and the target language making a continuum between them” (Lenard et al., 2015: np). Interlanguage “refers to a language intermediate between the native and the target language” (Lennon, 2008: 55). Trawinski (2005: 54) defines interlanguage as “the language system which the learner constructs in the process of SLA”. These definitions can be interpreted in the way that interlanguage is neither a mother tongue nor a perfect version of target language.

During the process of acquiring or learning a language, a learner tends to misuse some patterns of language, which later can be fossilized. According to Selinker (1972: 215), “fossilizable linguistic phenomena are linguistic items, rules, and subsystems which speakers of a particular NL will tend to keep in their IL relative to a particular TL, no matter what the age of the learner or amount of explanation and instruction he receives in the TL”. A learner may go through this misuse process; however, when she is not able to overcome those misuses, they become fossilized which undoubtedly reveals the feature of interlanguage. Selinker (1985, 1975, 1972, 1970, 1969) and Selinker and Lakshmanan (1992) defend that interlanguage is a mixture of the learner’s native language and her target language.

Interlanguage is considered as the learner language since it is actually the path the learners go through while they are trying to reach the target language in which they are to be competent enough to write scholarly. Therefore, it would not be unacceptable if the term is simplified into the phrase ‘learner language’. The continuum, namely, the learners’ language learning process can be regarded as an imperfect variety of the target language which needs to be acquired well enough in this context of writing academically. Interlanguage can be approached as imperfect just because only the natives of the target language have innate capability of the perfect variety.

To put it more explicitly, interlanguage refers to the non-native language of the learners with its mistakes and errors. When the learner is competent in the target language, it shows that s/he is successful in her/his interlanguage if s/he can convey what s/he has acquired or learned in an understandable way. Contrarily, the learner fails in her/his interlanguage when s/he cannot express what s/he has in mind. However, this success and failure are both key elements of interlanguage since it is what the learner has all the way through the dominance in the target language.

The non-native language user's language (IL) is an obstacle on the way to a native-like proficiency because it bears limitations within itself which create simplicity and uneasiness during the output production. For that reason, the language may sound ill-formed, impoverished, poor, deficient and of inferior quality. Improvement in interlanguage carries fluency and competency in both speaking and writing abilities with it. Grammatical proficiency which is a result of interlanguage development comes along in the output of the language user.

During the learning or acquisition of a language, the non-native user tends to employ incorrect usage of the target language. Namely, the incorrect language is the characteristic of the interlanguage. Interlanguage is formed by learners' mistakes and errors whose greatest example is first language transfer. Besides, learners employ simplification and generalization which result in applying the same rules to similar situations. This, unfortunately, creates failure in interlanguage. The reason is acquiring the language does not always come with its full and correct rules; on the contrary, learning comes with incorrect usage. Language goes through mistakes and errors which are the indicators of learning. To sum up, erroneous usage proves the naturalness of language development; because, the user will utilize his mistakes in order to learn the language.

In addition to all the above-mentioned explanations, the non-native user of language can improve himself if he is aware of the errors he makes after his teacher's feedback. The user can start to correct his mistakes and can create the right forms following the teacher or the instructors who help him achieve native-like proficiency.

The importance of interlanguage in EFL settings can be explained by Corder who states that non-native learners of target language that have the same mother tongue "speak more or less the same interlanguage at any point in their learning career" (1982: 20). This similar process they go through is of help for researching same level students in their verb+noun collocations; how they differ or how similarly they employ collocational usage.

Corder (1982: 57) expands his idea of the importance of interlanguage with the belief that the learner cannot be examined directly; therefore, the products of them can be utilized to infer a comparison. The products, in this present study argumentative essays, are investigated to compare same level students as they are taught the same syllabus. The same syllabus can refer to the input

they receive; thus, the comparison between TUCLE and KTUCLE data is likely to reveal the output differences the non-native learners display. This learner language version is named as interlanguage by Selinker (1972) and this term of his is used more commonly than the other terms mentioned above (Corder, 1982: 66).

The purpose of utilizing interlanguage process for this thesis is that the non-native argumentative essays “may quite regularly exhibit systematic properties which show no obvious resemblance to the mother tongue or any other language known to the learner” (Corder, 1982: 72). The importance of interlanguage is expressed by Zhang (2013: 551) as the term refers to “psychological structure latent in the brain” that goes active when the learner tries to learn a foreign language. Selinker’s definition for interlanguage as “a learner’s attempted production” (1972: 214) overlaps with this idea of Zhang (2013) in the way that the learner attempts to learn a second language during which the brain is activated.

The imperfectness of interlanguage can be considered necessary for language learning process, because as Ho (2003) argues errors are the indicators of learning, and therefore errors are inevitable for learning. Thus, this can lead to the inference that interlanguage process includes errors; that is to say, learners go through a way which owns errors in it and this is a necessary way to pass in order to acquire target language. By the same token, Hosseini and Rahmani (2015: 409) believe that interlanguage plays a significant role in language learning since it assists the teachers and learners to “fulfill the required necessities to achieve sooner, and in better manner, the destination of language learning”. Garcia (2006: 171) asserts that interlanguage is necessary as it “can help teachers and syllabus and ELT (English Language Teaching) material designers to pinpoint problem areas for second language learners”. In other words, interlanguage is necessary to see the drawbacks and positive features of the learners during the acquisition and learning processes. Therefore, interlanguage can be considered to be a research field which attempts to provide information for more proficient and competent language learning and teaching.

Popescu (2007: 182) describes interlanguage as “a third language, with its own lexicon, grammar and discourse structure, phonological traits, etc”. Therefore, it can be inferred that an interlanguage process goes through a unique language structure which later on closes to being native-like. However, there are some language transfer processes by which interlanguage is formed which are “negative transfer, positive transfer, avoidance, and overuse” (Popescu, 2007: 182). This present thesis study focuses on overuse and underuse which are named as avoidance by the abovementioned researcher, of the verb+noun collocations during the interlanguage process. This view of the abovementioned researcher is supported by Asikin (2017) who argues that there will be interferences of the mother tongue during the production of the target language.

When interlanguage is investigated in Turkish context, it can be seen that interlanguage corpus studies focused on written works (Can, 2009; Şanal, 2007; Kilimci, 2001), which is also a similar feature with this present thesis. As for foreign context, Navés (2006) lighted the way for implementing interlanguage studies and followed by Ruegg (2010) and Wong and Teo (2012) who also carried out interlanguage studies by investigating students' written products by both unfocused feedback and imitation.

Kil (2003) studied three kinds of errors in the products of five Korean learners of English in terms of word order, inversion and co-occurring articles. The study yielded to the result that the errors the students made were indicators of learning; that is, the students were in a process of improvement. This improvement of those Korean learners of English showed that they were in interlanguage process (as cited in Asikin, 2017). Another research carried out by Fauziati (2003) aimed at error finding in English course books for secondary schools. The results showed “that the English used in the textbooks is still at the level of IL. The features of systematicity, permeability, and fossilization of an interlanguage also existed in it” (2003: 179). In addition to the studies mentioned above, Luna (2010) similarly conducted a study which concentrated on the interlanguage of 27 university students in terms of their academic English. The study was carried out with Spanish undergraduate students' written products with the intention of discovering common linguistic patterns. The study revealed that interlanguage is of help for gaining “complex, unique and rich linguistic environment from where teachers can withdraw students' weak areas of development in L2” (Luna, 2010: 71). She also states that errors “are not perceived as something negative but as a construction of their own process of learning” (2010: 71).

Having mentioned the related studies, it would be of need to revise what interlanguage is; it is the form of target language which is neither perfect nor inadequate. The users of interlanguage do not have native-like competency and they occasionally employ L1 transfer during the production of the target language. Having mentioned the feature of interlanguage, the reason for utilizing interlanguage in this present study lies on this logic; in other words, the EFL learners' target language use is the basic of this research. The participants in this research are EFL students and their exposure to the target language is insufficient, so this creates a suitable atmosphere to research overuse and misuse of verb+noun collocations. The EFL students' collocational usages are compiled in this study with the purpose of comparing them to native corpus.

1.2.6. Contrastive Interlanguage Analysis

Interlanguage is “a learner's attempted production” (Selinker, 1972: 214) which yields to the meaning that the language used by the non-native users of the target language is inadequate or unsatisfactory. After presenting the definition of interlanguage, contrastive interlanguage analysis, hereafter CIA, is investigated in this chapter of the study. For Granger (1998b: 12), contrastive

interlanguage analysis differs from the traditional contrastive analyses which compare different languages; in other words, CIA works for comparing and contrasting native and non-native speakers of a language. CIA includes comparison of two major types:

- (1) NL vs. IL, i.e. comparison of native language and interlanguage;
- (2) IL vs. IL, i.e. comparison of different_ interlanguages (Granger, 1998a: 12).

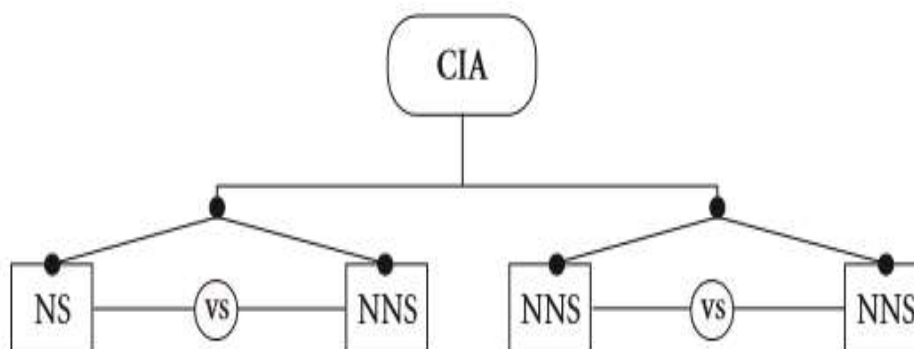
The comparisons of native language and interlanguage aims at discovering “the features of non-nativeness of learner language” (Granger, 1998b: 13). The importance of CIA for this thesis study depends on the features of not “plain errors, but differences in the frequency of use of certain words, phrases or structures, some being overused, others underused” (Granger, 1998b: 13). This investigation refers to the examination of “quantitative and qualitative comparisons between native language and learner language” (Granger 2009: 18) since quantitative usage signifies overuse and underuse of the specific patterns in a language whereas qualitative usage represents the misuse of the target language.

Granger (2002: 12) provides the feature of Contrastive Interlanguage Analysis as it is comparison of native and non-native learners’ works and it “consists in carrying out quantitative and qualitative comparisons between native (NS) and non-native (NNS) data or between different varieties of non-native data”. This corpus-based contrastive interlanguage analysis of the argumentative essays of the EFL learners in two different departments of tertiary levels assisted us in figuring out the qualitative and quantitative usages of collocations; in other words, misuses and over and underuses of the verb+noun collocations. Besides, it helped us comparing the non-native corpora with the purpose of presenting the similarities and differences in the EFL learners’ usages of collocations.

This present thesis study works on two learner corpora one of which is compiled by me with the purpose of examining the collocational competences of Turkish EFL students by employing the Contrastive Interlanguage Analysis. Non-native users of English tend to employ incorrect collocational usages and disregard the idiomatic nature of writing in English seems valid in Sinclair’s (1991) “Idiom Principle” which argues that language users apply pre-fabricated patterns or idioms intuitively. Therefore, CIA helps the research be more treasured since it will display the collocational usages of non-native tertiary level learners of English by employing CIA based corpus research.

This figure below represents the basic feature of CIA which can both compare native and non-native works, and non-native and non-native works.

Figure 2: Contrastive Interlanguage Analysis

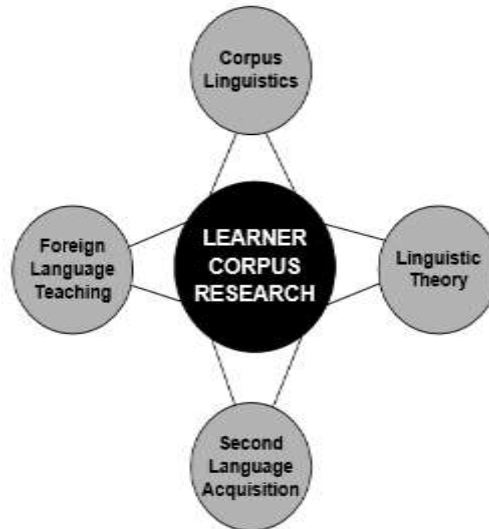


Reference: Granger, 2002: 12

Utilizing learners' works for a linguistic study requires contrastive interlanguage analysis which can compare varieties of a language, whether they are native or non-native works. Within the scope of this, employing a contrastive corpus analysis will assist us to grasp the proficiencies of non-native students in their collocational usages and the developmental stages they go through. Two non-native corpora, beside the native corpora, are investigated in this study as Granger (2002) asserts, the investigation of two different interlanguages provides the researcher to examine whether the foreignness features only belong to one single language group or varieties of different learner groups.

According to Granger (2009: 2) "learner corpora (LC) are electronic collections of foreign or second language learner texts assembled according to explicit design criteria. The fact that they contain data from language learners makes them a very special type of corpus, requiring from the analyst a wider range of expertise than is necessary for native corpora". She also provides a diagram about learner corpus which is directly related to corpus linguistics, linguistic theory, second language acquisition and foreign language teaching. The figure appeals to the meaning that researchers should be able to have linguistic expertise, a background in linguistic theory, knowledge of second language acquisition and understanding of foreign language teaching.

Figure 3: Core Components of Learner Corpus Research



Reference: Granger, 2009: 2

This thesis study makes use of native and two non-native (learner) corpora. One of the non-native corpora (TUCLE) is formed by my students' argumentative essays. TUCLE stands for Trabzon University Corpus of Learner English whereas the other non-native corpora KTUCLE stands for Karadeniz Technical University Corpus of Learner English. I have been teaching writing in the preparatory class at Trabzon University English Language Teaching Department, and it gave me the possibility to compile my own corpus after getting the students' consents. The similarities among the three corpora lie in the same type of genre and equivalent ages and experiences of the writers. The corpora include argumentative essays written by the approximately same age learners and users of English.

The CIA analyses interlanguages to determine the areas where learners have difficulties. Additionally, it can compare the interlanguage with native speaker language with the purpose of investigating the problems non-natives face in details (Lado, 1957, as cited in Özbay's doctoral dissertation, 2015: 7). Therefore, corpus based language studies are of vital help for further investigations and amendments in language teaching environments. By utilizing a corpus-based contrastive analysis, interlanguage and native language can be compared in an attempt to figure out the outstanding features between them (Granger, 1998b, 2002; Altenberg and Granger, 2001; McEney and Kifle, 2002).

Employing CIA when studying with the help of a learner corpora provides the advantage that interlanguage is compared with an entirely explicit type of target language standard (Hunston, 2002). When the non-native written works are examined with the help of corpus-based Contrastive Interlanguage Analysis, as Cobb (2003) states, the researcher is likely to comprehend the acquisition stages of learning the target language. Therefore, for Meunier (2002) it would be

possible to grasp the acquisition nature of the target language in addition to the development of curriculum design and pedagogic materials. In addition to understand the acquisition process, Contrastive Interlanguage Analysis helps the researchers perceive learner errors, strengths and weaknesses to figure out the common drawbacks the learners face during the interlanguage process. It also provides the researcher to observe their performances in both languages they use.

The strength of the native corpus comes from its explicit features; that is, the accurate and correct usages (Bloomfield, 1984) of the native speakers make it a useful base to research any linguistic features and to compare those features with non-native corpora. Chomsky (1965: 24) reveals the importance of the language when it is applied as the mother tongue: “A grammar is ...descriptively adequate to the extent that it correctly describes the intrinsic competence of the idealized native speaker”. This idealized native speaker utilizes the perfect version of the language, and the non-native speakers utilize their interlanguage during their developmental process. When a native corpus such as LOCNESS is compared to non-native corpora such as TUCLE and KTUCLE, the former refers to the ‘norm’ which means it will be the base to be compared with. In the light of these, Contrastive Interlanguage Analysis works on comparing the native and non-native corpora to reveal the ‘nativeness’ and ‘foreignness’ in the works produced.

Taking a native corpus as a norm to compare it with non-native corpus seems as a reasonable step as the native corpus is correct and accurate form of the target language (Bloomfield, 1933:43). Therefore, the native corpus is a reliable base to compare other forms of languages used by non-natives. In this present research, verb+noun collocations formed with ‘*do, make, have, give, take, get and go*’ by the non-natives (TUCLE & KTUCLE) are compared to the ones composed by natives in order to realize the differences in misuses and over and underuses of the collocations. Corpus-based Contrastive Analysis is the basis of this present study; it investigates native and non-native collocational usages with the help of three different corpora. The study examines verb+noun collocations employed in argumentative writings of native and non-native users of English.

Utilizing corpora for studying differences among three groups of speakers of English provides the researcher to grasp the developmental process of the non-natives. In addition; the common or consistent misuses, and overuse and underuses of verb+noun collocations are provided for the researchers and linguists to amend the existing teaching methods and systems. According to Kennedy (1998), corpus studies can also assist learners to process inductive learning; put differently, non-native users of a target language are likely to control their own improvements in learning and acquiring the language. This situation makes learners self-controlling; in other words, learners of a target language tend to examine their linguistic data, form conclusions for themselves and therefore be researchers in linguistics. This self-controlling and self-development can pave the way for fresh and up-to-date learning methods to employ.

Aside from the self-control and development employed by learners, corpus-based contrastive interlanguage analysis helps researchers to assess target language proficiency by utilizing statistical values (Altenberg and Granger, 2001). In brief, contrastive corpus-based analysis lets me compare native and non-native corpora to figure out similarities and differences of three different groups in terms of their collocational usages. The findings of this research are likely to pave the way for revealing the subjects or areas where students have most difficulty, and therefore teachers and instructors of those groups can pay more focused attention to the challenging parts during teaching their learners of English.

1.2.7. Foreignness

Foreignness, as the name suggests, is a sign of sounding incompetent and insufficient in a target language. Granger (1998b) state that learners of a target language do employ usage of collocations; however, their rate is so low that it sounds non-native and therefore it creates foreignness in target language usage. She argues that non-natives “use atypical word combinations” (1998: 152) which can refer to the meaning that atypical usages sound underdeveloped language proficiency. Similarly, as McArthur (1992) argues, collocations are necessary elements for being native-like; when one is not competent enough in collocational usage, she reveals foreignness which is definitely unwanted in scholar writing.

Linguist researchers (Lin and Fang, 2009; Siyanova and Schmitt, 2008; Erman, 2009; Granger, 1998; Howarth, 1998) found in their research that non-natives are weak in using collocations in their written works; that is, they underuse collocations and when they tend to be more active in applying collocations into their works they seem to sound foreign which leads to the inference that non-natives need to train more on collocational usage in order not to reveal foreignness.

This research employs Contrastive Interlanguage Analysis with the purpose of comparing one native and two non-native corpora to display the foreignness in the argumentative essays produced by tertiary level students in two different universities in Turkey. The EFL learners are not natives; therefore, the essays do not read native. In the light of these, this study will reveal the foreignness in the two non-native corpora. Following the revelation, new amendments can be pursued because in order to prevent and avoid foreignness, new methods of teaching vocabulary and collocations can be formed based on the results of this present research since as Granger (1998b: 158) states, these new methods can prevent learners from “foreignsounding”. This present study works on collocations and it is argued by many researchers (Laufer & Waldman, 2011; Nesselhauf, 2005; Siyanova & Schmitt, 2008; Yoon, 2016) that collocational usage is of help for forming a competent language; however, second or foreign language learners sound abnormal and foreign when they are not fully-competent in collocation use.

1.2.8. Rationale for Researching Writing Skill

Language is composed of four main skills as listening, speaking, reading, and writing. This study is based on non-native students' writings. The reason why writing skill is chosen can be strengthened by Asikin's (2017) statement which argues that doing research in "students' writing is considered important" and "writing is considered as the hardest skill for L2 learners" (Asikin, 2017: 47). Universities, high schools, and other educational programs value academic writing, so it has become necessary to examine writings in terms of their grammatical characteristics (Hinkel, 2003); the organizations of papers such as the topic sentences, supporting ideas, main ideas and concluding sentences; "strength and justification of claims" (Allison, Cooley, Lewkowicz & Nunan, 1998: 207). In other words; for being able to scholarly write, Hinkel (2003) takes a stand for grammatical competence whereas Allison, Cooley, Lewkowicz and Nunan (1998) are for physical organization of the paper.

For this thesis study, writing is taken into consideration since it is accepted as the most difficult and complex skill in a language as one cannot use gestures or emoticons in the written product (Asikin, 2017). The writer has to clearly deliver her message without the assistance of those kinds of obvious helpers. Therefore, academic or scholarly writing are a definite signal of competent learners. For a learner to be proficient, she needs a wide vocabulary, specific grammatical abilities, syntactic features, and divergent properties of language (Wolfe-Quintero, Inagaki & Kim, 1998; Peterkehan, 1998; Skehan, 2009). With all these taken into consideration, foreignness can be avoided with being able to proficiently use abovementioned characters of a language.

In our country, especially in prep classes I have taught so far, those abovementioned language skills are taught sometimes as separately sometimes as integratedly. Besides, writing skill is examined separately by asking students to write paragraphs or essays. Students are tested to see if they are grammatically competent, if they can organize their paper and if they can produce topic sentences, main ideas, supporting sentences and concluding sentences well enough. When they collect enough marks on a designed rubric, it means they have enough competence to pass the course. The reason I chose writing skill depends on all these logics. To survey these features of writing, I carried out a research which included tertiary level students' essay papers which were all in argumentative type.

Learners should know the purpose of their writing. Why do they write? As Bailey (2011) argues, the prominent reasons are as follows: "to report on a piece of research the writer has conducted, to answer a question the writer has been given or chosen, to discuss a subject of common interest and give the writer's view, to synthesise research done by others on a topic" (Bailey, 2011: 3). The researcher additionally states that the writer should smoothly leave her

message for a clear reading. The writer should be able to write clearly, so the receiver, in this case the reader, can infer the intended meaning without being led to other conclusions. So, this view can be summarized in a way that a writer should clearly write to imply a clear understanding. Therefore, foreignness can be avoided with being able to write smoothly, clearly, understandably, apparently, and comprehensibly.

1.2.9. Argumentative Writing and EFL Learners

I have been teaching English at universities for 11 years, and I have been teaching writing course for 7 years now. In preparatory classes, writing syllabus includes descriptive writing, proposal writing, opinion essays, plans, summaries, narrative essays, cause/effect essays and argumentative writing which is the main concern of my study. I decided on argumentative writing because as Coşkun and Tiryaki (2011) state, argumentative text writers need to produce a statement or a claim and support them with evidences. These procedures lead the reader to think in more details and may sometimes change his opinions by the argumentative writer's opposing ideas.

In addition to this, Hatch (1992) gives a definition to "argumentation as the process of supporting or weakening another statement whose validity is questionable or contentious" (p.185) (as cited in Özhan, doctoral dissertation, p.67). Besides, Özhan (2012) explains what argumentation is; and states "argumentation is a discourse mode where the intention is to persuade the audience to accept a proposition. To achieve that, an argumentative writing requires a debatable topic, a strong claim which is further supported by various forms of evidence, acknowledgement of opposing viewpoints and the refutation of these oppositions" (Özhan, 2012: 10). Again, Özhan (2012: 10) adds that in argumentative writing, the writer's purpose should be to persuade the readers to understand and admit her suggestion. This needs "a debatable topic, a strong claim which is further supported by various forms of evidence, acknowledgement of opposing viewpoints and the refutation of these oppositions".

Argumentative writing is a vital type of academic writing and therefore the learners of the target language should be aware of the importance and should pay utmost attention while producing a written work. Additionally, argumentative writing is essential as the visibility of the learners' defence in their opinions can be durable when the argumentation is strong. The main characteristic of argumentative writing is its inclusion of a language which holds phrases and combinations of contrast. In other words, contrastive language is the most essential feature of argumentative writing. How is this related to this present thesis? The usage of a contrastive language in argumentative essays by students has been examined in terms of verb+noun collocations with the verbs do, make, have, give, take, get and go. I tried to find out how strong students can utilize verbs while defending their claims. In the light of this information, I can state

my reason of choosing argumentative writing; it is to see how students can perform their arguments in their essays by using verb+noun collocations.

1.2.10. Verb+noun Combinations and Turkish EFL learners

As mentioned above, word combinations happen to be instinctively existent in a native speaker's mind, and collocations, which are one type of combinations, are already present, too. The question "why did I choose verb+noun collocations?" can be answered with Fontenelle's (1998) argument which tells that verb+noun combinations, infinitives and prepositional phrases constitute a huge part of collocations. According to Martelli (2006) verb+noun combinations are the most negatively affected collocations by the non-native learner's mother tongue; therefore, it can be stated that mostly occurring collocation errors in a target language are from verb+noun collocations. This can be expressed with L1 interference which implies that a learner can sometimes employ some characteristics of her native language while trying to produce a target language structure.

The verb+noun collocations in this present thesis will consist of the base verbs such as *do*, *make*, *have*, *give*, *take*, *get* and *go*. The reason to decide on these verbs is that these are the mostly known and used words in students' papers which I had analysed during my 11 years of teaching English as a foreign language. Sprenger (2003) states in his doctoral dissertation that if a multi-word combination occurs frequently, it means it could be examined in more detail. That is why these base verbs were chosen in order to study the verb+noun collocations.

1.2.11. The Use of Corpus as a Tool for Linguistic Investigation

This present study is based on two corpora formed with non-native learners of English in two different departments related to English, and a native corpus. Therefore, this study is existent under the term corpus linguistics which is for Granger et al (2002: 4) "a methodology which is based on the use of electronic collections of naturally occurring texts for various pedagogical purposes". The purpose of utilizing this kind of study is collecting students' argumentative essays in an electronic platform, and working in detail following the collection of the student products.

Corpus is defined as a language study which relies on computer assisted techniques that aims at examining naturally occurring language (Biber et al, 1998; Conrad, 2000; Granger, S., 2002; Kennedy, 1998; McEnery and Hardie, 2012). In addition to examining texts, "corpus is the most reliable source of evidence for such features as frequency" (McEnery and Wilson, 1996: 9).

Sinclair (1996: 2) puts a definition of learner corpora as follows:

...the electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardized and homogeneous way and documented as to their origin and provenance (as cited in Özbay & Kayaoğlu, 2016: 217).

Moon (1998b) reveals the strong connection between corpus and phraseology by her statement which argues that corpus data is a fruitful area for studying phraseological items because corpus can help the researcher identify collocational items, fixed expressions and idioms. Özbay (2015: 1) in his doctoral dissertation underlines the importance of corpus linguistics with the statement which says “that corpus linguistics presents us with profound changes in the way that we study, teach and learn languages all over the world due to its huge potential to present entirely authentic, genuine, qualitative and quantitative findings related to the nature of language”. Abovementioned opinion about the value of corpus can lead to the inference that corpus linguistics helps teachers and consequently students be more aware of the current situation of language teaching and learning.

Each corpus may include hundreds or thousands or millions of words. This thesis study includes three different corpora each of which has varied numbers of words. The largest corpus is a non-native one; that is KTUCLE, it consists of seven hundred and nine thousand and seven hundred and forty-eight words (709, 748). The second largest corpus is a native speaker corpus named LOCNESS which includes three hundred sixty-one thousand and fifty-four words (361,054). The smallest corpus is TUCLE, which is my own corpus formed with my preparatory students’ argumentative papers. This corpus embraces one hundred and ninety-five thousand, six hundred and eighty-four words.

Hunston (2002), Granger (1998a, 2002, 2003, 2009) and Sinclair (1991) argue that corpus is positively effective in language studies as the lexical units and certain combinations, which require more attention during target language teaching, can be reached and teaching methods can be amended accordingly. Additionally, how corpus can help in language teaching and learning can be answered again with Özbay’s (2015: 2) argument which states “that the frequency of certain words and lexical structures as well as the associated collocational preferences in texts pedagogically deserve immediate focus and research”. This argument can be summarized in this way; the instructor of a language class can employ a corpus-based study and instantly change or revise her view and method of teaching. Besides, she can determine which patterns the students are good or bad at; in this way, she can additionally compare her other classes with the ones she already teaches. Gabrielatos (2005) asserts that corpora studies and the awareness of their help pave the way for the teachers to form new activities and materials for effective teaching. This study scrutinizes argumentative essay papers by utilizing corpus because as Müller (2011: 2) asserts

corpus based studies are of high value as a result of “new methodological approaches as well as the availability of huge electronic corpora”; this modernization in linguistics is also significant for phraseological research.

1.2.12. Corpus Investigation and ELT

Corpus linguistics and English Language Teaching (ELT) has been working hand in hand since Sinclair started a project in the early 1980s (Flowerdew, 2004: 12). John Sinclair, who was a professor at the University of Birmingham in the English Language Department, worked together with Collins Publishing. Their common aim was preparing teaching materials and superior dictionaries which are parts of real English. They called it ‘real’ because the aim of this project was to let learners and users of English to be prepared to face actual language usages. With this purpose in mind, they formed Bank of English (BoE) which included millions of words corpus belonging of written and spoken English belonging to native-speakers (Flowerdew, 2004: 12; Römer, 2010: 19).

Corpus is in direct relation to language studies since “it has led to far-reaching new hypotheses about language such as the co-selection of lexis and syntax” (Stubbs, 1996: 232). Additionally, according to (Leech, 1992: 106) “it has the potential to change perspectives on language as a new research enterprise and a new philosophical approach to the subject”. Corpus studies are of great help for language instructors when teaching lexical and phraseological structures, because as McCarthy and Carter (2004) and Schmitt et al (2004) argue, these disciplines of language receive great attention and researchers can gain reliable evidence with the help of employing corpus based studies as efficient use of corpus based studies can be beneficial with its ability to compare situations or students, and its convenience to gain consistent and precise findings.

McEnery and Wilson (2001) assert that corpus is of great help to determine whether the target language usages of non-native speakers are correct or not by the employment of a corpus study. That is to say, the students may tend to form structures which they believe or feel are correct; however, a corpus based study can prove they are wrong or not by revealing concrete evidence. Learner corpora which are one area of corpus linguistics assist the researcher “to arrive at conclusions regarding the lexical development of the learners” (Özbay, 2015: 5). In addition to this argument, Granger’s (1998) statement which emphasizes that corpus helps the researcher receive objective data and utilize that data for academic and analytical aims seems to be logical and understandable.

1.2.13. The Importance of Corpus in ELT Research

Corpus Linguistics helps us study authentic materials (Allan, 2009), in this present research argumentative essays of the students. An answer to the question of the importance of corpus in language studies can be given by Chomsky who affirms that rules lead us to have the feeling of creating original and correct sentences (McEnery & Wilson, 2001). A supporting claim from Kennedy (1998) states that corpus is trustworthy as it can create “the basis for more accurate and reliable descriptions of how languages are structured and used” (1998: 88). To give it more clearly, corpus studies can assist the researchers and the language instructors figure out the correct and acceptable usages of target languages.

Corpus-based studies direct language instructors and researchers investigate language teaching environments (Campoy-Cubillo et al., 2010: 3). Technological improvements in language studies; in this present corpus-study, let the researcher gain a concrete display of importance of multi-word combinations because the corpus studies reveal concrete results which help the language be understood and utilized in a better way (Altenberg & Granger, 2001). Additionally, it can be noted that research and studies about multi-word expressions remains weak in terms of quantity and quality Nesselhauf (2005: 112); therefore, this current corpus-based study is carried out to fill at least one gap with the help of investigating multi-word combinations.

This present study deals with native and non-native corpora. It works on materials produced by learners; namely, they are the learners’ language (interlanguage) products. In order to study the interlanguage characteristics of learner products, Granger (1998a) first composed corpora, and since then working on learners’ corpora has gained attention (Altenberg & Granger 2001; Altenberg & Granger, 2002). The aim of this kind of study is to analyse errors in non-native works and trying to reveal their correct forms, and this would help EFL teachers and learners pay more attention to their studies. In other words, a corpus-based study can help the researcher to form a detailed investigation of the students’ argumentative essays and in this way, it can reveal the overuse and underuse of certain collocations.

This thesis study elaborates on three different corpora one of which is native and the other two are non-native corpora. The reason of working on three corpora is moving away from typical studies which include two corpora only to compare the target language and the native language features. This present research, additionally, looks for the details in two non-native corpora in terms of their characteristics.

Similar to this present study of mine, which includes three corpora to examine, in his doctoral dissertation, Tono (2002) studied three corpora as well. He worked with Japanese EFL learners to make comparisons on verbs. The corpora consisted of native works of Japanese, ELT textbooks,

and thirdly target language (English) works of native speakers of Japanese. The results of the research explain that the time which is spent on learning the target language has nothing to do with the correct usage of verb patterns. Besides, the verbs the learners used in their essays were the ones which they were taught in their textbooks. This clearly reveals that the types of verbs in textbooks are directly correlated with the production; to be precise, learners mostly used the verbs they faced in their textbooks. However, this does not mean that they do not make errors; they still make errors although the verbs they use are existent in their textbooks with their correct forms. Last but not least, the misuse of verbs resulted from the students' L1 interference.

Borin and Prütz (2004) again employed investigation of three corpora. However, they worked with native speakers of Swedish. The corpora included Swedish texts, target language English texts in different corpuses. Their research investigated syntactic structures of English language in "(1) the Uppsala Student English corpus (USE); (2) the written part of the British National Corpus Sampler (BNCS); (3) the Stockholm Umeå Corpus of written Swedish (SUC)" (Borin and Prütz (2004: 67). Their purpose was to reveal the L1 transfer in the target language usage in terms of overuse and underuse by comparing native and non-native corpora. The study found that students were in the tendency of making L1 transfer in their interlanguage processes. Additionally, this result of the study overlaps with the ones of Aarts and Granger's (1998) and Borin and Prütz's (2004) in the way that the participants in those studies implemented L1 transfer as well which confirms that interlanguage process is sturdy in target language production.

Comparing two corpora studies are also existent. Leńko-Szymańska (2004) compared only two corpora in which he elaborated on anaphora markers by analysing native and non-native (Polish) essays. He worked on two corpora named PELCRA corpus of learner English and the BNC Sampler. The results of the research revealed that demonstratives (this, that, these and those) were overused in the argumentative essays of native speakers of Polish in their English essays. Additionally, the research says that time spent on learning is not directly correlated with the correct use of the target language patterns. That is to say, the exposure to the target language is not significantly effective in creating correct usages.

Nesselhauf (2004) worked similarly to Leńko-Szymańska in the way that she employed native and non-native comparison as well. She worked with native speakers of German whose essays were analysed in terms of support verb constructions. The research took ICLE (the International Corpus of Learner English) as the basis to compare German essays. The study found out that non-native users of English use verbs in a problematic way not because the chosen verbs create problems but because they are already used incorrectly.

Thewissen (2013: 95) utilized an EFL learner corpus with the purpose of examining target language accuracy in students' essays. The results showed that EFL learners faced errors such as

progress-only errors, stabilization-only errors, progress and stabilization errors, and error types with marked regression.

In the light of the information and studies presented above, this research aims at comparing native and two non-native corpora one of which is compiled by my own students' argumentative essays they prepared during four months of study. The purpose of comparing three corpora is to reveal the usage differences by natives and non-natives in terms of over and underuse of verb+noun collocations formed with base verbs *do*, *make*, *have*, *give*, *take*, *get* and *go*.



CHAPTER TWO

2. METHODOLOGY

2.1. Introduction

This research aims at doing a corpus-based interlanguage analysis of verb+noun collocations in tertiary level EFL students' argumentative essays. The reason why this pattern of collocations is studied has been reiterated several times so far as the comparison among three corpora can yield to amendments in language or vocabulary teaching methods. Additionally, the results of the study is likely to increase an awareness on the part of the EFL learners towards their weakness, in this case EFL learners, be aware of their weaknesses and drawbacks; and therefore, they can work for a specific purpose which leads to awareness.

This section of the study includes a research design which is into four parts such as study design, sample selection, data collection instruments and procedures, learner corpus data extraction, data analysis and limitations of the study. In addition, and most importantly, the compilation of the researcher's own corpus (TUCLE) is described. The corpus size, participants, the corpus compilation method, the argumentative essays topics and receiving the consents of the students are mentioned in the following parts of this chapter.

2.2. Methodology

This study is based on a corpus analysis and this analysis will be carried out through a computer assistance and the naturally occurring data (Biber et al, 1998; Conrad, 2000; Granger, 2002; Kennedy, 1998; McEnery and Hardie, 2012) will be examined with a special focus to the use of collocations by tertiary level EFL learners. In the scope of the study, there will be some corpus based numeric analysis of the data and this makes the study, all the same, a quantitative one (Sinclair, 1991; Biber et al, 1998).

The study includes three corpora; one of them is a native corpus named LOCNESS (Louvain Corpus of Native English Essays) which includes 361,054 words. KTUCLE is one of the non-native corpora and stands for Karadeniz Technical University Corpus of Learner English. It is the largest of the three corpora compared; it consists of 709,748 words. The second non-native corpus is TUCLE which stands for Trabzon University Corpus of Learner English. This corpus seems to

be the first of its kind in Trabzon University since the researcher compiled it by herself with several argumentative essays papers written by her preparatory students at the department where she teaches reading and writing courses. TUCLE includes 195,684 words. TUCLE and KTUCLE are learner corpora which are formed by the tertiary level preparatory students at two different state universities in Trabzon. The two EFL learner groups are of the B2 level according to CEFR (Common European Framework Reference for Languages). Both non-native corpora include expository argumentative essay papers to be compared; however, the difference is the students' departments. KTUCLE is the corpus of English Language and Literature Department students whereas TUCLE is the corpus of English Language Teaching students.

Table 2: The Learner Corpora Used in This Study

| Corpora | Number of texts | Number of participants | Number of words |
|---------|-----------------|------------------------|-----------------|
| TUCLE | 344 | 21 | 195,684 |
| KTUCLE | 1600 | 480 | 709,748 |
| LOCNESS | 322 | unknown | 361,054 |
| Total | 1266 | 501+? | 1.266,486 |

Representativeness holds a significant place in corpus studies (Barnbrook, 1996). Biber (1993) argues that the representativeness of a sample “depends first of all on the extent to which it is selected from the range of text types in the target population” (1993: 243). Additionally, representativeness is a matter relying on the corpus size; that is to say, the corpus size is of utmost importance in generalizing the groups studied so this can lead to the inference that the larger the corpus size is the more generalizable the group is. Under the light of these, it can be asserted that representativeness is an issue which is attained by creating a balance in the corpus; in other words, sampling categories of essays or texts should primarily be described (Hunston, 2002; Sinclair, 2001; Tognini-Bonelli, 2001).

Atkins et al. (1992) provide a definition for texts; in this study they are the argumentative papers, “Texts are often assumed to be a series of coherent sentences and paragraphs” (1992: 2). Argumentative papers of the tertiary level EFL students were analyzed in this research as the reflectiveness of the writers are clearly seen while they are putting their opinions forward in order to let the readers be persuaded by their thoughts. As argued by Hatch (1992), argumentative works of writings are touchy since the producers weaken or support an existing idea with their own opinions and this makes the argumentation robust in claiming the validity of the rightness of the writers.

This research study utilized argumentative essays since the type of writing requires a “debatable topic” (Özhan, 2012:10) in which students were free to argue what their opinions were. During the compilation of TUCLE, the participants stated that they were really eager to be assigned

with their new writing projects as the topics were exciting and they could not wait for getting their arguments down on their papers. After putting their ideas down into the papers and handing in the papers, they were in discussion with their classmates in which they asked each other what their opinions were. With all these being put forward, the argumentative essay papers can be assented as written texts to be examined in this corpus study.

Table 3: The Study Design Steps Followed in the Methodology Section

| | |
|---------|---|
| Step 1: | Consent Form was given to all the preparatory students in the English Language Teaching Department. The total number of these students was 21. The form is given in Appendix 1. |
| Step 2: | Creation of TUCLE which is the compilation of essays by students of B2 levels. |
| Step 3: | Data Collection: Criteria for the Selection of Verb+noun Samples from the Learner Corpora: TUCLE and KTUCLE |
| Step 4: | Data Collection: Essays are sent to the researcher's e-mail address on Sundays. |

2.3. Setting and Sample Selection: Participants

This current research study, as mentioned above, includes two non-native corpora. Both are compiled by EFL learners' argumentative essays. TUCLE was compiled with the argumentative essay papers of 21 participants whose ages range from 18 to 22. One of the 21 students is Turkmen, one is originally Turkish but is an Austrian citizen, and the rest are all Turkish citizens. 8 of the participants (38,1%) in TUCLE are males; that is almost one third of the class size. This is a typical example of a language department in Turkey where English language departments are dominantly loaded with female students. Similarly, KTUCLE was compiled by the 600 papers of the tertiary level students aged between 18 and 22, and their number was 480. The native corpus LOCNESS, which is recognized as the reference corpus, includes 322 texts produced by native speakers of English who were between 17 and 23 years of age.

Table 4: Demographic Information of TUCLE Participants (Subjects)

| | | Number | % |
|-----------------------------------|--|--------|-------|
| Gender | Female | 13 | 61,9 |
| | Male | 8 | 38,1 |
| Age | 18 | 12 | 57,14 |
| | 19 | 7 | 33,33 |
| | 21 | 1 | 4,76 |
| | 22 | 1 | 4,76 |
| Total stay time in the department | Less than a year | 21 | 100 |
| Previous high school graduation | Public school | 19 | 90,47 |
| | School abroad | 2 | 9,53 |
| Other foreign languages known | German (A1 level) | 16 | 76,19 |
| | German (native) | 1 | 4,76 |
| | French (A1 level) | 1 | 4,76 |
| | French (native) Spanish and Dutch (B1 level) | 1 | 4,76 |
| | Turki Languages (native) | 1 | 4,76 |
| | Russian (B2 level) | 1 | 4,76 |
| Abroad experience | No foreign language | 1 | 4,76 |
| | Yes | 3 | 14,28 |
| | No | 18 | 85,72 |

The table below reveals the learner variables of the participants of TUCLE; their age, gender, mother tongue, region, level and learning contexts. There are two students whose Turkish accents and comprehension are a bit different from the rest of the classroom. There is a student who owns the citizenship of Austria and speaks German much better than Turkish; he even looks up the unknown words of English in an English-German dictionary. The second one is a female student who grew up in France and who also has difficulty in creating productive skills although to a lesser extent than the former two male students.

2.4. Compilation of TUCLE

The participants in TUCLE were given a placement test at the beginning of the fall term in the 2018-2019 academic year. This test was prepared and administered by the ELT department academics; and it included sections of evaluating skills as grammar and reading, writing, listening and speaking. The purpose of this placement test was to let the successful students start directly through the first grade of department whereas the unsuccessful ones had to complete two semesters of preparatory class with a degree of at least 80 out of 100. The students who passed the placement test got 70 or higher, so they could start their departmental education without having to study preparatory class for a full academic year.

The grammar and reading section of the placement test included 50 questions; in addition to that, writing, listening and speaking tests were applied separately. The listening test included 25 questions whereas the writing test required the attendants to write a paragraph about one of the topics given. The speaking test was open-ended; therefore, there was not any fixed number of questions. All the tests were rated out of 100 and the 25 percent of each section was taken out; so when each section is added up together they created the overall percentage. The listening and the written tests are attached in the appendix 2 and 3.

The successful ones passed the preparatory class and started their education right at the first grade. However, as mentioned above, the ones who failed studied preparatory class and they took courses such as main course, grammar, speaking, listening, and most importantly for this thesis study reading and writing. The researcher taught reading and writing in the preparatory class and those courses took 6 hours a week; 4 hours for the reading course and 2 hours for writing. The researcher assigned the students essays which had different topics ranging from social sciences to politics. It can be inferred that the students were eager to write as much as possible; because they asked the researcher what their new topic was before the teacher even mentioned about the homework. The participants also told that they loved writing because they were argumentative and expository in style which did not push them to be scholarly or strict.

It was compulsory for the participants to submit their argumentative papers every Sunday until 23:59 which is the last minute of the weekend. This deadline was fixed at the beginning of the fall term; thus, it was unnecessary to remind the deadline after every assignment. The late givers were ignored in order to create order for the class although a few attempted to submit their papers. The papers were argumentative in type and each essay topic was determined by the researcher so as not to create a disordered compilation. In addition to these, there was a limit of building at least 450 words in order to let students write as much as they can. Some students who were enthusiastic about writing went further and they created 600 words for each essay. However, there were few who wrote no more than 460 words. Yet, at first, as they said, they had difficulty in creating 450 words for one essay; however, as time passed they could write more than the required number of words and this made the corpus larger than expected.

Oakes (1992: 249) argues that the information which is already present in the producer's mind can physically exist by the use of computers and therefore the researcher collected the essay papers as written on word documents. In other words, students prepared their argumentative essay papers on computers and then sent them to the researcher by electronic mail which required extra work to draw the essays one by one from the e-mail platform.

Following the collection of all the documents, the researcher converted them into text file documents with AntFile Converter and uploaded into the corpus software concordance program

named Sketch Engine. This software can easily be reached by public since it is free online. The verbs *do, make, have, give, take, get* and *go* and their collocational noun partners were extracted, and the most commonly used ones were listed and grouped together. The verb+noun collocations that were most commonly utilized by the students were checked by referring to Oxford Dictionary of Collocations (2002) and BNC (British National Corpus) to check their acceptability and correctness.

The essays were collected from the students in two consecutive terms in the academic year 2018-2019. The preparatory students started writing in English first in paragraph form; they learned how to write a topic sentence, supporting sentences and a concluding sentence. Following the 10 weeks of paragraph writing, they were required to start writing in essay form. So, the researcher started asking them to write essays which included at least three paragraphs; the first one is the introductory paragraph which included a topic sentence; the second (third and fourth if any) is the body paragraph which supported the introductory paragraph and lastly the concluding paragraph which summarized the opinions in the former paragraphs. In other words, the students learned how to write essays and then they started essay writing. The scope of the topics was large enough to include different issues such as argumentations in social sciences and humanities. The students reported their enjoyment and satisfaction with some of the essay topics, though there were a limited number of students who, seemingly, expressed their concern related to the essay topics in the sense that the topics were not particularly relevant to their interests.

Table 5: The Argumentative Essay Topics of TUCLE

| |
|--|
| 1. Is it OK to date a younger male? |
| 2. The influence of internet: more harm than good or more good than harm. |
| 3. Should scientists test products on animals? |
| 4. Many teachers assign homework to students every day. Do you think that daily homework is necessary for students? Use specific reasons and details to support your answer. |
| 5. Cell phones: dangerous or useful for humans? |
| 6. Expertise or wisdom. |
| 7. Human cloning? Are you for or against? Discuss. |
| 8. A university in every town: are you for or against? Discuss. |
| 9. Which secondary languages are worth studying today? Why? |
| 10. Production and sales of tobacco must be made illegal. |
| 11. Death sentence should be activated in every country. |
| 12. Governments should forbid same-sex marriages. |
| 13. How can you impress someone? |
| 14. Should young men have the right to choose when it comes to military? |
| 15. People mustn't sleep more than 8 hours |
| 16. All students should wear uniforms to schools. |
| 17. The world cup should be organized every two years. |
| 18. Television can destroy communication among friends and family. |
| 19. China may become the biggest economic power. |
| 20. Men can wear skirts in the public. |
| 21. Is gun control effective in crime rates? |
| 22. Should court proceedings be documented for TV? |
| 23. Your opinions about abortion. |
| 24. Parents have the right to control their children above 18. |
| 25. What do you think are the pros and cons of social media? |

It should be kept in mind that the producers of those argumentative papers were not native speakers of English; therefore, they had errors in structures of sentences since they still are not native-like. They also had semantic and lexical errors which also reinforced the feature of interlanguage that affirms the imperfectness of the learner's language (Granger, 1998; Corder, 1982; Selinker, 1992).

The findings of this research study embody the view about the interlanguage feature with the lack of proficiency in grammatical and lexical usages of the students. The participants were unable to create fully-correct sentences in terms of grammar, syntax or semantics. As a result, this study is a drastic indicator of interlanguage characteristics. Therefore, TUCLE and KTUCLE do not hold scholar writing or native-like works of essays.

Table 6: The Corpus Design Criteria of TUCLE

| | |
|-----------------------------|---|
| Dates between | 19 th November 2018- 18 th March 2019 |
| Medium | Written |
| Genre | Expository argumentative |
| Topic names | Social sciences and politics. |
| Number of the topics | 25 |
| Technicality | EFL argumentation |
| Text type | Unpublished argumentative essays |
| Task Setting | Homework |
| Name of the authors | 21 preparatory students |
| Word count | 195,684 words |
| Level | B2 |

The design criteria table (Table 7) below for language variables of TUCLE participants presents the information that the students produced written works for the compilation of TUCLE. They were pre-informed that their papers were going to be utilized for the researcher's master thesis study. They were assured that their names would never be made public, and they also were not informed about what specific usages were going to be analyzed. In other words, they were felt free to write in the way they originally write. They were asked to write expository argumentations since they have not learned how to write academically, yet. The topics they were required to write about were in social sciences and politics disciplines. The participants were given a pre-determined deadline to hand their papers in.

Under normal circumstances, the researcher provides feedback for the first drafts of the papers; the students correct their mistakes accordingly and present their final drafts in the portfolios they prepare. However, following the submission of the essays, the researcher gave feedback for the students with their papers; yet, she did not utilize the final drafts, she utilized the first drafts of the papers in order to maintain the originality of the papers.

Table 7: Design Criteria for Language Variable Distribution of TUCLE

| Level B2 | |
|-----------------|------------------------------------|
| Medium | Written |
| Genre | Expository argumentation |
| Topic | Arts and Humanity, Social Sciences |
| Task Setting | Timed Essays |
| | Reference |

2.5. Reference Corpus: LOCNESS

Granger and Tyson (1996) argue that there should be a control native speaker corpus which is “exactly the same type of writing” (1996: 19). For this, a corpus of learner assembled by the Centre for English Corpus Linguistics at Université Catholique de Louvain in Belgium has a significant and a pioneering role in linguistics studies because this centre formed resources named Corpus of Learner English (ICLE) which includes large number of computerized language materials (Narita et al, 2004: 1171). ICLE holds two sub-corpora one of which possesses the pioneering role in corpus linguistics studies; that is LOCNESS (Louvain Corpus of Native English Essays). The other is a learner corpus of Swedish and French learners of English. The need for using a reference corpus is asserted by Granger and Tyson (1996) as the learner products should be controlled with the products of natives in order to compare those with regard to the genres they hold. Besides, it is of high importance to compare learner writing to the native writing since a definite conclusion cannot be withdrawn without it. Under the light of the abovementioned reasons, this thesis study utilizes LOCNESS as the reference corpus; the argumentative essays in TUCLE and KTUCLE are compared to the ones in the native corpus called LOCNESS. It includes three constituents: it has essays of A level essays by British students, essays of British university students and essays of American university students (retrived from the official website of LOCNESS, 2019). As for the size, LOCNESS holds 361,054 words, 322 texts and 90 topics. The ages of the writers of the essays range from 17 and 23.

The second learner corpus is KTUCLE which stands for Karadeniz Technical University Corpus of Learner English. It is compiled with the argumentative essays of Department of English Language and Literature students who also hold B2 level. The corpus includes 709,748 words, 600 texts and 480 students whose ages are between 18 and 22. This age range is the same with the ones of TUCLE participants. KTUCLE is the largest corpus of the three corpora compared.

2.6. Analysis of the Comparisons Among the Three Corpora

The comparison of the corpora LOCNESS, KTUCLE and TUCLE, and observed and normalized frequencies of the verb+noun combinations included extracting and comparing with the purpose of displaying noteworthy similarities and differences in terms of overuse and underuse. In addition, log likelihood measure was applied in order to figure out the amount and rate of the “overuse” and “underuse” of verb+noun combinations.

Log likelihood values of the combination samples were reflected in terms of overuse and underuse having been based on the values in the tables. If the log likelihood test result shows up more than 6.63, that means the difference among the corpora occurs at 99% level; and that yields to the meaning that it is $p < 0.01$. If the log likelihood is 3.84 or more, that means the difference among the three corpora occurs at the 95% level; and that is expressed as $p < 0.05$.

2.7. Criteria for the Selection of Verb+noun Collocations

TUCLE, which was designed and compiled by the researcher, includes expository argumentative essays whose writers are tertiary level students of English language. The verb+noun collocations with the base verbs *do, make, have, give, take, get* and *go* are extracted and listed one by one. The reason of utilizing the verbs above is that they are the most frequently used verbs with nouns (Sinclair, 1990; Akimoto, 1989). The verb+noun collocations were checked in collocation dictionaries to see if they are acceptable, one of them is the Collins Cobuild English Dictionary (1995) and the other is Oxford Dictionary of Collocations (2002). Following the dictionary checks, the collocations used by the writers of the essays were checked in the BNC (British National Corpus). In case the created collocations fail the acceptability test in dictionaries and in BNC, a native speaker of English would be consulted to do a proofread.

2.8. Data Collection

2.8.1. Demographic Information Questionnaire and Consent Form

The demographic information table is based on the answers directly given by the students of TUCLE subjects. They were asked in person in order to create a crystal-clear image of the participants. They were asked about their abroad experiences, the schools they studied in, the foreign languages they can speak and their mother tongue. Besides, students were pre-informed about the research study and their consents were taken by letting them sign a consent form. The consent includes the confidentiality which assures that their names would not be made public; and also it consists of the reminder of voluntariness which means the researcher did not push them to participate in the study. Attached is the abovementioned consent form in the appendix section.

2.9. Corpus Tool

This present research examined native and non-native corpora in which English language is the means of writing. One of the purposes of evaluating language products is that “language users never choose words randomly, and language is essentially non-random” (Kilgarriff, 2005: 263).

Native and non-native users of English language differed in employing collocations in their written products and the text analysis tool which was implemented in this research study was Sketch Engine whose results were examined for creating a quantitative analysis of the verb+noun collocations formed with the base verbs *do, make, have, give, take, get* and *go*.

Sketch Engine was developed by Kilgarriff and his colleagues in 2004 as “a corpus tool which takes as input a corpus of any language and a corresponding grammar patterns and which generates word sketches for the words of that language” (Kilgarriff et al, 2004: 105). Sketch Engine is of value in quantitative analysis of collocations since “by clicking on a collocate of interest in the word sketch, the user is taken to a concordance of the corpus evidence giving rise to that collocate in that grammatical relation” (Kilgarriff et al, 2004: 108). Sketch Engine is available to be used in many different language domains, and during the analysis “the items with the highest frequency in the domain corpus in comparison to the reference corpus will be the top term candidates (Kilgarriff et al, 2014: 53). The Sketch Engine works with a keyword function while comparing any corpus existing on the system. The easiest method to compare the frequency of a word in a text and its frequency in another text lies in the employment of Sketch Engine. Occurrences in the compared corpora is counted, then each number is divided by the word number in the corpus, and it is multiplied by 1,000 or 1,000,000 in order to create the frequency in terms of thousand or a million, and finally the first number is divided by the second number to reveal the ratio.

Following the collection of all the documents, the researcher converted them into text file documents with the help of a program named AntFile Converter. Then, the compiled corpus by the researcher and the other two corpora were uploaded into the Sketch Engine which is a corpus software concordance program. The noun collocations of the target verbs were found with the help of the word sketch feature of the program. Sketch Engine already provides both raw and normalized frequencies of any searched item. It makes it easier for the user to calculate and compare the uploaded data. In order to search the target word by the word sketch feature, some limitations were done such as keeping the minimum frequency as 1, number of findings as 999, and ordered by frequency. Verb+noun collocations were gathered from the section of ‘objects of target word’.

Having provided the concordances of the target verb+noun collocations, the three corpora were compared in terms of overuse and underuse with the help of Log-likelihood (LL) measure. LL measures were taken by using an online free calculator by Lancaster University which is one of the pioneers in corpus studies. The total size of the corpora and raw references of the verb+noun collocations of the target verbs are given into the calculator, and it provides the overuse and underuse measures.

CHAPTER THREE

3. FINDINGS AND DISCUSSION

3.1. Overview of the Study

This corpus-based research study is quantitative in principle and computer based analytical techniques were employed to interpret the verb+noun collocations in the argumentative essay papers of tertiary level EFL students and a native corpus named LOCNESS. The target collocations were extracted with the help of the corpus tool sketch engine which helps the researcher obtain raw and normalized frequencies of the items searched.

3.2. Analysis of the Comparisons across TUCLE, KTUCLE and LOCNESS

During the analysis of the three corpora, raw and normalized frequencies of the verb+noun collocations with base verbs *do*, *make*, *have*, *give*, *take*, *get* and *go* were extracted and compared with the purpose of uncovering similarities and differences which may be important in the interpretation of the overall data. Additionally, log likelihood (LL) measure was employed to scrutinize the overuse and underuse amounts of verb+noun collocations. The extracted verb+noun collocation samples were displayed in terms of their overuse and underuse patterns. When the log likelihood measure is more than 6.63, it indicates that the difference between the corpora happens to be at 99% level. It also points to the meaning that it is $p < 0.01$. When the log likelihood measure is 3.84 or more, it is the evidence that the difference among the corpora happens to be at 95% level which is revealed as $p < 0.05$.

Table 8: Distribution of “do” in TUCLE, KTUCLE and LOCNESS

| DO | | | | | | | | |
|------------------|-----|---------------|-------------------|-----|---------------|--------------------|-----|--------------|
| TUCLE 195,684 | | | KTUCLE 709,748 | | | LOCNESS 361,054 | | |
| OBJECTS | RAW | NORM | OBJECTS | RAW | NORM | OBJECTS | RAW | NORM |
| homework | 94 | 480,37 | homework | 448 | 631,21 | thing | 25 | 69,24 |
| thing | 47 | 240,18 | something | 70 | 98,63 | job | 21 | 58,16 |
| something | 44 | 224,85 | thing | 68 | 95,81 | something | 18 | 49,85 |
| military service | 35 | 178,86 | work | 57 | 80,31 | nothing | 13 | 36,01 |
| anything | 31 | 158,42 | everything | 51 | 71,86 | anything | 11 | 30,47 |
| everything | 26 | 132,87 | research | 43 | 60,58 | study | 7 | 19,39 |
| job | 18 | 91,99 | experiment | 41 | 57,77 | everything | 7 | 19,39 |
| work | 17 | 86,87 | anything | 41 | 57,77 | work | 5 | 13,85 |
| research | 16 | 81,76 | test | 31 | 43,68 | favour | 5 | 13,85 |
| duty | 16 | 81,76 | shopping | 24 | 33,81 | research | 4 | 11,08 |
| experiment | 10 | 51,10 | job | 23 | 32,41 | rest | 3 | 8,31 |
| abortion | 8 | 40,88 | protest | 21 | 29,59 | literature | 3 | 8,31 |
| business | 6 | 30,66 | exercise | 17 | 23,95 | | | |
| crime | 6 | 30,66 | activity | 17 | 23,95 | | | |
| nothing | 4 | 20,44 | sport | 15 | 21,13 | | | |
| activity | 4 | 20,44 | nothing | 14 | 19,73 | | | |
| rest | 1 | 5,11 | testing | 11 | 15,50 | | | |
| | | | practice | 7 | 9,86 | | | |
| | | | action | 7 | 9,86 | | | |
| | | | study | 7 | 9,86 | | | |
| | | | favour | 2 | 2,82 | | | |
| | | | literature | 1 | 1,41 | | | |

The first base verb of verb+noun collocations to be investigated is “do” with which students produced “do homework” collocation more than any other combinations in the three different corpora. One outstanding verb+noun collocation is “do homework” which is considerably frequently used by the two non-native groups. TUCLE participants placed “do homework” combination into their essays with 94 raw references and the normalized frequency of 480,37. Though, KTUCLE participants use the mentioned combination more than the other non-native speakers with the normalized frequency of 631,748 which is higher than TUCLE. However, what is most frequently used in the native corpus is “do thing” with the normalized frequency of 69,24; besides, the aforementioned collocation is the second most frequently used combination in TUCLE, the researcher’s corpus.

The reason for employing “do homework” more than any other collocation seem to be stemming from the incompetency of collocational usage which is according to Nation (2001) “one important aspect of vocabulary knowledge” (Nation, 2001: 529). In other words, the more collocations the non-natives use in their works the more competent they seem in the target language they are trying to be proficient. Additionally, non-natives may not employ fully correct

grammatical structures which may be due to their lack of target language exposure, and years of experiences that are also important to be a fluent and a competent user of a language (Carter, 1987). For this reason, using one single collocation pattern with a large number seems to be the cause of not being competent enough in English language; in other words, placing a specific collocation more than any other combination may be associated with the incompetency.

One striking difference among the collocational usages in the three corpora is “do nothing” which is the fourth most used do+noun collocation in the native corpus with the normalized frequency of 36,01 whereas it is used among the last rows of the two non-native corpora. However, the normalized frequency of “do nothing” in TUCLE is 20,44 while it is 19,73 in KTUCLE; and this shows that natives use numerous verb+noun collocations which can be inferred from the normalized frequencies of the collocations whereas non-natives stick to some specific combinations and the normalized frequencies seem similar; therefore, it is possible to infer that this is a result of the incompetency of non-natives in their collocational usages.

The least frequently used do+noun collocation is “do rest” in TUCLE with only 1 usage; however, it is a misused collocation since BNC and Oxford Collocation Dictionary search reveal that there are not any references of “do rest”. In addition, it seems possible to infer from the table that the do+noun collocation which does not have any examples in KTUCLE is “do rest”; that is to say, “do rest” has the raw reference of 1 in TUCLE whereas the reference is 3 in LOCNESS, besides, the normalized frequency of the aforesaid collocation is 5,11 in TUCLE and 8,31 in LOCNESS. However, KTUCLE does not hold any evidences of “do rest” in the argumentative papers.

Additionally, TUCLE participants were required to write an essay about ‘abortion’ and their papers included the do+abortion collocation 8 times with the normalized frequency of 40,88; yet, it is an example of misuse since it has a completely different meaning from the one implied by the student(s). This shows that the student(s) may not have searched for the correct collocation and formed an incorrect combination.

Granger (1998b), Wray (2002) and Nizonkiza (2011) argue that effective and correct collocational usage is the sign of native-like proficiency and Granger (1996) also asserts that the interlanguage of non-native users of a target language is neither L1 nor L2 and therefore examples of native language transfer are possible to be encountered in their works. One reference of “do literature” with the normalized frequency of 1,41 in KTUCLE is an example of L1 transfer which reduces the quality of the target language work.

As an overall investigation of do+noun collocations, it is possible to infer that “do homework”, “do something” and “do thing” are the three most used collocations in the two non-

native corpora. “Do thing” and “do something” are also in the most frequently used collocations list in LOCNESS; however, one combination is different from the non-native corpora; natives employed “do job” while non-natives employed “do homework”. TUCLE holds 94 raw references of “do homework” while KTUCLE involves 448 raw references, so this makes the aforementioned collocation have the normalized frequency of 480,37 in TUCLE and 631,21 in KTUCLE. Additionally, the normalized frequency “do thing” in TUCLE is higher since TUCLE participants used the collocation for 47 times whereas KTUCLE students applied it for 68 times; therefore, TUCLE includes 240,18 as the normalized frequency of “do thing” while KTUCLE has 95,81 of it. Furthermore, the last of the three most commonly used do+noun collocations is “do something” which again has a higher normalized frequency with 224,85 in TUCLE whereas the frequency is 95,81 in KTUCLE.

Table 9: Log-likelihood Ratio and Significance Level with “do” in TUCLE and LOCNESS

| DO – TUCLE vs LOCNESS - LL SCORE | | | | | | |
|---|-------------------------|---------------|---------------------------|--------------|-----------|------------|
| | TUCLE 195,684 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| something | 44 | 224,85 | 18 | 49,85 | 32,90 | + |
| thing | 47 | 240,18 | 25 | 69,24 | 26,96 | + |
| everything | 26 | 132,87 | 7 | 19,39 | 26,33 | + |
| anything | 31 | 158,42 | 11 | 30,47 | 26,05 | + |
| research | 16 | 81,76 | 4 | 11,08 | 16,91 | + |
| work | 17 | 86,87 | 5 | 13,85 | 16,30 | + |
| job | 18 | 91,99 | 21 | 58,16 | 2,00 | |
| nothing | 4 | 20,44 | 13 | 36,01 | 1,07 | |
| rest | 1 | 5,11 | 3 | 8,31 | 0,19 | |

The comparison of the native corpus and one of the non-native corpora witnesses a significant finding which reveals that “do something” is overused by the non-native students with the log likelihood measure of +32,90 which is higher than the average value. Although TUCLE is smaller in size, its participants employed “do something” many more times (44 raw references in TUCLE and 18 raw references in LOCNESS) than the native corpus which is bigger in size. The use of “do something” by TUCLE participants has a normalized frequency of 224,85 with raw references of 44 whereas the aforementioned collocation has a normalized frequency of 49,85 in the native corpus. In addition to “do something”, “do thing” is another verb+noun collocation which is significantly overused by the non-native tertiary level students with the raw reference of 47 whereas it has 25 references in LOCNESS. “Do everything” and “do anything” are the two other significantly overused combinations when compared to the native corpus. The LL scores of “do something”, “do thing”, “do everything” and “do anything” display that they are the most frequently overused verb+noun combinations of TUCLE participants; additionally, they have very

close LL measures with +26.96, +26.33, and +26.05 respectively. “Do research” and “do work” are also examples of overuse in TUCLE with the LL measure of +16,91 and +16,30 respectively when they are compared to the native users of English language.

Non-natives do employ overuse of “do+noun” collocations; yet, they can use some of the do+noun collocations in normal frequencies when compared to native speakers of the target language. “Do job”, “do nothing”, and “do rest” are the do+noun combinations which are neither overused nor underused collocations in TUCLE as they have the LL measures of 2.00, 1.07 and 0.19 respectively when compared to native corpus participants. These measures show that the combinations are not overused since they remain within the absolute value of 3.84; because as explained in the beginning of this chapter and in the methodology section, when the log likelihood measure occurs to be 3.84 or more, it means that the investigated collocations are the evidences of overuse. However, “do rest”, which has the raw reference of 1 and the normalized frequency of 5,11 has the feature of misuse which is applied by a TUCLE student.

Table 10: Log-likelihood Ratio and Significance Level with “do” in KTUCLE and LOCNESS

| DO – KTUCLE vs LOCNESS - LL SCORE | | | | | | |
|-----------------------------------|-------------------|--------------|--------------------|--------------|-------|-----|
| | KTUCLE 709,748 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| work | 57 | 80,31 | 5 | 13,85 | 22,99 | + |
| research | 43 | 60,58 | 4 | 11,08 | 16,70 | + |
| everything | 51 | 71,86 | 7 | 19,39 | 14,45 | + |
| something | 70 | 98,63 | 18 | 49,85 | 7,54 | + |
| favour | 2 | 2,82 | 5 | 13,85 | 4,14 | - |
| anything | 41 | 57,77 | 11 | 30,47 | 3,98 | + |
| job | 23 | 32,41 | 21 | 58,16 | 3,67 | |
| literature | 1 | 1,41 | 3 | 8,31 | 2,85 | |
| nothing | 14 | 19,73 | 13 | 36,01 | 2,39 | |
| thing | 68 | 95,81 | 25 | 69,24 | 2,02 | |
| study | 7 | 9,86 | 7 | 19,39 | 1,57 | |

TUCLE is one of the non-native corpora and it is also compiled by the researcher herself. The other non-native corpus is KTUCLE which is compiled by the participants of another state university in Trabzon. KTUCLE participants are different from TUCLE participants only in their department discipline; the former is the literature department whereas the latter stands for the language teaching department. KTUCLE holds the feature of being the largest of the three corpora compared.

When the collocation usages are calculated by normalizing the number into a million, it seems possible to argue that KTUCLE participants employ “do work” combination more than any

other do+noun collocation in their argumentative papers with 57 raw references and the normalized frequency of 80,31 whereas the native corpus holds the raw reference of 5 and the normalized frequency of 13,85 which lead to the inference that the LL measure is +22,99. To make it more explicitly, “do work” is the most frequently used combination by the KTUCLE students; besides, it is also revealed that the mentioned collocation is overused when compared to the native corpus with the log likelihood measure of +22,99. Additionally, “do research”, “do everything”, “do something” and “do anything” are the overused do+noun collocations when compared to the native corpus with LL measures of 16.70, 14.45, 7.54 and 3.98 respectively.

KTUCLE participants employ some do+noun collocations in average frequencies; “do job”, “do literature”, “do nothing”, “do thing” and “do study” are the collocations which are neither over nor underused combinations by the KTUCLE participants; they all remain within the LL limits of normal measure. However, there is an example of L1 transfer which leads to misuse in writing; it is “do literature” which sounds too Turkish.

Thus far, only one example of underuse has been observed in the two non-native corpora. “Do favour” is employed only two times in the largest corpus of 709,748 words whereas it is used 5 times in the smaller native corpus. “Do favour” has a LL measure of -4,14 which remains under the normal accepted measures and therefore it is an example of underuse.

Table 11: Distribution of “make” in TUCLE, KTUCLE and LOCNESS

| MAKE | | | | | | | | |
|------------------|-----|---------------|-------------------|-----|--------------|--------------------|-----|---------------|
| TUCLE 195,684 | | | KTUCLE 709,748 | | | LOCNESS 361,054 | | |
| OBJECTS | RAW | NORM | OBJECTS | RAW | NORM | OBJECTS | RAW | NORM |
| impression | 33 | 168,64 | protest | 54 | 76,08 | decision | 38 | 105,20 |
| decision | 26 | 132,87 | test | 34 | 47,90 | money | 24 | 66,47 |
| friend | 17 | 86,87 | decision | 33 | 46,50 | choice | 18 | 49,85 |
| contact | 8 | 40,88 | thing | 21 | 29,59 | mistake | 16 | 44,31 |
| mistake | 8 | 40,88 | experiment | 20 | 28,18 | profit | 12 | 33,24 |
| sense | 7 | 35,77 | mistake | 18 | 25,36 | argument | 12 | 33,24 |
| thing | 6 | 30,66 | choice | 17 | 23,95 | discovery | 10 | 27,70 |
| choice | 6 | 30,66 | difference | 15 | 21,13 | statement | 7 | 19,39 |
| clone | 5 | 25,55 | effort | 14 | 19,73 | claim | 7 | 19,39 |
| experiment | 5 | 25,55 | use | 12 | 16,91 | point | 7 | 19,39 |
| call | 4 | 20,44 | research | 12 | 16,91 | law | 7 | 19,39 |
| progress | 4 | 20,44 | friend | 12 | 16,91 | difference | 6 | 16,62 |
| test | 4 | 20,44 | homework | 12 | 16,91 | change | 6 | 16,62 |
| law | 4 | 20,44 | call | 11 | 15,50 | call | 5 | 13,85 |
| money | 4 | 20,44 | contribution | 11 | 15,50 | case | 5 | 13,85 |
| contribution | 3 | 15,33 | change | 10 | 14,09 | error | 4 | 11,08 |
| communication | 3 | 15,33 | everything | 9 | 12,68 | advance | 4 | 11,08 |
| | | | progress | 8 | 11,27 | effort | 4 | 11,08 |
| | | | money | 6 | 8,45 | contribution | 3 | 8,31 |
| | | | profit | 6 | 8,45 | comparison | 3 | 8,31 |
| | | | impact | 5 | 7,04 | assumption | 3 | 8,31 |
| | | | point | 4 | 5,64 | reference | 3 | 8,31 |
| | | | comparison | 4 | 5,64 | attack | 3 | 8,31 |
| | | | statement | 2 | 2,82 | progress | 3 | 8,31 |
| | | | claim | 1 | 1,41 | demand | 3 | 8,31 |
| | | | error | 1 | 1,41 | impact | 3 | 8,31 |

When the three different corpora are scrutinized, it seems possible to assert that “make impression”, “make protest” and “make decision” are the most frequently used make+noun collocations used by TUCLE, KTUCLE and LOCNESS participants respectively.

The table above holds a significant finding which shows the difference of collocational usages between native and non-native participants. LOCNESS comprises of 8 make+noun collocations whose evidences are none in the two non-native corpora; “make argument”, “make discovery”, “make case”, “make advance”, “make assumption”, “make reference”, “make attack”

and “make demand” are the collocations which hold no examples neither in TUCLE nor in KTUCLE. Additionally, “make argument” and “make discovery” are the sixth and seventh most used make+noun collocations in LOCNESS with 12 and 10 raw references correspondingly.

Another noteworthy situation about the collocations is that LOCNESS includes 10 make+noun collocations which have no evidences of use in TUCLE; “make profit”, “make statement”, “make claim”, “make point”, “make difference”, “make change”, “make error”, “make effort”, “make comparison” and “make impact” are the collocations whose examples are not evident in TUCLE.

One common make+noun collocation employed by the three corpora participants is “make decision”. The mentioned collocation stands among the three most frequently used make+noun collocations in the three corpora; nevertheless, TUCLE participants hold the usage of highest normalized frequency with 132,87 and 26 raw references. KTUCLE students used “make decision” with the raw reference of 33. However, “make decision” is the most frequently employed make+noun collocation by LOCNESS participants; it is referred 38 times in the 361,054-word native corpus.

“Make impression” has the raw reference of 33 in TUCLE; which makes it the most frequently used make+noun collocation in it. Yet, the interesting thing is that it is not used by the two other corpora for even once. “Make money” stands in the last rows of TUCLE and KTUCLE with LL measures of 20,44 and 8,45. On the other hand, it is striking to observe that the aforementioned collocation is the second most used collocation in the native corpus with LL measure of 66,47.

TUCLE and LOCNESS have a similarity in the LL measure of “make mistake” with 40,88 and 44,31 respectively. In TUCLE, it is the fifth most frequently used collocation while it is the fourth in LOCNESS. On the other hand, TUCLE and KTUCLE have a likeness in the usage of “make contribution” with a very close LL measure of 15,33 and 15,50 respectively. This can pave the way for the inference that non-natives have a similar employment reference of “make contribution”. However, the mentioned collocation is listed in the second last row in TUCLE whereas it stands in the middle rows in KTUCLE.

The least frequently used make+noun collocations in TUCLE are “make money”, “make contribution” and “make communication” with respectively 4, 3 and 3 raw references. KTUCLE has “make statement”, “make claim” and “make error” with 2, 1 and 1 raw references respectively. Finally, LOCNESS holds collocations of “make progress”, “make demand” and “make impact” in the last three rows with 3 raw references for each.

Table 12: Log-likelihood Ratio and Significance Level with “Make” in TUCLE and LOCNESS

| MAKE – TUCLE vs LOCNESS - LL SCORE | | | | | | |
|------------------------------------|------------------|---------------|--------------------|---------------|------|-----|
| | TUCLE 195,684 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| money | 4 | 20,44 | 24 | 66,47 | 6,19 | - |
| progress | 4 | 20,44 | 3 | 8,31 | 1,40 | |
| choice | 6 | 30,66 | 18 | 49,85 | 1,15 | |
| decision | 26 | 132,87 | 38 | 105,20 | 0,82 | |
| contribution | 3 | 15,33 | 3 | 8,31 | 0,55 | |
| call | 4 | 20,44 | 5 | 13,85 | 0,33 | |
| mistake | 8 | 40,88 | 16 | 44,31 | 0,03 | |
| law | 4 | 20,44 | 7 | 19,39 | 0,01 | |

TUCLE, one of the two non-native corpora, and the native corpus LOCNESS include eight collocations in common. Additionally, LOCNESS and TUCLE hold the same make+noun collocation as the most frequently used combinations. TUCLE holds 26 raw references of “make decision” with the normalized frequency of 132,87 which makes the aforementioned collocation the most frequently applied make+noun collocation in TUCLE. The most frequent use of make+noun collocation is employed with again “make decision” in LOCNESS with the raw reference of 38 and normalized frequency of 105,82.

One outstanding collocation usage is of “make money” since it is the most frequently used collocation in the native corpus whereas it is underused in the non-native corpus. LOCNESS holds 24 raw references and normalized frequency of 66,47 of “make money” while TUCLE includes 4 raw references and normalized frequency of 20,44 of the collocation mentioned; therefore, it occurs after the calculation that the aforementioned collocation has the LL measure of -6,19 which makes it underused since it remains under the absolute value of average measure of 3,84.

The remaining make+noun collocations which are used in both corpora are “make progress”, “make choice”, “make contribution”, “make call”, “make mistake” and “make law” among which “make contribution” is one of the least frequently used make+noun collocation in both corpora. There seems a striking evidence of collocational usage in the corpora compared; the least frequently used collocation is the same, “make contribution” collocation has the normalized frequency of 15,33 in TUCLE while it holds the normalized frequency of 8,31 in LOCNESS.

Table 13: Log-likelihood Ratio and Significance Level with “Make” in KTUCLE and LOCNESS

| MAKE – KTUCLE vs LOCNESS - LL SCORE | | | | | | |
|-------------------------------------|-------------------|--------------|--------------------|---------------|-------|-----|
| | KTUCLE 709,748 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| money | 6 | 8,45 | 24 | 66,47 | 27,09 | - |
| decision | 33 | 46,50 | 38 | 105,20 | 11,69 | - |
| claim | 1 | 1,4 | 7 | 19,39 | 10,01 | - |
| profit | 6 | 8,45 | 12 | 33,24 | 8,11 | - |
| statement | 2 | 2,82 | 7 | 19,39 | 7,33 | - |
| choice | 17 | 23,95 | 18 | 49,85 | 4,63 | - |
| error | 1 | 1,4 | 4 | 11,08 | 4,52 | - |
| point | 4 | 5,63 | 7 | 19,39 | 4,09 | - |
| mistake | 18 | 25,36 | 16 | 44,31 | 2,58 | |
| effort | 14 | 19,73 | 4 | 11,08 | 1,14 | |
| contribution | 11 | 15,50 | 3 | 8,31 | 1,02 | |
| difference | 15 | 21,13 | 6 | 16,62 | 0,26 | |
| comparison | 4 | 5,64 | 3 | 8,31 | 0,25 | |
| progress | 8 | 11,27 | 3 | 8,31 | 0,21 | |
| change | 10 | 14,09 | 6 | 16,62 | 0,10 | |
| impact | 5 | 7,04 | 3 | 8,31 | 0,05 | |
| call | 11 | 15,50 | 5 | 13,85 | 0,04 | |

KTUCLE, the second non-native corpus, and the native LOCNESS hold 17 common make+noun collocations whose most frequently used ones are “make decision” in both corpora. KTUCLE includes 33 raw references and the normalized frequency of 46,50 of the aforementioned collocations whereas LOCNESS consists of 38 raw references and the normalized frequency of 105,20 “make decision”. Nevertheless, the usage of “make decision” remains underused in KTUCLE with the LL measure of -11,69 although it is the most frequently used collocation.

The collocations “make choice” and “make mistake” have similar raw references in both corpora; KTUCLE holds 17 raw references of “make choice” whereas LOCNESS includes 18 raw references. However, the aforementioned collocation remains underused in KTUCLE with the LL measure of -4,63. Another collocation used with similar raw references is “make mistake”, it is included 18 times in KTUCLE while involved 16 times in LOCNESS. Yet, it does not seem to be an underused collocation as “make choice”; it remains in normal accepted measure. The least frequently used make+noun combinations are “make claim” and “make error” with 1 raw reference; they additionally are underused according to the LL measure which displays that the former collocation had the measure -10,01 while the latter has -4,52 measure.

The underused collocations in KTUCLE are “make money”, “make decision”-which is the most frequently used one as well-, “make claim” which is one of the two least frequently used ones, “make profit”, “make statement”, “make choice”, “make error” which is also one of the two least frequently used ones, and “make point”. In other words, KTUCLE holds 8 underused make+noun collocations. The collocations “make comparison”, and “make impact” are not used as frequently as the other collocations; however, they do not occur as underused since the native corpus participants do not apply the usage of the aforementioned collocations in the same number.

Table 14: Distribution of “have” in TUCLE, KTUCLE and LOCNESS

| HAVE | | | | | | | | |
|------------------|-----|--------|-------------------|-----|--------|--------------------|-----|--------|
| TUCLE 195,684 | | | KTUCLE 709,748 | | | LOCNESS 361,054 | | |
| OBJECTS | RAW | NORM | OBJECTS | RAW | NORM | OBJECTS | RAW | NORM |
| right | 100 | 511,03 | right | 153 | 215,60 | child | 71 | 196,60 |
| abortion | 39 | 199,30 | problem | 128 | 180,35 | right | 62 | 171,72 |
| knowledge | 29 | 148,20 | effect | 110 | 154,98 | effect | 45 | 124,64 |
| gun | 29 | 148,20 | phone | 100 | 140,90 | chance | 24 | 66,47 |
| advantage | 27 | 137,98 | information | 90 | 126,81 | choice | 23 | 63,70 |
| child | 25 | 127,76 | university | 82 | 115,53 | sex | 20 | 55,39 |
| university | 23 | 117,54 | life | 78 | 109,90 | power | 20 | 55,39 |
| idea | 22 | 112,43 | time | 74 | 104,26 | time | 20 | 52,62 |
| baby | 20 | 102,21 | child | 71 | 100,04 | problem | 18 | 49,85 |
| difficulty | 18 | 91,99 | knowledge | 56 | 78,90 | idea | 15 | 41,55 |
| time | 17 | 86,87 | advantage | 54 | 76,08 | money | 15 | 41,55 |
| problem | 17 | 86,87 | idea | 50 | 70,45 | abortion | 12 | 33,24 |
| chance | 16 | 81,76 | role | 45 | 63,40 | baby | 11 | 30,47 |
| feature | 15 | 76,65 | chance | 44 | 62 | system | 11 | 30,47 |
| benefit | 15 | 76,65 | responsibility | 41 | 57,77 | impact | 10 | 27,70 |
| phone | 15 | 76,65 | benefit | 41 | 57,77 | nothing | 10 | 27,70 |
| harm | 14 | 71,54 | friend | 40 | 56,36 | access | 9 | 24,93 |
| feeling | 13 | 66,43 | money | 34 | 47,90 | advantage | 9 | 24,93 |
| effect | 13 | 66,43 | place | 33 | 46,50 | reason | 9 | 24,93 |
| job | 12 | 61,32 | homework | 33 | 46,50 | strength | 8 | 22,16 |
| experience | 11 | 56,21 | ability | 31 | 43,68 | control | 8 | 22,16 |
| opportunity | 11 | 56,21 | fun | 30 | 42,27 | life | 8 | 22,16 |
| opinion | 11 | 56,21 | difficulty | 29 | 40,86 | purpose | 7 | 19,39 |
| life | 11 | 56,21 | opportunity | 29 | 40,86 | sympathy | 7 | 19,39 |
| place | 10 | 51,10 | reason | 29 | 40,86 | aids | 7 | 19,39 |
| reason | 9 | 45,99 | job | 28 | 39,45 | interest | 7 | 19,39 |
| responsibility | 9 | 45,99 | power | 18 | 25,36 | consequence | 7 | 19,39 |
| power | 8 | 40,88 | impact | 17 | 23,95 | opinion | 7 | 19,39 |
| impact | 7 | 35,77 | experience | 17 | 23,95 | responsibility | 7 | 19,39 |
| choice | 6 | 30,66 | family | 16 | 22,54 | need | 7 | 19,39 |
| money | 6 | 30,66 | system | 15 | 21,13 | desire | 6 | 16,62 |
| nothing | 6 | 30,66 | way | 13 | 18,32 | experience | 6 | 16,62 |
| control | 6 | 30,66 | thing | 13 | 18,32 | name | 6 | 16,62 |
| relationship | 6 | 30,66 | choice | 12 | 16,91 | respect | 6 | 16,62 |

Table 14-continued

| HAVE | | | | | | | | |
|------------------|-----|--------------|-------------------|-----|--------------|--------------------|-----|--------------|
| TUCLE 195,684 | | | KTUCLE 709,748 | | | LOCNESS 361,054 | | |
| OBJECTS | RAW | NORM | OBJECTS | RAW | NORM | OBJECTS | RAW | NORM |
| parent | 6 | 30,66 | marriage | 12 | 16,91 | influence | 6 | 16,62 |
| disability | 5 | 25,55 | access | 11 | 15,50 | meaning | 6 | 16,62 |
| fun | 5 | 25,55 | purpose | 11 | 15,50 | opportunity | 6 | 16,62 |
| structure | 5 | 25,55 | baby | 11 | 15,50 | rate | 6 | 16,62 |
| history | 5 | 25,55 | interest | 10 | 14,09 | knowledge | 6 | 16,62 |
| clone | 5 | 25,55 | opinion | 10 | 14,09 | sense | 6 | 16,62 |
| rate | 5 | 25,55 | value | 10 | 14,09 | team | 6 | 16,62 |
| friend | 5 | 25,55 | nothing | 9 | 12,68 | man | 6 | 16,62 |
| way | 5 | 25,55 | sense | 9 | 12,68 | diploma | 5 | 13,85 |
| homework | 5 | 25,55 | freedom | 8 | 11,27 | difficulty | 5 | 13,85 |
| history | 5 | 25,55 | sex | 7 | 9,86 | history | 5 | 13,85 |
| way | 5 | 25,55 | respect | 7 | 9,86 | implication | 5 | 13,85 |
| family | 5 | 25,55 | influence | 7 | 9,86 | authority | 5 | 13,85 |
| influence | 4 | 20,44 | facility | 5 | 7,04 | boy | 5 | 13,85 |
| ability | 4 | 20,44 | control | 4 | 5,64 | ability | 5 | 13,85 |
| freedom | 4 | 20,44 | consequence | 4 | 5,64 | room | 5 | 13,85 |
| law | 4 | 20,44 | need | 4 | 5,64 | feeling | 5 | 13,85 |
| thing | 4 | 20,44 | desire | 4 | 5,64 | aspect | 5 | 13,85 |
| respect | 3 | 15,33 | aim | 4 | 5,64 | freedom | 5 | 13,85 |
| system | 3 | 15,33 | anything | 4 | 5,64 | job | 5 | 13,85 |
| access | 3 | 15,33 | loss | 4 | 5,64 | law | 5 | 13,85 |
| meaning | 3 | 15,33 | something | 4 | 5,64 | place | 5 | 13,85 |
| aspect | 3 | 15,33 | feeling | 3 | 4,23 | way | 5 | 13,85 |
| purpose | 2 | 10,22 | car | 3 | 4,23 | facility | 4 | 11,08 |
| name | 2 | 10,22 | meaning | 2 | 2,82 | marriage | 4 | 11,08 |
| man | 2 | 10,22 | rate | 2 | 2,82 | aim | 4 | 11,08 |
| diploma | 2 | 10,22 | parent | 2 | 2,82 | parent | 4 | 11,08 |
| authority | 2 | 10,22 | support | 2 | 2,82 | support | 4 | 11,08 |
| marriage | 2 | 10,22 | team | 1 | 1,41 | anything | 4 | 11,08 |
| aim | 2 | 10,22 | history | 1 | 1,41 | attitude | 4 | 11,08 |
| something | 2 | 10,22 | implication | 1 | 1,41 | program | 4 | 11,08 |
| sex | 1 | 5,11 | authority | 1 | 1,41 | car | 4 | 11,08 |
| sympathy | 1 | 5,11 | boy | 1 | 1,41 | value | 4 | 11,08 |
| interest | 1 | 5,11 | aspect | 1 | 1,41 | loss | 4 | 11,08 |
| sense | 1 | 5,11 | law | 1 | 1,41 | family | 4 | 11,08 |
| | | | attitude | 1 | 1,41 | something | 4 | 11,08 |
| | | | program | 1 | 1,41 | thing | 4 | 11,08 |

The distribution of have+noun collocations occupies the largest place of the tables; the table includes 70 rows of collocations some of which are used in common among the three corpora. The outstanding thing is that the have+noun combination “have right” is the most frequently used collocation in both non-native corpora. In TUCLE the mentioned collocation is applied for 100

times whereas it is used for 153 times in the second non-native corpus KTUCLE. The attention-grabbing thing is that “have right” is the second most used collocation in the native corpus LOCNESS with 62 raw references.

Although the two non-native corpora have the same have+noun collocation as the most frequently used ones, the second and third most used ones are not similar; TUCLE involves “have abortion” and “have knowledge” as the second and third frequent usages respectively whereas KTUCLE holds “have problem” and “have effect” as the second and third most frequently used have+noun collocations. Additionally, KTUCLE’s third most used have+noun collocation is also standing in the third row of LOCNESS; in other words, KTUCLE and LOCNESS hold the same collocation as the third most frequently used one in their lists. However, the raw reference of “have effect” remains distinct since KTUCLE participants applied the aforementioned collocation for 110 times while LOCNESS contributors used it for 45 times. Their normalized frequencies are 154,98 in KTUCLE and 124,64 in LOCNESS.

The native corpus holds 17 have+noun collocations some of which are inexistent in either TUCLE or KTUCLE, or in neither non-native corpora. As an evidence, “have strength” is only present in LOCNESS; the other two corpora do not include the mentioned collocation. For further examples, “have aids” and “have room” take place in the native corpus whereas there are not any usages of those collocations in TUCLE and KTUCLE. In addition, there are a few have+noun collocations which occur in TUCLE and LOCNESS whereas no examples of those collocations can be encountered in KTUCLE; “have abortion”, “have sympathy”, “have name”, “have man” and “have diploma” are instances of have+noun collocations applied in both LOCNESS and TUCLE. Yet, their raw references and normalized frequencies differ; “have abortion” holds the normalized frequency of 199,30 in TUCLE while LOCNESS includes the frequency as 33,24. One outstanding difference of raw reference belongs to the collocation “have sympathy” since it is applied only once in TUCLE whereas the raw reference of the aforementioned collocation is 7 in LOCNESS.

Having mentioned the common collocations between TUCLE and LOCNESS, the mutual collocations between KTUCLE and LOCNESS are as follows:

- “have consequence”,
- “have need”,
- “have desire”,
- “have team”,
- “have implication”,
- “have body”,
- “have facility”,
- “have support”,

“have anything”,
“have attitude”,
“have program”,
“have car”,
“have value”
“have loss”

Whereas the raw references of the commonly used collocations stand similar, their normalized frequencies differ; “have facility”, “have anything” and “have car” hold alike references around 3 and 4, however, “have facility” carries the normalized frequency of 7,04 in KTUCLE whereas LOCNESS includes the mentioned feature as 11,08. When it comes to “have anything” it seems possible to assert that it has the normalized frequency of 5,64 in KTUCLE and 11,08 in LOCNESS. The main difference lies in the collocation “have car”; KTUCLE includes the normalized frequency as 4,23 while LOCNESS consists it as 11,08.

TUCLE comprises “have sympathy”, “have interest”, “have sex” and “have sense” as the least frequently used have+noun collocations with only 1 raw references. One notable situation arises here; the most frequently used have+noun collocation in TUCLE is “have right” with 100 raw references and the normalized frequency of 511,03 whereas “have sympathy” has the normalized frequency of 5,11 which seems disparate when compared.

Though TUCLE and KTUCLE involve “have right” as the most frequently used have+noun collocation, KTUCLE comprises 153 raw references whereas TUCLE includes 100 raw references. In spite of the closeness of references of “have right”, the normalized frequencies create a diverse position with 511,03 in TUCLE and 215,60 in KTUCLE. TUCLE participants applied “have right” more than double when compared to KTUCLE.

TUCLE includes “have knowledge” and “have gun” with raw references of 29 and the normalized frequency of 148,20. A similar raw reference and a normalized frequency in KTUCLE comes from “have effect” and “have phone”; yet, the raw references of 110 and 100. This examination reveals the fact that although raw reference numbers in KTUCLE are much higher than the ones in TUCLE, this is the evidence that the size of the corpus matters in forming the normalized frequencies.

The least frequently used have+noun collocations of KTUCLE “have team”, “have history”, “have implication”, “have authority”, “have boy”, “have aspect”, “have law”, “have attitude” and “have program” hold only 1 references for each; and their normalized frequencies are 1,41. However, “have boy” is an example of misuse since there are not any occurrences of it in neither

BNC nor the collocation dictionary. The have+noun collocation table includes only one evidence of misuse with “have boy”.

Table 15: Log-likelihood Ratio and Significance Level with “have” in TUCLE and LOCNESS

| HAVE – TUCLE vs LOCNESS - LL SCORE | | | | | | |
|---|-------------------------|---------------|---------------------------|---------------|-----------|------------|
| | TUCLE 195,684 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| right | 100 | 511,03 | 62 | 171,72 | 47,24 | + |
| abortion | 39 | 199,30 | 12 | 33,24 | 36,30 | + |
| knowledge | 29 | 148,20 | 6 | 16,62 | 33,77 | + |
| advantage | 27 | 137,98 | 9 | 24,93 | 23,77 | + |
| difficulty | 18 | 91,99 | 5 | 13,85 | 17,89 | + |
| sex | 1 | 5,11 | 20 | 55,39 | 11,37 | - |
| baby | 20 | 102,21 | 11 | 30,47 | 11,03 | + |
| feeling | 13 | 66,43 | 5 | 13,85 | 10,25 | + |
| idea | 22 | 112,43 | 15 | 41,55 | 9,04 | + |
| job | 12 | 61,32 | 5 | 13,85 | 8,83 | + |
| place | 10 | 51,10 | 5 | 13,85 | 6,15 | + |
| experience | 11 | 56,21 | 6 | 16,62 | 6,13 | + |
| opportunity | 11 | 56,21 | 6 | 16,62 | 6,13 | + |
| opinion | 11 | 56,21 | 7 | 19,39 | 5,01 | + |
| effect | 13 | 66,43 | 45 | 124,64 | 4,44 | - |
| life | 11 | 56,21 | 8 | 22,16 | 4,07 | + |
| child | 25 | 127,76 | 71 | 196,60 | 3,66 | |
| responsibility | 9 | 45,99 | 7 | 19,39 | 2,95 | |
| choice | 6 | 30,66 | 23 | 63,70 | 2,90 | |
| problem | 17 | 86,87 | 18 | 49,85 | 2,65 | |
| parent | 6 | 30,66 | 4 | 11,08 | 2,55 | |
| sympathy | 1 | 5,11 | 7 | 19,39 | 2,13 | |
| interest | 1 | 5,11 | 7 | 19,39 | 2,13 | |
| time | 17 | 86,87 | 20 | 52,62 | 1,82 | |
| reason | 9 | 45,99 | 9 | 24,93 | 1,66 | |
| family | 5 | 25,55 | 4 | 11,08 | 1,56 | |
| sense | 1 | 5,11 | 6 | 16,62 | 1,55 | |
| system | 3 | 15,33 | 11 | 30,47 | 1,25 | |
| history | 5 | 25,55 | 5 | 13,85 | 0,92 | |

Table 15-continued

| HAVE – TUCLE vs LOCNESS - LL SCORE | | | | | | |
|------------------------------------|------------------|--------------|--------------------|--------------|------|-----|
| | TUCLE 195,684 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| way | 5 | 25,55 | 5 | 13,85 | 0,92 | |
| thing | 4 | 20,44 | 4 | 11,08 | 0,74 | |
| purpose | 2 | 10,22 | 7 | 19,39 | 0,71 | |
| access | 3 | 15,33 | 9 | 24,93 | 0,57 | |
| power | 8 | 40,88 | 20 | 55,39 | 0,55 | |
| rate | 5 | 25,55 | 6 | 16,62 | 0,49 | |
| chance | 16 | 81,76 | 24 | 66,47 | 0,41 | |
| money | 6 | 30,66 | 15 | 41,55 | 0,41 | |
| name | 2 | 10,22 | 6 | 16,62 | 0,38 | |
| man | 2 | 10,22 | 6 | 16,62 | 0,38 | |
| control | 6 | 30,66 | 8 | 22,16 | 0,35 | |
| ability | 4 | 20,44 | 5 | 13,85 | 0,33 | |
| freedom | 4 | 20,44 | 5 | 13,85 | 0,33 | |
| law | 4 | 20,44 | 5 | 13,85 | 0,33 | |
| impact | 7 | 35,77 | 10 | 27,70 | 0,26 | |
| diploma | 2 | 10,22 | 5 | 13,85 | 0,14 | |
| authority | 2 | 10,22 | 5 | 13,85 | 0,14 | |
| influence | 4 | 20,44 | 6 | 16,62 | 0,10 | |
| nothing | 6 | 30,66 | 10 | 27,70 | 0,04 | |
| aspect | 3 | 15,33 | 5 | 13,85 | 0,02 | |
| respect | 3 | 15,33 | 6 | 16,62 | 0,01 | |
| meaning | 3 | 15,33 | 6 | 16,62 | 0,01 | |
| marriage | 2 | 10,22 | 4 | 11,08 | 0,01 | |
| aim | 2 | 10,22 | 4 | 11,08 | 0,01 | |
| something | 2 | 10,22 | 4 | 11,08 | 0,01 | |

As scrutinized in detail in the distribution table of “have” in the three corpora, it stood out that the collocation “have right” is the most frequently used one in both TUCLE and KTUCLE. When TUCLE is compared to LOCNESS in terms of the have+noun collocations, it can be stated that “have right” is overused since it holds the log likelihood measure of +47,24. LOCNESS involves 62 raw references of “have right” even though it is a larger corpus than TUCLE in size. There stand 14 have+noun collocations in total which are overused when compared to LOCNESS; they are listed as “have right”, “have abortion”, “have knowledge”, “have advantage”, “have

difficulty”, “have baby”, “have feeling”, “have idea”, “have job”, “have place”, “have experience”, “have opportunity”, “have opinion” and “have life”.

The LL measure of “have abortion” is +36,30 which makes it an overused have+noun collocation in TUCLE. The reason for making it overused rests in the argumentative essay topic which asked participants to write about their opinions on abortion. One another overused have+noun collocation is “have knowledge” with the LL measure of +33,77; in addition, “have advantage” holds +23,77 LL measure and “have difficulty” carries the LL measure of +17,89 which leads to the inference that they are all overused have+noun collocations.

In addition to the aforementioned overused collocations, there stand 2 have+noun collocations in TUCLE that are underused when compared to LOCNESS; “have sex” and “have effect” are the have+noun collocations which are underused with the LL measures of -11,37 and -4,44 respectively. The collocation “have sex” has only 1 raw reference with the normalized frequency of 5,11 in TUCLE while it has 20 raw references and the normalized frequency of 55,39 in the native corpus LOCNESS. Besides, “have effect” has 13 raw references in TUCLE whereas it is used 45 times in the native corpus. This table does not hold any examples of misuse.

Table 16: Log-likelihood Ratio and Significance Level with “have” in KTUCLE and LOCNESS

| HAVE – KTUCLE vs LOCNESS - LL SCORE | | | | | | |
|-------------------------------------|-------------------|---------------|--------------------|---------------|-------|-----|
| OBJECTS | KTUCLE 709,748 | | LOCNESS 361,054 | | LL | + - |
| | RAW | NORM | RAW | NORM | | |
| problem | 128 | 180,35 | 18 | 49,85 | 35,38 | + |
| life | 78 | 109,90 | 8 | 22,16 | 28,32 | + |
| knowledge | 56 | 78,90 | 6 | 16,62 | 19,68 | + |
| sex | 7 | 9,86 | 20 | 55,39 | 18,34 | - |
| child | 71 | 100,04 | 71 | 196,60 | 15,92 | - |
| choice | 12 | 16,91 | 23 | 63,70 | 14,87 | - |
| advantage | 54 | 76,08 | 9 | 24,93 | 12,31 | + |
| responsibility | 41 | 57,77 | 7 | 19,39 | 9,06 | + |
| place | 33 | 46,50 | 5 | 13,85 | 8,42 | + |
| ability | 31 | 43,68 | 5 | 13,85 | 7,36 | + |
| time | 74 | 104,26 | 20 | 52,62 | 7,04 | + |
| difficulty | 29 | 40,86 | 5 | 13,85 | 6,33 | + |
| history | 1 | 1,41 | 5 | 13,85 | 6,29 | - |
| law | 1 | 1,41 | 5 | 13,85 | 6,29 | - |
| authority | 1 | 1,41 | 5 | 13,85 | 6,29 | - |
| aspect | 1 | 1,41 | 5 | 13,85 | 6,29 | - |
| job | 28 | 39,45 | 5 | 13,85 | 5,83 | + |
| power | 18 | 25,36 | 20 | 55,39 | 5,72 | - |
| rate | 2 | 2,82 | 6 | 16,62 | 5,69 | - |
| meaning | 2 | 2,82 | 6 | 16,62 | 5,69 | - |
| control | 4 | 5,64 | 8 | 22,16 | 5,41 | - |
| opportunity | 29 | 40,86 | 6 | 16,62 | 4,83 | + |
| idea | 50 | 70,45 | 15 | 41,55 | 3,51 | |
| nothing | 9 | 12,68 | 10 | 27,70 | 2,86 | |
| feeling | 3 | 4,23 | 5 | 13,85 | 2,75 | |
| parent | 2 | 2,82 | 4 | 11,08 | 2,70 | |
| baby | 11 | 15,50 | 11 | 30,47 | 2,47 | |
| right | 153 | 215,60 | 62 | 171,72 | 2,35 | |
| family | 16 | 22,54 | 4 | 11,08 | 1,84 | |
| reason | 29 | 40,86 | 9 | 24,93 | 1,82 | |
| effect | 110 | 154,98 | 45 | 124,64 | 1,56 | |
| access | 11 | 15,50 | 9 | 24,93 | 1,09 | |
| aim | 4 | 5,64 | 4 | 11,08 | 0,90 | |
| something | 4 | 5,64 | 4 | 11,08 | 0,90 | |
| influence | 7 | 9,86 | 6 | 16,62 | 0,86 | |
| respect | 7 | 9,86 | 6 | 16,62 | 0,86 | |
| thing | 13 | 18,32 | 4 | 11,08 | 0,84 | |
| system | 15 | 21,13 | 11 | 30,47 | 0,83 | |
| experience | 17 | 23,95 | 6 | 16,62 | 0,63 | |
| marriage | 12 | 16,91 | 4 | 11,08 | 0,57 | |
| opinion | 10 | 14,09 | 7 | 19,39 | 0,41 | |
| interest | 10 | 14,09 | 7 | 19,39 | 0,41 | |
| way | 13 | 18,32 | 5 | 13,85 | 0,29 | |
| sense | 9 | 12,68 | 6 | 16,62 | 0,26 | |
| money | 34 | 47,90 | 15 | 41,55 | 0,21 | |
| purpose | 11 | 15,50 | 7 | 19,39 | 0,21 | |
| impact | 17 | 23,95 | 10 | 27,70 | 0,13 | |
| freedom | 8 | 11,27 | 5 | 13,85 | 0,13 | |
| chance | 44 | 62 | 24 | 66,47 | 0,08 | |

When it comes to the comparison between KTUCLE and LOCNESS, it can be withdrawn that there are 11 overused and 11 underused have+noun collocations in KTUCLE when compared to the native corpus. “Have problem”, “have life”, “have knowledge”, “have advantage”, “have responsibility”, “have place”, “have ability”, “have time”, “have difficulty”, “have job” and “have opportunity” are the overused have+noun collocations with their log likelihood measures.

The biggest difference lies between the usage of “have problem”; in KTUCLE it includes the raw reference of 128 and the normalized frequency of 180,35 whereas LOCNESS holds 18 raw references and the normalized frequency of 49,85. Therefore, the LL measure happens to be +35,38 in KTUCLE that makes the aforementioned collocation an overused one. The second outstanding difference is on the use of “have life” which involves raw references of 78 in KTUCLE and 8 in LOCNESS; thus, it makes the collocation overused with the LL measure of +28,32.

As opposed to TUCLE, KTUCLE holds 11 underused have+noun collocations when compared to LOCNESS as TUCLE has only 2 of them. “Have sex” and “have effect” are the underused have+noun collocations in TUCLE; similarly, “have sex” is underused in KTUCLE as well. KTUCLE includes 7 raw references of the abovementioned collocation whereas TUCLE includes only 1 with the normalized frequencies of 9,86 and 5,11 respectively. The aforesaid have+noun collocation has the LL measure of -18,34 in KTUCLE. Additionally, “have child” and “have choice” are the two other have+noun collocations which carry the LL measure similar to “have sex” with -15,92 and -14,87 correspondingly.

A notable have+noun collocation occurs to be “have child” since it holds the same number of raw references both in KTUCLE and LOCNESS; both corpora include 71 usages of the aforesaid collocation. Yet, the normalized frequency is 100,04 in KTUCLE whereas it is 196,60 in LOCNESS. So, the LL measure happens to be -15,92 which is a result of the size difference between the corpora; KTUCLE is a corpus of 709,748 words while LOCNESS includes 361,054 words that make the latter a smaller corpus in size.

Table 17: Distribution of “give” in TUCLE, KTUCLE and LOCNESS

| GIVE | | | | | | | | |
|------------------|-----|---------------|-------------------|-----|---------------|--------------------|-----|--------------|
| TUCLE 195,684 | | | KTUCLE 709,748 | | | LOCNESS 361,054 | | |
| OBJECTS | RAW | NORM | OBJECTS | RAW | NORM | OBJECTS | RAW | NORM |
| homework | 107 | 546,80 | homework | 106 | 149,30 | money | 11 | 30,47 |
| birth | 21 | 107,32 | example | 89 | 125,40 | chance | 9 | 24,93 |
| example | 20 | 102,21 | information | 30 | 42,27 | birth | 8 | 22,16 |
| money | 8 | 40,88 | importance | 21 | 29,59 | power | 8 | 22,16 |
| result | 7 | 35,77 | result | 21 | 29,59 | reason | 6 | 16,62 |
| answer | 6 | 30,66 | birth | 19 | 26,77 | time | 6 | 16,62 |
| punishment | 4 | 20,44 | money | 15 | 21,13 | life | 5 | 13,85 |
| information | 4 | 20,44 | opportunity | 14 | 19,73 | rise | 4 | 11,08 |
| life | 4 | 20,44 | message | 13 | 18,32 | evidence | 4 | 11,08 |
| importance | 3 | 15,33 | damage | 13 | 18,32 | example | 4 | 11,08 |
| chance | 3 | 15,33 | chance | 11 | 15,50 | permission | 3 | 8,31 |
| phone | 3 | 15,33 | answer | 10 | 14,09 | priority | 3 | 8,31 |
| assignment | 2 | 10,22 | harm | 10 | 14,09 | reasoning | 3 | 8,31 |
| portion | 2 | 10,22 | assignment | 9 | 12,68 | statistic | 3 | 8,31 |
| support | 2 | 10,22 | permission | 8 | 11,27 | hope | 3 | 8,31 |
| point | 2 | 10,22 | love | 8 | 11,27 | treatment | 3 | 8,31 |
| drug | 2 | 10,22 | decision | 7 | 9,86 | respect | 3 | 8,31 |
| lesson | 2 | 10,22 | sense | 6 | 8,45 | something | 3 | 8,31 |
| education | 2 | 10,22 | education | 6 | 8,45 | idea | 3 | 8,31 |
| idea | 2 | 10,22 | time | 6 | 8,45 | right | 3 | 8,31 |
| time | 2 | 10,22 | power | 5 | 7,04 | | | |
| permission | 1 | 5,11 | rise | 4 | 5,64 | | | |
| rise | 1 | 5,11 | right | 4 | 5,64 | | | |
| right | 1 | 5,11 | reason | 2 | 2,82 | | | |
| | | | hope | 2 | 2,82 | | | |
| | | | idea | 2 | 2,82 | | | |
| | | | respect | 1 | 1,41 | | | |

When give+noun collocations are distributed in tables, it can be withdrawn that the two non-native corpora hold the same two give+noun collocations in the row of the first three most used collocations. They are seen as “give homework” and “give example”. However, “give homework” stands with almost identical raw reference numbers; TUCLE includes 107 raw references when KTUCLE involves 106 references. Yet, their normalized frequencies differ a lot since KTUCLE is much bigger than TUCLE in size; therefore, the normalized frequency of “give homework” in TUCLE is 546,80 whereas it is 149,30 in KTUCLE. This example is a strong proof of the importance of normalized frequency values; although raw reference numbers are too much alike, the normalized frequencies are at variance.

Having mentioned the first most frequently used give+noun collocation in the two non-native corpora, the second and third most used ones need evince; “give example” is the second common

used collocation while their third most frequent ones differ; TUCLE includes “give birth” whereas KTUCLE holds “give information”. When it comes to the scrutiny of “give example”, their normalized frequencies seem similar despite the difference in raw reference numbers; TUCLE includes 20 usages of “give example” while KTUCLE takes in 89 usages. However, since KTUCLE is much bigger than TUCLE, the normalized frequency of “give example” is revealed to be 125,40 in the former corpus whereas it is 102,21 in the latter one. The two non-native corpora contain “give birth” and “give information” in their three most frequently used give+noun collocations.

A similarity takes place between LOCNESS and TUCLE with the “give birth” collocation whose raw reference is 21 in TUCLE and 9 in LOCNESS. KTUCLE does not include a common give+noun collocation with LOCNESS whereas TUCLE holds one; “give birth”. Nevertheless, the normalized frequencies diverge as the aforementioned collocation in TUCLE has the normalized frequency of 107,32 whereas LOCNESS has the normalized frequency of 22,16. However, this does not change the fact that “give birth” is included in the three most used give+noun collocations in TUCLE and LOCNESS.

LOCNESS holds 6 give+noun collocations of which no evidence can be seen in the two non-native corpora; “give evidence”, “give priority”, “give reasoning”, “give statistic”, “give treatment” and “give something” are the give+noun collocations that have no examples in TUCLE and KTUCLE. In addition to these, there are 4 give+noun collocations which has no usages in TUCLE although they can be encountered in KTUCLE; “give power”, “give reason”, “give hope” and “give respect” are examples of the collocations which cannot be faced in TUCLE. Additionally, there is 1 give+noun collocation in LOCNESS which is not existent in KTUCLE; “give life” is visible in TUCLE although no examples can be seen in KTUCLE.

The least used give+noun collocations in TUCLE are “give permission”, “give rise” and “give right” with 1 raw references for each; and the normalized frequency of the mentioned collocation is 5,11. The second non-native corpus KTUCLE includes “give hope”, “give idea” and “give respect” as the least frequently used give+noun collocations with the normalized frequencies of 2,82 for the first two and 1,41 respectively. LOCNESS consists of “give something”, “give idea” and “give right” as the least used give+noun collocations. A noteworthy situation takes place in this situation; “give right” is among the three least frequently used give+noun collocations both in TUCLE and LOCNESS. In addition, “give idea” is among the three least frequently used give+noun collocations both in KTUCLE and LOCNESS.

One example of misuse of give+noun collocation is encountered in TUCLE with “give phone” that seems to be an evidence of L1 transfer; it sounds Turkish. Moreover, the student/s used

it for 3 times. The participant may have intended to mean “give phone number”; however, s/he directly translated from Turkish.

Table 18: Log-likelihood Ratio and Significance Level with “give” in TUCLE and LOCNESS

| GIVE – TUCLE vs LOCNESS - LL SCORE | | | | | | |
|------------------------------------|------------------|---------------|--------------------|--------------|-------|-----|
| | TUCLE 195,684 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| example | 20 | 102,21 | 4 | 11,08 | 23,66 | + |
| birth | 21 | 107,32 | 8 | 22,16 | 16,68 | + |
| chance | 3 | 15,33 | 9 | 24,93 | 0,57 | |
| rise | 1 | 5,11 | 4 | 11,08 | 0,55 | |
| money | 8 | 40,88 | 11 | 30,47 | 0,39 | |
| time | 2 | 10,22 | 6 | 16,62 | 0,38 | |
| life | 4 | 20,44 | 5 | 13,85 | 0,33 | |
| permission | 1 | 5,11 | 3 | 8,31 | 0,19 | |
| right | 1 | 5,11 | 3 | 8,31 | 0,19 | |
| idea | 2 | 10,22 | 3 | 8,31 | 0,05 | |

As mentioned in the interpretation of the previous table, “give birth” is in the first three most frequently used give+noun collocations in TUCLE and LOCNESS. The raw frequencies of the collocation are 21 and 8 respectively; however, their normalized frequencies differentiate a lot as TUCLE has the normalized frequency of 107,32 whereas LOCNESS includes the normalized frequency as 22,16. Therefore, when the two corpora are compared in terms of normalized frequencies LL measure occurs to be +16,68 which means that “give birth” is overused in TUCLE.

Additionally, “give example” is another evidence of overuse since it has the raw reference of 20 in TUCLE whereas LOCNESS consists of 4 raw references. Although LOCNESS is larger than TUCLE, the usage of “give example” is fewer in LOCNESS than of TUCLE; therefore, the normalized frequency is 102,21 in TUCLE and 11,08 in LOCNESS which can lead to the result that “give example” is overused in TUCLE with the LL measure of +23,66.

Table 19: Log-likelihood Ratio and Significance Level with “give” in KTUCLE and LOCNESS

| GIVE – KTUCLE vs LOCNESS - LL SCORE | | | | | | |
|-------------------------------------|-------------------|---------------|--------------------|--------------|-------|-----|
| | KTUCLE 709,748 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| example | 89 | 125,40 | 4 | 11,08 | 48,90 | + |
| reason | 2 | 2,82 | 6 | 16,62 | 5,69 | - |
| power | 5 | 7,04 | 8 | 22,16 | 4,18 | - |
| respect | 1 | 1,41 | 3 | 8,31 | 2,85 | |
| hope | 2 | 2,82 | 3 | 8,31 | 1,44 | |
| idea | 2 | 2,82 | 3 | 8,31 | 1,44 | |
| time | 6 | 8,45 | 6 | 16,62 | 1,35 | |
| chance | 11 | 15,50 | 9 | 24,93 | 1,09 | |
| rise | 4 | 5,64 | 4 | 11,08 | 0,90 | |
| money | 15 | 21,13 | 11 | 30,47 | 0,83 | |
| right | 4 | 5,64 | 3 | 8,31 | 0,25 | |
| birth | 19 | 26,77 | 8 | 22,16 | 0,21 | |
| permission | 8 | 11,27 | 3 | 8,31 | 0,21 | |

When KTUCLE and LOCNESS are compared in terms of give+noun collocations, it seems possible to state that there is only one give+noun collocation which is overused; “give example” has the LL measure of +48,90. The raw reference of the aforesaid collocation is 89 in KTUCLE with the normalized frequency of 125,40 whereas LOCNESS has 4 references and the normalized frequency of 11,08 of “give example”. The huge difference in the normalized frequencies create the LL measure as +48,90 which make it an overused give+noun collocation.

In addition to the 1 overused give+noun collocation, there are 2 underused give+noun collocations in KTUCLE. “Give reason” and “give power” are evidences of underuse with the normalized frequencies of 2,82 and 7,04 in KTUCLE while the normalized frequencies are 16,62 and 22,16 in LOCNESS. The difference in “give reason” creates the LL measure as -5,69 and “give power” forms the log likelihood measure as -4,18 which make the two abovesaid give+noun collocations underused in KTUCLE. The other give+noun collocations remain within the average values; therefore, it means that they are neither overused nor underused examples.

Table 20: Distribution of “take” in TUCLE, KTUCLE and LOCNESS

| TAKE | | | | | | | | |
|------------------|-----|--------------|-------------------|-----|--------------|--------------------|-----|--------------|
| TUCLE 195,684 | | | KTUCLE 709,748 | | | LOCNESS 361,054 | | |
| OBJECTS | RAW | NORM | OBJECTS | RAW | NORM | OBJECTS | RAW | NORM |
| precaution | 11 | 56,21 | responsibility | 30 | 42,30 | responsibility | 15 | 41,55 |
| responsibility | 10 | 51,10 | time | 26 | 36,63 | life | 15 | 41,55 |
| photo | 8 | 40,88 | education | 24 | 33,81 | advantage | 13 | 36,01 |
| life | 8 | 40,88 | precaution | 22 | 31 | time | 10 | 27,70 |
| education | 5 | 25,55 | photo | 22 | 31 | action | 8 | 22,16 |
| advantage | 5 | 25,55 | exam | 20 | 28,18 | course | 6 | 16,62 |
| time | 5 | 25,55 | risk | 17 | 23,95 | risk | 6 | 16,62 |
| gun | 5 | 25,55 | advantage | 17 | 23,95 | decision | 6 | 16,62 |
| measure | 4 | 20,44 | measure | 14 | 19,73 | precaution | 5 | 13,85 |
| photograph | 3 | 15,33 | note | 14 | 19,73 | refuge | 5 | 13,85 |
| revenge | 3 | 15,33 | lesson | 14 | 19,73 | egg | 5 | 13,85 |
| picture | 3 | 15,33 | hour | 13 | 18,32 | notice | 4 | 11,08 |
| risk | 3 | 15,33 | attention | 13 | 18,32 | year | 4 | 11,08 |
| weapon | 3 | 15,33 | look | 13 | 18,32 | way | 4 | 11,08 |
| lesson | 3 | 15,33 | phone | 12 | 16,91 | money | 4 | 11,08 |
| baby | 3 | 15,33 | step | 10 | 14,09 | bath | 3 | 8,31 |
| notice | 2 | 10,22 | decision | 9 | 12,68 | exam | 3 | 8,31 |
| hostage | 2 | 10,22 | action | 8 | 11,27 | class | 3 | 8,31 |
| prevention | 2 | 10,22 | life | 8 | 11,27 | hour | 3 | 8,31 |
| advice | 2 | 10,22 | animal | 7 | 9,86 | position | 3 | 8,31 |
| action | 2 | 10,22 | control | 6 | 8,45 | control | 3 | 8,31 |
| exam | 2 | 10,22 | course | 6 | 8,45 | everything | 3 | 8,31 |
| year | 2 | 10,22 | money | 5 | 7,04 | view | 3 | 8,31 |
| job | 2 | 10,22 | everything | 5 | 7,04 | drug | 3 | 8,31 |
| decision | 1 | 5,11 | drug | 5 | 7,04 | job | 3 | 8,31 |
| refuge | 1 | 5,11 | year | 3 | 4,23 | effect | 3 | 8,31 |
| control | 1 | 5,11 | way | 3 | 4,23 | woman | 3 | 8,31 |
| everything | 1 | 5,11 | effect | 3 | 4,23 | | | |
| | | | notice | 1 | 1,41 | | | |
| | | | position | 1 | 1,41 | | | |
| | | | view | 1 | 1,41 | | | |

Having scrutinized the give+noun collocations, now it is the turn for examining take+noun collocations used in the three corpora. When take+noun collocations are dealt in detail, it seems possible to infer that “take responsibility” is the only combination which is commonly used among the first three most frequently used take+noun collocations in the three corpora. The other two take+noun collocations are different from each other. TUCLE, which is the smallest corpus, holds 10 raw references of “take responsibility” with the normalized frequency of 51,10 whereas KTUCLE and LOCNESS do not include the same collocation with similar reference and frequencies. Although KTUCLE consists of the aforesaid collocation with more raw references, that is 30, the normalized frequency occurs to be 42,30 which is the result of the corpus size.

Additionally, the native corpus involves a very similar normalized frequency with 41,55 in spite of the fewer number of raw references of 15.

One remarkable difference between the two non-native corpora is that KTUCLE participants applied more collocations in number than TUCLE. When KTUCLE holds its first most used collocation with 30 raw references, TUCLE includes only 11 of it. Similarly, the second most frequently applied take+noun collocations in KTUCLE and TUCLE contain raw references of 26 and 10 respectively. For further detail, it looks possible to state that the third most used take+noun collocation in TUCLE has raw references of 8 whereas KTUCLE has 24 of it; yet, the normalized frequencies do not differ greatly since KTUCLE is the largest corpus in this study while TUCLE is the smallest. The third most used take+noun collocation in TUCLE is “take photo” with the normalized frequency of 40,88 while KTUCLE’s third most used collocation is “take education” with the normalized frequency of 33,81.

A noteworthy situation arises here because although “take education” is frequently used in KTUCLE, it is an evidence of misuse since BNC does not hold even one example of it. The reason for this probably derives from the native language interference which is also termed L1 transfer. This may lead to the inference that the third most used take+noun collocation in KTUCLE is an example of misuse.

The native corpus and TUCLE hold a common feature; the raw references of the take+noun collocations are similar, yet the normalized frequencies differ. As an example, “take advantage” and “take precaution” are used 5 times in the two corpora; however, the normalized frequency of “take advantage” in TUCLE is 25,55 whereas “take precaution” has the normalized frequency of 13,85 in LOCNESS.

LOCNESS includes 4 take+noun collocations whose evidences are not encountered in the two non-native corpora; “take egg”, “take bath”, “take class” and “take woman” has examples in LOCNESS, yet, they are not evident in neither TUCLE nor KTUCLE. Additionally, there are 8 take+noun collocations which are not existent in TUCLE; they are “take course”, “take way”, “take money”, “take hour”, “take position”, “take view”, “take drug” and “take effect”. In addition to these, there are 2 take+noun collocations that have no evidence in KTUCLE; “take refuge” and “take job” are the inexistent collocations in KTUCLE.

The least frequently used take+noun collocations in the three corpora differ; TUCLE involves “take decision”, “take refuge”, “take control” and “take everything” as its take+noun collocations that are only referenced once. “Take notice”, “take position” and “take view” are the take+noun collocations in KTUCLE which have 1 raw references for each. Finally and noticeably, LOCNESS consists of 12 take+noun collocations which have 3 raw references for each of them; “take bath”,

“take exam”, “take class”, “take hour”, “take position”, “take control”, “take everything”, “take view”, “take drug”, “take job”, “take effect” and “take woman” all have 3 references and the normalized frequencies of 8,31.

Table 21: Log-likelihood Ratio and Significance Level with “take” in TUCLE and LOCNESS

| TAKE – TUCLE vs LOCNESS - LL SCORE | | | | | | |
|------------------------------------|------------------|--------------|--------------------|--------------|------|-----|
| OBJECTS | TUCLE 195,684 | | LOCNESS 361,054 | | LL | + - |
| | RAW | NORM | RAW | NORM | | |
| precaution | 11 | 56,21 | 5 | 13,85 | 7,46 | + |
| decision | 1 | 5,11 | 6 | 16,62 | 1,55 | |
| action | 2 | 10,22 | 8 | 22,16 | 1,10 | |
| refuge | 1 | 5,11 | 5 | 13,85 | 1,02 | |
| advantage | 5 | 25,55 | 13 | 36,01 | 0,45 | |
| responsibility | 10 | 51,10 | 15 | 41,55 | 0,25 | |
| control | 1 | 5,11 | 3 | 8,31 | 0,19 | |
| everything | 1 | 5,11 | 3 | 8,31 | 0,19 | |
| exam | 2 | 10,22 | 3 | 8,31 | 0,05 | |
| job | 2 | 10,22 | 3 | 8,31 | 0,05 | |
| time | 5 | 25,55 | 10 | 27,70 | 0,02 | |
| risk | 3 | 15,33 | 6 | 16,62 | 0,01 | |
| notice | 2 | 10,22 | 4 | 11,08 | 0,01 | |
| year | 2 | 10,22 | 4 | 11,08 | 0,01 | |
| life | 8 | 40,88 | 15 | 41,55 | 0 | |

When TUCLE and LOCNESS are compared in terms of their take+noun collocations, it stands out that there is only one combination which is overused in TUCLE; “take precaution” is used 11 times in TUCLE whereas it has the raw reference of 5 in LOCNESS even though LOCNESS is a larger corpus in size. The normalized frequency of “take precaution” is 56,21 in TUCLE and 13,85 in LOCNESS; therefore, the LL measure of the aforesaid collocation occurs to be +7,46 which makes it an overused take+noun collocation in TUCLE. On the contrary, there do not seem any evidences of underuse of take+noun collocations in TUCLE.

Table 22: Log-likelihood Ratio and Significance Level with “take” in KTUCLE and LOCNESS

| TAKE – KTUCLE vs LOCNESS - LL SCORE | | | | | | |
|-------------------------------------|-------------------|-------|--------------------|-------|------|-----|
| | KTUCLE 709,748 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| life | 8 | 11,27 | 15 | 41,55 | 9,47 | - |
| exam | 20 | 28,18 | 3 | 8,31 | 5,16 | + |
| notice | 1 | 1,41 | 4 | 11,08 | 4,52 | - |
| precaution | 22 | 31 | 5 | 13,85 | 3,09 | |
| control | 1 | 5,11 | 3 | 8,31 | 2,85 | |
| everything | 1 | 5,11 | 3 | 8,31 | 2,85 | |
| action | 8 | 11,27 | 8 | 22,16 | 1,79 | |
| hour | 13 | 18,32 | 3 | 8,31 | 1,77 | |
| year | 3 | 4,23 | 4 | 11,08 | 1,60 | |
| way | 3 | 4,23 | 4 | 11,08 | 1,60 | |
| job | 2 | 10,22 | 3 | 8,31 | 1,44 | |
| course | 6 | 8,45 | 6 | 16,62 | 1,35 | |
| advantage | 17 | 23,95 | 13 | 36,01 | 1,19 | |
| risk | 17 | 23,95 | 6 | 16,62 | 0,63 | |
| time | 26 | 36,63 | 10 | 27,70 | 0,59 | |
| money | 5 | 7,04 | 4 | 11,08 | 0,44 | |
| decision | 9 | 12,68 | 6 | 16,62 | 0,26 | |
| responsibility | 30 | 42,30 | 15 | 41,55 | 0 | |

Having observed the 1 overused take+noun collocation in TUCLE, it is also possible to infer that KTUCLE holds 1 overused take+noun collocation as well. However, this time the overused take+noun collocation in KTUCLE is “take exam” with 20 raw references and the normalized frequency of 28,18 whereas the aforementioned collocation is used 3 times in LOCNESS with the normalized frequency of 8,31 which make the collocation overused with the LL measure of +5,16.

As distinct from TUCLE, KTUCLE holds two underused take+noun collocations when compared to LOCNESS. “Take life” and “take notice” are used in KTUCLE with 8 and 1 raw references respectively, also their normalized frequencies are 11,27 and 1,41. The largest corpus with 709,748 words includes only one example of “take notice” and this feature of the aforesaid collocation makes it underused when compared to LOCNESS as the LL measure occurs to be -4,52. In addition to “take notice”, the other evidence of underuse is “take life” since it has 8 raw references in KTUCLE while it is included in LOCNESS for 15 times. Additionally, the normalized frequencies of the mentioned collocation are 11,27 in KTUCLE and 41,55 in LOCNESS; therefore, this situation creates an underuse condition with the LL measure of -9,47.

Table 23: Distribution of “get” in TUCLE, KTUCLE and LOCNESS

| GET | | | | | | | | |
|------------------|-----|--------------|-------------------|-----|--------------|--------------------|-----|--------------|
| TUCLE 195,684 | | | KTUCLE 709,748 | | | LOCNESS 361,054 | | |
| OBJECTS | RAW | NORM | OBJECTS | RAW | NORM | OBJECTS | RAW | NORM |
| job | 16 | 81,76 | information | 61 | 85,95 | money | 6 | 16,62 |
| information | 10 | 51,10 | result | 15 | 21,13 | education | 4 | 11,08 |
| education | 9 | 45,99 | cure | 12 | 16,91 | job | 4 | 11,08 |
| increase | 4 | 20,44 | marriage | 12 | 16,91 | respect | 3 | 8,31 |
| license | 4 | 20,44 | wisdom | 9 | 12,68 | gun | 3 | 8,31 |
| result | 4 | 20,44 | money | 7 | 9,86 | information | 2 | 5,54 |
| phone | 3 | 15,33 | chance | 7 | 9,86 | idea | 2 | 5,54 |
| diploma | 2 | 10,22 | point | 7 | 9,86 | | | |
| treatment | 2 | 10,22 | divorce | 6 | 8,45 | | | |
| reaction | 2 | 10,22 | increase | 6 | 8,45 | | | |
| food | 2 | 10,22 | cancer | 6 | 8,45 | | | |
| training | 2 | 10,22 | something | 6 | 8,45 | | | |
| disadvantage | 2 | 10,22 | knowledge | 5 | 7,04 | | | |
| experience | 2 | 10,22 | life | 5 | 7,04 | | | |
| harm | 2 | 10,22 | training | 5 | 7,04 | | | |
| sentence | 2 | 10,22 | news | 5 | 7,04 | | | |
| idea | 2 | 10,22 | opportunity | 3 | 4,23 | | | |
| homework | 2 | 10,22 | time | 1 | 1,40 | | | |
| money | 1 | 5,11 | | | | | | |

When the three corpora are examined in terms of their get+noun collocation usages, it stands out that TUCLE holds three most used get+noun collocations one of which is existent in KTUCLE while two of which are included in LOCNESS. “Get job” is the first most frequently applied get+noun collocation in TUCLE, and the same collocation stands among the three most frequently used combinations with get+noun. The mentioned collocation holds 16 raw references and 81,76 of the normalized frequency in TUCLE; besides, the same combination comprises 4 raw references and the normalized frequency of 11,08 in LOCNESS. Although the references and frequencies differ, “get job” is the first most used get+noun collocation in TUCLE and the third most used get+noun collocation in LOCNESS.

Additionally, LOCNESS and TUCLE have one more get+noun collocation which remains in the first three most used collocations; that is “get education” which has the raw reference of 9 and the normalized frequency of 45,99 in TUCLE while having the raw reference of 4 and the normalized frequency of 11,08 in LOCNESS. TUCLE includes one get+noun collocation common in the first three with KTUCLE; “get information” that has 10 raw references in TUCLE and 61 in KTUCLE when the normalized frequencies are 51,10 and 85,95 correspondingly.

There are 13 least frequently used get+noun collocations in TUCLE, 11 of which have 2 raw references; “get phone” has 3 references while “get diploma”, “get treatment”, “get reaction”, “get

food”, “get training”, “get disadvantage”, “get experience”, “get harm”, “get sentence”, “get idea” and “get homework” carry 2 raw references and the normalized frequency of 10,22 for each. The get+noun collocation with 1 raw reference and the normalized frequency of 5,11 is “get money”.

KTUCLE includes 6 get+noun collocations as the least used ones with 5, 3, and 1 raw references; “get knowledge”, “get life”, “get training” and “get news” hold 5 usages while “get something” and “get time” have 6 and 1 raw references respectively. The least frequently applied get+noun collocations in LOCNESS are “get respect”, “get gun”, “get information” and “get idea” with the normalized frequencies of 8,31 and 5,54 correspondingly.

LOCNESS consists three get+noun collocations which are encountered only in TUCLE, “get education”, “get job” and “get idea” have the raw references of 4,4 and 2 correspondingly in LOCNESS while TUCLE holds the aforementioned collocations with raw references of 9, 16 and 2 respectively. One common significant feature of “get respect” and “get gun” are only existent in LOCNESS, which means that there are not any evidences of aforesaid collocations in TUCLE and KTUCLE.

Similarly, the first three most used get+noun collocations in TUCLE do not carry any examples of themselves in KTUCLE; “get job”, “get information” and “get education” are existent both in TUCLE and LOCNESS, yet, KTUCLE does not hold any examples of them. In the same way, KTUCLE involves two combinations that cannot be seen in TUCLE but can be faced in LOCNESS; “get information” and “get money” are the evidences with the normalized frequencies of 85,95 and 9,86 respectively.

Table 24: Log-likelihood Ratio and Significance Level with “get” in TUCLE and LOCNESS

| GET – TUCLE vs LOCNESS - LL SCORE | | | | | | |
|-----------------------------------|------------------|--------------|--------------------|--------------|-------|-----|
| | TUCLE 195,684 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| job | 16 | 81,76 | 4 | 11,08 | 16,91 | + |
| information | 10 | 51,10 | 2 | 5,54 | 11,83 | + |
| education | 9 | 45,99 | 4 | 11,08 | 6,24 | + |
| money | 1 | 5,11 | 6 | 16,62 | 1,55 | |
| idea | 2 | 10,22 | 2 | 5,54 | 0,37 | |

TUCLE and LOCNESS consist of common get+noun collocations two of which stand in the first three most used combinations in both corpora. As mentioned in the interpretation of the previous table, “get job” and “get education” stand out with their feature of being overused in TUCLE. “Get job” has the raw reference of 16 and the normalized frequency of 81,76 in TUCLE while LOCNESS consists 4 raw references and the normalized frequency of 11,08 although it is a

larger corpus; therefore, the LL measure occurs to be +16,91 which lead the “get job” collocation to be overused in TUCLE.

Similarly, “get education” is the second combination which is overused when compared to LOCNESS since while TUCLE holds the normalized frequency of 45,99, LOCNESS involves 11,08 of the normalized frequency. Thus, the LL measure counts as +6,24 and therefore “get education” becomes an overused get+noun collocation in TUCLE.

Table 25: Log-likelihood Ratio and Significance Level with “get” in KTUCLE and LOCNESS

| GET – KTUCLE vs LOCNESS - LL SCORE | | | | | | |
|------------------------------------|-------------------|--------------|--------------------|--------------|-------|-----|
| | KTUCLE 709,748 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| information | 61 | 85,95 | 2 | 5,54 | 36,79 | + |
| money | 7 | 9,86 | 6 | 16,62 | 0,86 | |

LOCNESS and KTUCLE include the collocation “get information” with the raw references of 2 and 61, and the normalized frequencies of 5,54 and 85,95 respectively. The huge diversity in the normalized frequencies forms the LL measure as +36,79; therefore “get information” happens to be the only overused get+noun collocation in KTUCLE.

Table 26: Distribution of “go” in TUCLE, KTUCLE and LOCNESS

| GO | | | | | | | | |
|------------------|-----|--------------|-------------------|-----|--------------|--------------------|-----|--------------|
| TUCLE 195,684 | | | KTUCLE 709,748 | | | LOCNESS 361,054 | | |
| OBJECTS | RAW | NORM | OBJECTS | RAW | NORM | OBJECTS | RAW | NORM |
| university | 7 | 35,77 | university | 13 | 18,32 | mile | 4 | 11,08 |
| place | 5 | 25,55 | shopping | 10 | 14,09 | home | 3 | 8,31 |
| city | 4 | 20,44 | library | 9 | 12,68 | downtown | 2 | 5,54 |
| recruitment | 2 | 10,22 | school | 8 | 11,27 | way | 2 | 5,54 |
| school | 2 | 10,22 | home | 6 | 8,45 | | | |
| military service | 2 | 10,22 | city | 4 | 5,64 | | | |
| | | | way | 4 | 5,64 | | | |
| | | | mall | 3 | 4,23 | | | |
| | | | bank | 3 | 4,23 | | | |
| | | | hospital | 3 | 4,23 | | | |
| | | | place | 3 | 4,23 | | | |

The table of go+noun collocations seems different from the previous tables as LOCNESS has two go+noun collocations which are existent neither in TUCLE nor KTUCLE; “go mile” and “go downtown” have no evidences in the two non-native corpora although “go mile” is the most

frequently used go+noun combination in LOCNESS with 4 raw references and the normalized frequency of 11,08.

Besides, “go home” and “go way” are only encountered in KTUCLE; “go home” stands with the raw references of 6 in KTUCLE and 3 in LOCNESS. The normalized frequencies of the aforementioned collocation are 8,45 in KTUCLE while 8,31 in the native corpus. The collocation “go way” holds 4 raw references in KTUCLE while holding 2 in LOCNESS; likewise, the normalized frequencies of the aforesaid collocation are 5,64 and 5,54 respectively.

The common go+noun collocations between TUCLE and KTUCLE are “go university” in the first row. “Go place”, “go city” and “go school” are mutually used collocation in the two non-native corpora with their normalized frequencies of 25,55 in TUCLE and 4,23 in KTUCLE. The second common go+noun collocation “go city” holds the normalized frequency as 20,44 in TUCLE and 5,64 in KTUCLE. Despite the big difference in the normalized frequencies of “go place” and “go city”, “go school” does not hold a huge diversity between the two corpora mentioned. “Go school” is used 2 times in TUCLE while it is applied 8 times in KTUCLE, which pave the way for the normalized frequencies of 10,22 and 11,27 respectively.

TUCLE holds an example of misuse with the collocation “go recruitment” whose evidences are not existent in native corpora when checked for the acceptability. The reason for creating this misused collocation seems to be deriving from negative transfer from the mother tongue whose examples are encountered in previous tables as well.

Table 27: Log-likelihood Ratio and Significance Level with “go” in TUCLE and LOCNESS

| GO – TUCLE vs LOCNESS - LL SCORE | | | | | | |
|----------------------------------|------------------|------|--------------------|--------------|----|-----|
| | TUCLE 195,684 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| mile | - | | 4 | 11,08 | | |
| home | - | | 3 | 8,31 | | |
| downtown | - | | 2 | 5,54 | | |
| way | - | | 2 | 5,54 | | |

TUCLE does not hold any common evidences of go+noun collocations with LOCNESS; therefore, it is impossible to create a LL measure to compare the corpora.

Table 28: Log-likelihood Ratio and Significance Level with “go” in KTUCLE and LOCNESS

| GO – KTUCLE vs LOCNESS - LL SCORE | | | | | | |
|-----------------------------------|-------------------|-------------|--------------------|-------------|----|-----|
| | KTUCLE 709,748 | | LOCNESS 361,054 | | | |
| OBJECTS | RAW | NORM | RAW | NORM | LL | + - |
| home | 6 | 8,45 | 3 | 8,31 | 0 | |
| way | 4 | 5,64 | 2 | 5,54 | 0 | |

Although TUCLE and LOCNESS do not include any mutual go+noun collocations, KTUCLE holds two of them; “go home” and “go way” are existent with the raw references of 6 and 4 respectively. LOCNESS includes 3 of “go home” and 2 of “go way”; the normalized frequencies of “go home” are almost identical with the rates of 8,45 in TUCLE and 8,31 in LOCNESS. Similarly, “go way” has the normalized frequency of 5,64 in TUCLE and 5,54 in LOCNESS and therefore no LL measure difference can be withdrawn between the two corpora compared.

3.3. Discussion of the Research Questions

This study employed corpus-driven quantitative research by comparing three corpora two of which are non-native and compiled in Trabzon although they belong to different state universities. Corpus tool sketch engine was utilized with the purpose of extracting verb+noun collocations formed with the seven base verbs *do*, *make*, *have*, *give*, *take*, *get* and *go*. The verb+noun collocations were examined in the argumentative essay papers produced by native and non-native speakers of English. This last chapter of the research aims at revising the significant points and answering the research questions listed in the introduction chapter. In addition to answering the questions, this chapter holds the mission of suggesting further research and pedagogical implications. The research questions which pave the way for conducting this research are as follows:

1. How do native and non-native users differ in verb+noun collocation usages? Is there any statistically significant difference between natives and non-natives in terms of interlanguage developmental levels?
2. What are some of the most common overused and underused and misused patterns of verb+noun collocations in non-native corpora?
3. What are some of the most common “base verbs” that are used as complementary to in the verb+noun: collocations?

3.3.1. Discussion of the First Research Question

This present research study aimed at revealing the usage difference between native and non-natives in terms of collocational usage; therefore, it asked the questions ‘How do native and non-native users differ in verb+noun collocation usages? Is there any statistically significant difference between natives and non-natives in terms of interlanguage developmental levels?’

The first main question of the present study asked for the difference between native and non-native corpora in terms of their use of verb+noun collocations. Two non-native corpora included TUCLE and KTUCLE that both hold argumentative essays by tertiary level EFL students. Raw references and normalized frequencies of the verb+noun collocations applied by the participants were extracted with the help of sketch engine.

Hill (2000), Wray (2002) and Sadeghi (2009) have an argument in common which supports that if learners have a considerable amount of collocations in their works, they can be proficient and competent users of the target language they are trying to improve in. Additionally, this current study of mine worked on collocations as they are “fundamental units in the stylistic description of text” (Gledhill, 2000: 234)

The most significant result of this study seems to be the number of verb+noun collocations used in the three corpora; that is, native participants did not differ greatly in the use of verb+noun collocations in number from the two non-native corpora participants. The comparison of raw references and normalized frequencies implies that tertiary level learners of English use similar number of verb+noun collocations with LOCNESS participants. When all the verb+noun collocations formed with seven base verbs are scrutinized, it seems possible to infer that tertiary level EFL students and native speakers of English have alike features of verb+noun collocation numbers.

Table 29: The Total Number of Verb+noun Collocations in the Three Corpora

| Investigated Corpora | do+noun | make+noun | have+noun | give+noun | take+noun | get+noun | go+noun |
|----------------------|---------|-----------|-----------|-----------|-----------|----------|---------|
| TUCLE (195,684) | 17 | 17 | 69 | 24 | 28 | 19 | 6 |
| KTUCLE (709,748) | 22 | 26 | 70 | 27 | 31 | 18 | 11 |
| LOCNESS (361,054) | 12 | 26 | 70 | 20 | 37 | 7 | 4 |

Although the three corpora do not display a big difference in terms of collocation numbers in general, do+noun collocations carry an outstanding dissimilarity since the three corpora take

diverse number of collocations. However, when generally regarded, it is possible to infer that the rest of the verb+noun collocations do not contrast in terms of numbers.

KTUCLE and LOCNESS hold a quite similar feature in terms of make+noun collocations; although KTUCLE is much larger than LOCNESS in size, they include related number of make+noun collocations. However, one thing should be taken into consideration; some make+noun collocations are included with only one raw reference whereas some are involved with at least three raw references.

The most frequently used base verb in the three corpora is have with 69 different noun combinations in TUCLE, 70 combinations for each in both KTUCLE and LOCNESS. One other striking feature of have+noun collocations is that they have almost identically same numbers in the three corpora compared. Although TUCLE is the smallest corpus, it holds a very similar number of have+noun collocations both with KTUCLE and LOCNESS. Give+noun collocations take similar space as the do+noun and make+noun collocations with their alike number of combinations. Being the largest corpus, KTUCLE holds 27 different give+noun collocations while TUCLE and LOCNESS comprise 24 and 20 combinations respectively.

Similar to the abovementioned verb+noun combinations, take+noun collocations take more place in KTUCLE again. TUCLE and LOCNESS hold very similar numbers with 28 and 27 combinations respectively. The get+noun collocations reveal a difference; although till now KTUCLE included the most number of combinations, this time TUCLE held more collocations in number than the other two corpora. Additionally, LOCNESS involved much fewer get+noun collocations than the two non-native corpora. The base verb “go” holds the fewest combinations in the three corpora compared. The native corpus comprises only four go+noun collocations that is similar to TUCLE; however, KTUCLE involves 11 different combinations of go.

When all the verb+noun collocations are examined, it seems possible to infer that *have* is the most frequently used base verb in terms of verb+noun collocations. It takes place in TUCLE with 69 noun combinations while KTUCLE and LOCNESS both include 70 noun combinations with it. However, *go* is the least frequently applied base verb which is existent in TUCLE with 6 and in LOCNESS with 4 combinations although KTUCLE includes 11 noun combinations with it.

Under the light of the scrutiny of the verb+noun collocations in the three corpora, it occurs that non-native speakers of English reveal a significant difference in terms of interlanguage levels. Additionally, non-natives are not fully competent users of English thus they generally use fewer verb+noun combinations than natives however, they may use more combinations with some base verbs. Yet, this does not lead to the inference that non-natives apply as many combinations as natives since the collocations with 3 references at least were included in native corpus table.

The noteworthy feature of the collocation tables is that the two non-native corpora do not greatly differ in numbers of combinations from each other. However, the scrutiny of the collocation tables may reveal that the two non-native corpora participants do not have the ability to use almost as many collocations as the native corpus participants. In this manner, the general argument of Biskup (1992), Kaszubski (2000), Howarth (1996), which implies that even advanced level learners of English may not employ as many collocations as natives, can be supported; because this study reveals that non-natives, whose levels are intermediate to advanced, practice the verb+noun collocations with alike numbers; however, non-natives still lack the number of collocations natives hold.

3.3.2. Discussion of the Second Research Question

The second research question of this present study asks for the overuse, underuse and misuse of collocations in the two non-native corpora. The question is as follows: What are some of the most common overused, underused and misused patterns of verb+noun collocations in non-native corpora?

3.3.2.1. Overuse Evidences of Verb+noun Collocations in TUCLE and KTUCLE

3.3.2.1.1. Overuse Evidences of Verb+noun Collocations in TUCLE

As Selinker (1972: 214) asserts, non-natives may not be able to employ fully correct target language usage since they have the attempt of producing a target language work. Therefore, the results of the collocational usage difference may answer why natives and non-natives are dissimilar in their uses.

TUCLE and KTUCLE, whose participants are non-native tertiary level learners of English, include significant verb+noun collocations that are overused and underused when compared to the native corpus LOCNESS. “Do something” can be provided for a clear evidence of overuse in a non-native corpus; TUCLE participants apply the aforementioned collocation with the LL measure of +32,90. Additionally, “do thing”, “do everything”, “do anything”, “do research” and “do work” collocations are overused in TUCLE.

In TUCLE, there are some have+noun collocations that are overused as well; “have right” is the most frequently overused have+noun combination with the LL measure of +47,24 when compared to the native corpus. 13 more have+noun collocations are overused in TUCLE that are “have abortion”, “have knowledge”, “have advantage”, “have difficulty”, “have baby”, “have feeling”, “have idea”, “have job”, “have place”, “have experience”, “have opportunity”, “have opinion” and “have life”.

Table 30: The Overused Verb+noun Collocations in TUCLE

| | | | |
|---------------|-----------------|------------------|-----------------|
| do something | have experience | have life | have job |
| do thing | have right | have place | get education |
| do everything | have abortion | have idea | get job |
| do anything | have knowledge | have feeling | take precaution |
| do research | have advantage | have opportunity | give example |
| do work | have difficulty | have opinion | give birth |
| have baby | | | |

Tono (2002) argues that non-native learners of English apply some certain patterns more often than the others because they face those in their textbooks or they are directly taught in class. As an evidence; although “have abortion” is an example of misuse, it is overused in the essays of TUCLE participants since the researcher assigned them an argumentative essay about their opinions on *abortion*.

“Give birth” and “give example” are two examples of give+noun collocations that are overused in TUCLE. “Give example” holds the log likelihood measure of +23,66 when compared to the native corpus users while “give birth” involves usage with the LL measure +16,68. In addition to the overused collocations above, “take precaution” is another combination that is overused with the LL measure +7,46 when compared to LOCNESS. Get+noun collocations include examples of overuse in TUCLE; “get job” “get education” are the two other combinations that are overused by TUCLE participants with the LL measures of +16,91 and +6,24 respectively.

3.3.2.1.2. Overuse Evidences of Verb+noun Collocations in KTUCLE

When it comes to the examination of KTUCLE in terms of overuse, it is possible to state that “do work” is the most frequently used do+noun collocation and it is also one of the overused collocations in KTUCLE with the LL measure of +22,99. In addition to “do work”, the collocations “do research”, “do everything”, “do something” and “do anything” are the overused combinations in KTUCLE when compared to LOCNESS.

As for the have+noun collocations, KTUCLE holds 11 overused combinations; “have problem” is the most frequently overused have+noun collocation with the LL measure +35,38. The remaining overused have+noun combinations can be listed as “have life”, “have knowledge”, “have advantage”, “have responsibility”, “have place”, “have ability”, “have time”, “have difficulty”, “have job” and “have opportunity”.

While TUCLE holds the collocation “give example” with the LL measure of +23,66, KTUCLE participants overused the aforementioned collocation as well. However, the measure of KTUCLE seems significantly higher than TUCLE since KTUCLE holds the LL measure as +48,90 when compared to LOCNESS. In addition to the overused collocations, “take exam” is another evidence of overuse in KTUCLE with the LL measure of +5,16 when it is compared to the papers of native corpus participants. In addition to the overused combinations mentioned above, “get information” is an evidence of overuse in KTUCLE with the LL measure of +36,79 when compared to LOCNESS.

Table 31: The Overused Verb+noun Collocations in KTUCLE

| | | | |
|---------------|---------------------|-----------------|------------------|
| do work | have problem | have place | have opportunity |
| do research | have life | have ability | give example |
| do everything | have knowledge | have time | take exam |
| do something | have advantage | have difficulty | get information |
| do anything | have responsibility | have job | |

This overuse feature of collocations can overlap with the result of Leńko-Szymańska’s (2004) research which argues that some grammatical patterns are overused in the works of non-native learners of English in their English essays.

3.3.2.2. Underuse Evidences of verb+noun Collocations in TUCLE and KTUCLE

3.3.2.2.1. Underuse Evidences of verb+noun Collocations in TUCLE

There do not seem any examples of underuse in do+noun, give+noun, take+noun, get+noun and go+noun collocations in TUCLE. The corpus comprises of one example of underuse in terms of make+noun collocations; “make money” is the evidence of underuse in TUCLE with the LL measure of -6,19. “Have sex” and “have effect” are the two instances of underuse in have+noun collocations in TUCLE with their LL measures of -11,37 and -4,44 respectively. The reason lying behind this may be related to the negative transfer; the participants may not relate ‘have’ and ‘sex’ as used together since in Turkish they use another verb for that.

Table 32: The Underused Verb+noun Collocations in TUCLE

| |
|-------------|
| make money |
| have sex |
| have effect |

3.3.2.2.2. Underuse Evidences of verb+noun Collocations in KTUCLE

Table 33: The Underused Verb+noun Collocations in KTUCLE

| | | | |
|--------------|----------------|--------------|----------------|
| Do favor | Make decision | Make money | Make claim |
| Make profit | Make statement | Make choice | Make error |
| Make point | Have sex | Have child | Have choice |
| Have history | Have law | Have meaning | Have authority |
| Have aspect | Have power | Have rate | Have control |
| Give reason | Give power | Take life | Take notice |

While TUCLE does not hold any examples of underuse in terms of do+noun collocations, KTUCLE includes one which is “do favour”; it has the log likelihood measure of -4,14 when compared to the native corpus. Additionally, although TUCLE includes only one evidence of underuse in make+noun collocations, KTUCLE holds 8 of them. There stands a noteworthy situation here, in spite of the fact that “make decision” is the most frequently used make+noun collocation in KTUCLE, it still remains underused when compared to LOCNESS with the LL measure of -11,69. The other remaining 7 underused make+noun collocations are “make money”, “make claim”, “make profit”, “make statement”, “make choice”, “make error” and “make point”.

The second non-native corpus KTUCLE includes 11 examples of underuse of have+noun collocations whereas TUCLE holds only two of them. One mutually underused collocation between the two non-native corpora is “have sex”, yet, the others seem different and more in number in KTUCLE. Karadeniz Technical University Corpus of Learner English hold evidences of underuse as “have child”, “have choice”, “have history”, “have law”, “have authority”, “have aspect”, “have power”, “have rate”, “have meaning” and “have control”.

KTUCLE comprises of two underused give+noun collocations which are “give reason” and “give power” whose LL measures are -5,69 and -4,18 correspondingly. One noteworthy situation arises here; although TUCLE does not hold any evidences of underuse in give+noun and take+noun collocations, KTUCLE does include examples. “Take life” and “take notice” are the underused take+noun collocations in KTUCLE with their LL measures of -9,47 and -4,52 respectively. Besides, no underused examples of get+noun and go+noun collocations can be encountered in KTUCLE.

3.3.2.3. Misuse Evidences of Verb+noun Collocations in TUCLE and KTUCLE

In addition to the overused and underused patterns of verb+noun collocations in the two non-native corpora, there occurred 7 examples of misuses after checking them in the BNC and Oxford Collocation Dictionary. 4 of the misused patterns belonged to TUCLE; “do rest”, “do abortion”,

“give phone” and “go recruitment” are the evidences of misuses which sound too Turkish or which are directly effects of L1 transfer.

Table 34: The Misused Verb+noun Collocations in the Non-native Corpora

| TUCLE | KTUCLE |
|----------------|----------------|
| do rest | do literature |
| do abortion | have boy |
| give phone | take education |
| go recruitment | |

In addition, the remaining 3 misuse evidences are of KTUCLE which are “do literature”, “have boy” and “take education”. This outcome of the results overlaps with the ones of Aarts and Granger (1998b) and Borin and Prütz (2004) since their research revealed that their participants applied L1 transfer and made mistakes as well; thus, this seems to be the proof of interlanguage process which is the evidence of learning.

3.3.3. Discussion of the Third Research Question

The third research question of the present study asked for the most common base verbs which are used as complementary in the verb+noun collocations; ‘What are some of the most common “base verbs” that are used as complementary to in the verb+noun: collocations?’

During the analysis of the sketch engine results, one verb overwhelmingly took much more room than the other verbs; *have*. Both native and non-native corpora included the most frequently used base verb as “*have*”. KTUCLE and LOCNESS included 70 examples of *have*+noun collocations while TUCLE included 69 of them. The second most commonly used base verb is “*take*” which is utilized in 28 combinations in TUCLE, 31 in KTUCLE and 28 in LOCNESS. As for the least commonly used base verb, “*go*” can be given as an evidence since TUCLE holds 6 combinations whereas KTUCLE includes 11 and LOCNESS consists of 4 combinations with the aforesaid base verb.

CONCLUSION AND SUGGESTIONS

The main purpose of this present research is to reveal the differences and similarities among the three corpora (native and non-native) in terms of collocational usage. With this aim, the researcher compiled an argumentative corpus with a strict design criteria and that is composed of 344 argumentative essays that were written by tertiary level Turkish EFL learners of English Language Teaching Department at Trabzon University. The corpus is named as TUCLE and it currently holds 195,684 words. Thus, a native corpus and two non-native corpora were compared and contrasted for collocational patterns.

The second non-native corpus belonged to a state university in Trabzon as well; KTUCLE included 600 texts and 709,748 words which make it the largest corpus of the three corpora compared. As for the native corpus, LOCNESS took place in the study with 322 texts and 361,054 words. The three corpora were investigated by revealing the verb+noun collocations with the base verbs *do*, *make*, *have*, *give*, *take*, *get* and *go* in all corpora. Following the display of the raw references and the normalized frequencies of the collocations formed with the aforesaid base verbs, non-native corpora and the native corpus were compared by revealing the log-likelihood measure. The comparisons were done in comparing one non-native and the native corpus; in other words, for instance, TUCLE and LOCNESS were compared in terms of analyzing the verb+noun collocations.

This research study may have the potential of revealing the overuses, underuses and misuses of verb+noun collocations by non-native participants as it embarked on displaying the difference in collocational usages of natives and non-natives. The results offered some misuses as well as the underuses and overuse patterns of collocations by non-native participants. These findings have potential to pave the way for reconsidering the teaching methods of collocations, if any, by the teachers, course book writers and the other involved parties such as course syllabus experts and other educational authorities within the country. In addition to the authority consideration on teaching methods; I, as a teacher, feel the responsibility and the need to guide my students for avoiding L1 transfer since it may have potential to bring collocational misuses based on the findings of the current study. For this purpose, the students may be encouraged to raise their perceptions and awareness levels to see the differences between target language and the mother tongue in an attempt to be proficient and competent users of English.

The results revealed that Turkish EFL learners overuse some verb+noun collocations; however, TUCLE and KTUCLE participants seem to show some overuse patterns with a limited

number of words when compared to LOCNESS. TUCLE holds 25 overused verb+noun collocations six of which are formed with “do”; and fourteen are formed with “have”. When it comes to the collocational overuses in KTUCLE, it seems possible to argue that there are 19 overused verb+noun collocations eleven of which are created with “have”, and five are formed with “do”. There occurs a noteworthy situation that puts forward the difference in the two non-native corpora. While TUCLE holds 25 overused collocations, KTUCLE includes 19 of them. Additionally, both corpora exclude any overuses of make+noun; in other words, there are no evidences of overuses in make+noun collocations in both non-native corpora.

The reason of overuse of collocations can be based on the incompetency in the target language; in other words, participants stick to a limited number of collocations and apply them over and over again in their works. They may hesitate to take risks of forming form newer collocations; thus, their essays include some fixed collocations more than the essays of the native participants. The lack of new collocations may be considered as a signal for the need of integrating teaching collocational patterns into the curricula. Following the analysis of overused verb+noun collocations in both non-native corpora; the underused ones may be displayed. TUCLE includes 3 underused verb+noun collocations two of which are mutual with KTUCLE participants.

This study mainly focused on the overuses and underuses of collocations by non-native participants. Nevertheless, there occurred 7 evidences of misuse during the analysis of the essays; TUCLE included four of them while KTUCLE held 3 of them. The misuses can be considered as the results of L1 transfer which is a sign of interlanguage process.

The results of this research imply the clarity of the need for integrating collocational patterns into teaching since they reveal that non-native tertiary level EFL students are not as competent and proficient as the natives in writing process. Non-natives may have the tendency to overuse and underuse some collocations, and additionally, they sometimes may apply misused patterns which are results of direct translation or transfer from the mother tongue. With all these in mind, curriculum developers and teachers as well may feel the need to consider modifying already existing programs with the purpose of raising more lexically competent and proficient users of English.

Pedagogical Implications for EFL Students

The study carries some results which may pave the way for amending pedagogical methods as tertiary level EFL students are incompetent and improficient in integrating fully-correct collocations into their writing. The learners of English may be provided with multi-word combination lists which are formed with collocations of verb+noun, adjective+noun, adverb+adjective, verb+adverb, etc. The students may also be supplied with pre-fabricated chunks

and lexical items in their reading passages or speaking classes in order to create fluency and native-like competency in their works. For these aims, the curricula may be modified and may include collocations or multi-word combinations as chapters or units with the purpose of avoiding negative transfers from mother tongue.

Limitations of the Study

This study bears some limitations in itself as it is limited to writing; in other words, the study could not inspect the speaking or reading performances of tertiary level EFL students. Additionally, the research is limited to the state universities in Trabzon; therefore, the generalization for the competencies of Turkish university students may not be possible.

In addition, the study did not focus on any other grammatical patterns apart from verb+noun collocations since it would take too much time and effort which might let the study be incomplete and insufficient in terms of validity. Focusing and investigation on only one pattern of collocations is a limitation of this research.

Suggestions for Further Research

This present research embarked on verb+noun collocations; therefore, focusing on the other collocational patterns would be a wise step. New research may be conducted for analyzing other grammatical aspects such as usage of passives, conditionals, reported speech, etc. Additionally, a similar study can be employed following the completion of the students' university education; the participants may be asked to write argumentative essays about the topics they were assigned in their preparatory class, and the essays might be analyzed in the same way they were done in this study. This could let the research be a longitudinal study.

REFERENCES

- Aarts, Jan and Granger, Sylviane (1998), "Tag Sequences in Learner Corpora: A Key to Interlanguage Grammar and Discourse", Sylviane Granger (Ed.), **Learner English on Computer**, *in* (132-141), Longman: Essex.
- Ackermann, Kirsten and Chen, Yu-Hua (2013), "Developing the Academic Collocation List (ACL) – A Corpusdriven and Expert-judged Approach", **Journal of English for Academic Purposes**, 12, 235–247.
- Akimoto, Minoji (1989), **A Study of Verbo-nominal Structures in English**, Shinozaki Shorin, Tokyo.
- Al Ghazali, Fawzi (2006), "Collocations and Word-Combinations in English: Considerations, Classifications, and Pedagogic Implications", **The Centre for English Language Studies (CELS) TESOL Arabia**, 1-20.
- Allan, Rachel (2009), "Can a Graded Reader Corpus Provide 'Authentic' Input?", **ELT Journal** Volume 63(1), 23-32, doi:10.1093/elt/ccn011.
- Allison, Desmond et al. (1998), "Dissertation Writing in Action: The Development of a Dissertation Writing Support Program for ESL Graduate Research Students", **English for Specific Purposes**, 17(2), 199-217.
- Altenberg, Bengt and Granger, Sylviane (2001), "The Grammatical and Lexical Patterning of Make in Native and Non-native Student Writing", **Applied Linguistics**, 22, 173–194.
- _____ (2002), "Recent Trends in Cross-linguistic Lexical Studies", Bengt Altenberg et al. (Eds.), **Lexis in Contrast. Corpus-based Approaches. Studies in Corpus Linguistics**, *in* (3-48), Benjamins, Amsterdam & Philadelphia.
- Asikin, Nida Amalia (2017), "The Analysis of Interlanguage Produced by 3rd Grade High School Students in Narrative Writing Text", **Indonesian EFL Journal**, 3(1), 45-50.
- Atkins, Sue et al. (1992), "Corpus Design Criteria", **Literary and Linguistic Computing**, 7(1), 1-16.
- Bahns, Jens and Eldaw, Moira (1993), "Should We Teach EFL Students Collocations?" **System**, 21(1), 101–114.
- Bailey, Stephen (2011), **Academic Writing, A Handbook for International Students**, (3rd edition), Routledge, Oxon.

- Bardovi-Harlig, Kathleen and Bofman, Theodora (1989), "Attainment of Syntactic and Morphological Accuracy by Advanced Language Learners", **Studies in Second Language Acquisition**, 11, 17-34.
- Barfield, Andrew William (2006), **An Exploration of Second Language Collocation Knowledge and Development**. Unpublished Doctoral Dissertation, University of Swansea, the United Kingdom.
- Barnbrook, Geoffrey (1996), **Language and Computers: A Practical Introduction to the Computer Analysis of Language**, Edinburgh University Press, Edinburgh.
- Benson, Morton, et al. (1986), **The BBI Combinatory Dictionary of English: A Guide to Word Combinations**, John Benjamins, Amsterdam/Philadelphia.
- Biber, Douglas (1993), "Representativeness in Corpus Design", **Literary and Linguistic Computing**, 8(4), 243-257.
- Biber, Douglas et al. (1998), **Corpus Linguistics: Investigating Language Structure and Use**, Cambridge University Press, Cambridge
- _____ (2004), "If You Look at ...:Lexical Bundles in University Teaching and Textbooks", **Applied Linguistics**, 25(3), 371–405.
- Biskup, Danuta (1992), "L1 Influence on Learners' Renderings of English Collocations. A Polish/German Empirical Study" Pierre J. L. Arnaud and Henri Béjoint (Eds.), **Vocabulary and Applied Linguistics in** (85–93), Macmillan, London.
- Bloomfield, Leonard (1933), **Language**, The University of Chicago Press, Chicago.
- Borin, Lars and Prütz, Klas (2004), "New Wine in Old Skins? A Corpus Investigation of L1 Syntactic Transfer in Learner Language", Guy Aston et al. (Eds.) **Corpora and Language Learners**, in (67-87), John Benjamins, Amsterdam/Philadelphia.
- Brown, Dorothy F. (1974), "Advanced Vocabulary Teaching: The Problem of Collocation", **RECL Journal**, 5:111.
- Campoy-Cubillo, Mari et al. (2010), "Introduction to Corpus Linguistics and ELT", Mari Carmen et al. (Eds.), **Corpus-based Approaches to English Language Teaching. Research in Corpus and Discourse**, in (3-17), Continuum International, London.
- Can, Cem (2009), "İkinci Dil Edinimi Çalışmalarında Bilgisayar Destekli Bir Türk Öğrenci İngilizcesi Derlemi: İcle'nin Bir Altderlemi Olarak Ticle", **Dil Dergisi**, 144, Nisan-Mayıs-Haziran, np.
- Chan, Alice Y.W. (2010), "Toward a Taxonomy of Written Errors: Investigation into the Written Errors of Hong Kong Cantonese ESL learners", **TESOL Quarterly**, 44(2), 295-319.
- Chomsky, Noam (1965), **Aspects of the Theory of Syntax**, MA: MIT Press, Boston.

- Cobb, Tom (2003), "Analyzing Late Interlanguage with Learner Corpora: Québec Replications of Three European Studies", **The Canadian Modern Language Review**, 59(3), 393-423.
- Collins Cobuild English Dictionary (1995), Harper Collins Publishers. London.
- Conrad, Susan (2000), "Will Corpus Linguistics Revolutionize Grammar Teaching in the 21st Century?", **TESOL Quarterly**, 3 (34), 548-560.
- Corder, Stephen PIT. (1971), "Idiosyncratic Dialects and Error Analysis", **International Review of Applied Linguistics**, Vol 9, 147-159.
- _____ (1982), **Error Analysis and Interlanguage**, Oxford University Press, Great Britain.
- Coşkun, Eyyup and Tiryaki, Esra Nur (2011), "Argumentative Text Structure and Teaching It", **Mustafa Kemal University Journal of Social Sciences Institute**, 8(16), 63-73.
- Cowie, Anthony Paul (1998), "Past Achievements and Current Trends", Anthony Paul Cowie (Ed.), **Phraseology: Theory, Analysis, and Applications**, np, Oxford University Press, Oxford.
- _____ (1999), **English Dictionaries for Foreign Learners: A History**, Oxford University Press, the United States.
- Coxhead, Averil (2000), "A New Academic Word List", **TESOL Quarterly**, 34(2), 213-238.
- Demir, Cüneyt (2016), **Lexical Proficiency of Collocation, Boosting, and Hedging in Academic Discourse: A Comparative Study**, Unpublished Coctoral Dissertation, Atatürk University.
- _____ (2018), "Word Combinations of English in Academic Writing", **Journal of Language and Linguistic Studies**, 14(1), 293-327.
- Durrant, Philip and Schmitt, Norbert (2009), "To What Extent Do Native and Non-native Writers Make Use of Collocations?", **IRAL**, 47,157-177, doi: 10.1515/iral.2009.007.
- Erman, Britt (2009), "Formulaic Language from a Learner Perspective: What the Learner Needs to Know", B. Corrigan et al. (Eds.), **Formulaic Language**, in (27–50), John Benjamins, Amsterdam.
- Erman, Britt and Warren, Beatrice (2000), "The Idiom Principle and the Open Choice Principle", **Text**, 20(1), 29-62.
- Farghal, Mohammed and Hussein, Obiedat (1995), "Collocations: A Neglected Variable in EFL", **International Review of Applied Linguistics**, 33, 315–331.
- Fauziati, Endang (2003), "Interlanguage Errors in English Textbooks for Junior High School Students in Surakarta", **Teflin Journal**, 14(2), 179-192.

- Flowerdew, Lynne (2004), "The Argument for Using English Specialized Corpora", Ulla Connor et al. (Eds.), **Discourse in the Professions: Perspectives from Corpus Linguistics**, in (11-33), John Benjamins, Amsterdam/Philadelphia.
- Fontenelle, Thierry (1998), "Phraseology and the Dictionary. Discovering Significant Lexical Functions in Dictionary Entries", Anthony Paul Cowie (Ed.), **Phraseology: Theory, Analysis, and Applications**, np, Oxford University Press, Oxford.
- Gabrielatos, Costas (2005), "Corpora and Language Teaching: Just a Fling or Wedding Bells?", **TESL-EJ**, 8(4), 1-35.
- Garcia, Ana Llinares (2006), "A Pragmatic Analysis of Children's Interlanguage in EFL Preschool Contexts", **Intercultural Pragmatics**, 3(2), 171-193.
- Gilquin, Gaetanelle (2000/2001), "The Integrated Contrastive Model. Spicing up Your Data", **Languages in Contrast**, 3(1), 95-123.
- Gledhill, Christopher (2000), "Collocations in Science Writing", **Language in Performance Series**, 22, 7-20.
- Granger, Sylviane (1998a), "The Computerized Learner Corpus: A Versatile New Source of Data for SLA Research", Sylviane Granger (Ed.), **Learner English on Computer**, in (3-18), Addison Wesley Longman, London and New York.
- _____ (1998b), "Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae", Anthony Paul Cowie (Ed.), **Phraseology: Theory, Analysis, and Applications**, np, Oxford University Press, Oxford.
- _____ (2002), "A bird's-eye View of Learner Corpus Research" in Sylviane Granger et al. (Eds.), **Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching**, in (3-33), John Benjamins, Amsterdam.
- _____ (2003), "The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research", **TESOL Quarterly**, 538-546.
- _____ (2009), "The Contribution of Learner Corpora to Second Language Acquisition and Foreign Language Teaching: A Critical Evaluation", Karin Aijmer (Ed.), **Corpora and Language Teaching**, in (13-32), Benjamins, Amsterdam & Philadelphia.
- Granger, Sylviane and Paquot, Magali (2008), "Disentangling the Phraseological Web", Sylviane Granger et al. (Eds.), **Phraseology, an interdisciplinary perspective**, in (27-49), John Benjamins Publishing Company, Amsterdam / Philadelphia.
- Granger, Sylviane and Tyson, Stephanie (1996), "Connector Usage in the English Essay Writing of Native and Non-native EFL Speakers of English", **World Englishes**, IS(I), 17-21.

- Granger, Sylviane et al. (2002), **Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching**, Benjamins, Amsterdam.
- Hamed, Muftah (2014), "Conjunctions in Argumentative Writing of Libyan Tertiary Students", **English Language Teaching**, 7(3), 108-120.
- Hill, Jimmie (2000), "Revising Priorities: From Grammatical Failure to Collocational Success", M. Lewis (Ed.), **Teaching Collocations: Further developments in the lexical approach**, in (47-69), Language Teaching Publications, Hove.
- Hinkel, Eli (2003), "Simplicity Without Elegance: Features of Sentences in L1 and L2 Academic Texts", **TESOL Quarterly**, 37(2), 275-301.
- _____ (2004), **Teaching Academic ESL Writing: Practical Techniques in Vocabulary and Grammar**, Lawrence Erlbaum Associates, Mahwah, NJ.
- Ho, Caroline Mei Lin (2003), "Empowering English Teachers to Grapple with Errors in Grammar", **The Internet TESL Journal**, 9(3), (http://iteslj.org/Techniques/Ho_Grammar_Errors.html, accessed 22nd January, 2019).
- Hosseini, Seyyed Sajjad, and Sangani, Hamid Rahmani (2015), "Studying the Pre-Intermediate Iranian EL Learners` Interlanguage and The Contribution of Their Innate System to The Development of Their Oral Communicative Proficiency", **Procedia - Social and Behavioral Sciences**, 192, 408-418.
- Howarth, Peter (1996), **Phraseology in English Academic Writing. Some Implications for Language Learning and Dictionary Making**, Niemeyer, Tübingen.
- _____ (1998), "Phraseology and Second Language Proficiency", **Applied Linguistics**, 19(1), 24-44.
- Hunston, Susan (2002), "Pattern Grammar, Language Teaching, and Linguistic Variation" Randi Reppen et al. (Eds.), **Using Corpora to Explore Linguistic Variation**, in (167-86), John Benjamins, Amsterdam.
- Kennedy, Graeme (1998), **An Introduction to Corpus Linguistics**, Longman, London and New York.
- Kilgarriff, Adam (2005), "Language is Never, Ever, Ever, Random", **Corpus Linguistics and Linguistic Theory**, 1(2), 263-276.
- Kilgarriff, Adam et al. (2004), "The Sketch Engine", **Proceedings of EURALEX, European Association for Lexicography**, July, 105-116.
- _____ (2014), "Finding Terms in Corpora for Many Languages With the Sketch Engine", **Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics**, April, 53-56.

- Kilimci, Abdurrahman (2001), "Automatic Extraction of the Lexical Profile of EFL Learners Through Corpus Query Techniques", **Journal of Faculty of Education**, Çukurova University, 37-47.
- Knoch, Ute et al. (2015), "What Happens to ESL Students' Writing After Three Years of Study at an English Medium University?", **Journal of Second Language Writing**, 28, 39-52.
- Lamidi, Mufutau Temitayo (2017), "The Syntax of Multi-Word Expressions in Yorulish Code-Mixing", **Ghana Journal of Linguistics**, 6(2), 30-55.
- Larsen-Freeman, Diane and Long, Michael H. (1991), **An Introduction to Second Language Acquisition Research**, Taylor & Francis Group.
- Laufer, Batia and Waldman, Tina (2011), "Verb-noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English", **Language Learning**, 61(2), 647-672.
- Leech, Geoffrey (1992), "Corpora and Theories of Linguistic Performance", Jan Svartvik (Ed.), **Directions in Corpus Linguistics**, in (105-122), Mouton de Gruyter, Berlin.
- Lenard, Dragana Božić et al. (2015), "A Corpus-based Linguistic Analysis of Errors in Final Paper Abstracts. The Second International Conference on Teaching English for Specific Purposes and New Language Learning Technologies", Faculty of Electronic Engineering, University of Niš, Serbia May, 22nd - 24th, 2015. No Page.
- Leńko-Szymańska, Agnieszka (2004), "Demonstratives as Anaphora Markers in Advanced Learners' English", In Guy Aston et al. (Eds.), **Corpora and Language Learners**, in (89-107), John Benjamins, Amsterdam/Philadelphia.
- Lennon, Paul (2008), "Contrastive Analysis, Error Analysis, Interlanguage", S. Gramley et al. (Eds.), **Bielefeld Introduction to Applied Linguistics**, in (51-60), Aisthesis, Bielefeld.
- Lewis, Michael (2000), **Teaching Collocation. Further Developments in the Lexical Approach**, LTP, Hove.
- Lin, Huifen and Fang, Yuehchiu (2009), "Virtual Dialogue: Native and Non-native English Speakers' Input on Measured Written Performance", **The International Journal of Learning**, 16(3), 451-463.
- Lopez Rodriguez, Jesus (2005), **Exploring Recurrent Word Combinations in A Business English Learner Corpus: A Parallel Corpus Analysis and Its Curricular Implication**. Doctoral Dissertation, University of Puerto Rico, UMI Number: 3198873.
- Luna, Rosa Muñoz (2010), "Interlanguage in Undergraduates' Academic English: Preliminary Results From Written Script Analysis", **Encuentro**, 19, 60-73, ISSN 1989-0796.
- Manning, Christopher, D and Schütze, Hinrich (1999), **Foundations of Statistical Natural Language Processing**, Massachusetts Institute of Technology.

- Manning, Christopher, et al. (2008), **Introduction to Information Retrieval**, Cambridge University Press, the USA.
- Martelli, Aurelia (2006), “A Corpus Based Description of English Lexical Collocations Used by Italian Advanced Learners”, **Atti del XII Congresso Internazionale di Lessicografia**, *in* (1005-1011), Torino.
- Martyńska, Małgorzata (2004), “Do English Language Learners Know Collocations?”, **Investigationes Linguisticae**, 11, 1-12.
- Masini, Francesca (2005), “Multi-word Expressions Between Syntax and the Lexicon: the Case of Italian Verb-particle Constructions”, **SKY Journal of Linguistics**, 18, 145–173.
- McArthur, Tom (1992), **The Oxford Companion to the English Language**, Oxford University Press, Oxford.
- McCarthy, Michael and Carter, Ronald (2004), “This That and the Other: Multi-word Clusters in Spoken English as Visible Patterns of Interaction”, **Teanga**, 21, 30-52.
- McCarthy, Michael and O'Dell, Felicity (2005), **English Collocations in Use**, Cambridge University Press, Cambridge.
- McEnery, Tony and Hardie, Andrew (2012), **Corpus Linguistics: Method, Theory and Practice**, Cambridge University Press, Cambridge.
- McEnery, Tony and Kifle, Nazareth A (2002), “Epistemic Modality in Argumentative Essays of Second-language Writers”, John Flowerdew (Ed.), **Academic Discourse**, *in* (182-195), Longman, Pearson Education, London.
- McEnery, Tony and Wilson, Andrew (2001), **Corpus Linguistics** 2nd Edition, Edinburgh University Press, Edinburgh.
- Meunier, Fanny (2002), “The Pedagogical Value of Native and Learner Corpora in EFL Grammar Teaching”, Sylviane Granger et al. (Eds.), **Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching**, *in* (119-142), John Benjamins, Amsterdam.
- Moon, Rosamund (1998a), **Fixed Expressions and Idioms in English: A Corpus-Based Approach**, New York.
- _____ (1998b), “Phraseology in Written and Spoken Corpora. Frequencies and Forms of Phrasal Lexemes in English”, Anthony Paul Cowie (Ed.), **Phraseology: Theory, Analysis, and Applications**, *in* (77-122), Oxford University Press, Oxford.
- Müller, Peter O. (2011), “Multi-word Expressions”, Peter O. Müller et al. (Eds.), **Word Formation, An International Handbook of the Languages of Europe** (HSK series), np, De Gruyter, Berlin.

- Narita, Masumi et al. (2004), **Connector Usage in the English Essay Writing of Japanese EFL Learners**, Paper presented at the Fourth International Conference on Language Resources and Evaluation.
- Nation, ISP (2001), **Learning Vocabulary in Another Language**, Cambridge University Press, Cambridge.
- Nattinger, James R. (1980), "A Lexical Phrase Grammar for ESL", **TESOL Quarterly**, 14(3), 337-344.
- Navés, Teresa (2006), **Analytical Measures of Learners' Written Interlanguage**, Paper presented at AESLA. XXV International AESLA Convention. Universidad de Murcia. Barcelona, Catalonia.
- Nemser, William (1971), "Approximative Systems of Foreign Language Learners", **International Review of Applied Linguistics**, 9, 115-124.
- Nesselhauf, Nadja (2005), **Collocations in a Learner Corpus**, John Benjamins B.V. Amsterdam · The Netherlands.
- Nicolás-Conesa, Florentina et al. (2014), "Development of EFL Students' Mental Models of Writing and Their Effects on Performance", **Journal of Second Language Writing**, 24, 1-19.
- Nizonkiza, Déogratias (2011), "The Relationship Between Lexical Competence, Collocational Competence, and Second Language Proficiency", **English Text Construction**, 4(1), 113-145.
- Oakes, Michael (1998), **Statistics for Corpus Linguistics**, Edinburgh University Press, Edinburgh.
- Özbay, Ali Şükrü (2015), **Corpus Analysis of the Support Verb Construction Development and Use by EFL Learners**, Unpublished Doctoral Dissertation, Karadeniz Technical University, Institute of Social Sciences.
- Özbay, Ali Şükrü and Kayaoğlu, Mustafa Naci (2016), "The Use of Support Verb Constructions (SVC) in the Argumentative Essays of the Turkish EFL learners: Ktucle vs Locness", Katarzyna Papaja et al. (Eds.), **Language in Focus: Exploring The Challenges and Opportunities in Linguistics**, in (215-237), Cambridge Scholars Publishing.
- Peterkehan (1998), **A Cognitive Approach to Language Learning**, Oxford University Press, Oxford.
- Popescu, Teodora (2007), "College Essay-Writing: A Corpus-Based Analysis", **Translation Studies: Retrospective and Prospective Views**, 182-188.

- Reynolds, Dudley W. (2005), "Linguistic Correlates of Second Language Literacy Development: Evidence from Middle-grade Learner Essays", **Journal of Second Language Writing**, 14, (19-45).
- Römer, Ute (2010), "Using General and Specialized Corpora in English Language Teaching: Past, Present and Future", Campoy-Cubillo et al. (Eds.), **Corpus-Based Approaches to English Language Teaching**, in (18-35), Continuum, London.
- Ruegg, Rachael (2010), "Interlanguage Development: The Effect of Unfocused Feedback on L2 Writing", **Kanda University of International Studies**, 247-254.
- Sadeghi, Karim (2009), "Collocational Differences Between L1 and L2: Implications for EFL Learners and Teachers", **TESL Canada Journal/Revue Tesl Du Canada**, 26(2), 100-124, DOI: <https://doi.org/10.18806/tesl.v26i2.417>.
- Şanal, Fahrettin (2007), **A Learner Corpus Based Study on Second Language Lexicology of Turkish Students of English**, Unpublished Doctoral Dissertation, Çukurova University.
- Schmitt, Norbert et al. (2004), "Are Corpus-derived Recurrent Clusters Psycholinguistically Valid?", Norbert Schmitt (Ed.), **Formulaic Sequences**, in (127-153), John Benjamins, Amsterdam.
- Seidlhofer, Barbara (2009), "Accommodation and the Idiom Principle in English as a Lingua Franca", **Intercultural Pragmatics**, 6(2), 195–215.
- Selinker, Larry (1969), "Language Transfer", **General Linguistics**, 9(2), 67-92.
- _____ (1970), "Second-Language Learning Experiments and Mathematical Learning Theory", **Language Learning**, 22(2), 291-299.
- _____ (1972), "Interlanguage", **International Review of Applied Linguistics**, 10, 209–231.
- _____ (1992), **Rediscovering Interlanguage**, Longman, London.
- Selinker, Larry and Lakshmanan, Usha (1992), "Language Transfer and Fossilization: the 'Multiple Effects Principle'", Susan et al. (Eds.), **Language Transfer in Language Learning**, 2nd Edition in (197-217), John Benjamins, Amsterdam.
- Selinker, Larry and Douglas, Dan (1985), "Wrestling with "Context" in Interlanguage Theory", **Applied Linguistics**, 6, 190-204.
- Sinclair, John (1990), **Collins COBUILD English Grammar**, Harper Collins, London.
- _____ (1991), **Corpus, Concordance, Collocation**, Oxford University Press, Oxford.
- _____ (2001), "Preface", Mohsen Ghadessy et al. (Eds.), **Small Corpus Studies and ELT**, in (vii–xv), John Benjamins, Amsterdam.

- _____ (2004), **Trust the Text**, Oxford University Press, Oxford.
- Sinclair, John and Renouf, Antoinette (1988), “A Lexical Syllabus for Language Learning” Ronald Carter et al. (Eds.), np, **Vocabulary and Language Teaching**, Taylor & Francis: London and New York.
- Siyanova, Anna and Schmitt, Norbert (2008), “L2 Learner Production and Processing of Collocation: A Multi-study Perspective”, **The Canadian Modern Language Review**, 64(3), 429-458.
- Skehan, Peter (2009), “Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis”, **Applied Linguistics**, 30, 510–532.
- Sprenger, Simone A. (2003), **Fixed Expressions and the Production of Idioms**, Doctoral Dissertation, The Netherlands, University of Groningen, ISBN: 90-9016943-1.
- Stubbs, Michael (1996), **Text and Corpus Analysis: Computer-assisted Studies of Language and Culture**, Blackwell, Oxford.
- _____ (2001), **Words and Phrases**, Blackwell, Oxford.
- Tedick, Diane J. (1990), “ESL Writing Assessment: Subject-matter Knowledge and its Impact on Performance”, **Studies in Second Language Acquisition**, 9, 123–143.
- Thewissen, Jennifer (2013), “Capturing L2 Accuracy Developmental Patterns: Insights from an Error Tagged EFL Learner Corpus”, **The Modern Language Journal**, 97, 1, 77-101.
- Tognini-Bonelli, Elena (2001), **Corpus Linguistics at Work**, John Benjamins, Amsterdam.
- Tono, Yukio (2002), **The Role of Learner Corpora in SLA Research and Foreign Language Teaching: The Multiple Comparison Approach**, Unpublished Ph.D. thesis. Lancaster University, the United Kingdom.
- Trawinski, Mariuzs (2005), **An Outline of Second Language Acquisition Theories**, Krakow: Wydawnictwo Naukowe Akademii Pedagogicznej.
- Wei, Yaoyu and Lei, Lei (2011), “Lexical Bundles in the Academic Writing of Advanced Chinese EFL Learners”, **RELC Journal**, 42(2), 155–166.
- Wei-yu Chen, C. (2006), “The Use of Conjunctive Adverbials in the Academic Papers of Advanced Taiwanese EFL learners”, **International Journal of Corpus Linguistics**, 11(1), 113-130.
- Wolfe-Quintero, Kate et al. (1998), **Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity**, University of Hawai’i at Manoa, Hawai’i.

- Wong, Bee Eng and Teo, Pauline Hwa Ling (2012), “Elicited Imitation as a Measure of L2 English Learners’ Interlanguage Representation of Relative Clauses”, **Electronic Journal of Foreign Language Teaching**, 9(1), 91–107.
- Woolard, George (2000), “Collocation – Encouraging Learner Independence”, Michael Lewis (Ed.), **Teaching Collocation: Further Developments in the Lexical Approach**, in (28-46), Language Teaching Publications, Hove.
- Wouden, Ton Van Der (1997), **Negative Contexts: Collection, Polarity, and Multiple Negation**, Routledge, New York.
- Wray, Alison (2002), **Formulaic Language and the Lexicon**, Cambridge University Press, Cambridge.
- Xiao, Richard Zhonghua (2007), “What Can Sla Learn from Contrastive Corpus Linguistics? The Case of Passive Constructions in Chinese Learner English”, **Indonesian Journal of English Language Teaching**, 3(1), 1-19.
- Yoon, Hyung-Jo (2016), “Association Strength of Verb-noun Combinations in Experienced NS and Less Experienced NNS Writing: Longitudinal and Cross-sectional Findings”, **Journal of Second Language Writing**, 34, 42-57.
- Zareva, Alla et al. (2005), “Relationship Between Lexical Competence and Language Proficiency”, **Studies in Second Language Acquisition**, 27(4), 567-595.
- Zhang, Ronggen (2013), “A Corpus-based Error Analysis of Students' Writing in Graded Teaching Classes”, **Journal of Convergence Information Technology (JCIT)**, 8(10), 551-557.



APPENDICES

Appendix 1:

TUCLE Project Consent Form

Trabzon University Corpus of Learner English

TUCLE is a Learner Corpus Project in the Department of English Language Teaching at Trabzon University. The aim is to investigate the use of multi-word combinations in the academic essays that you will produce for the next two semesters (spring and fall) of 2018-2019. The project compiles a database of learner English, which will be transcribed and stored in electronic form. The written material will be used for research purposes only. Proper names and other identifying information will not be made public.

I hereby give my consent for my essay scripts to be used for research purposes.

Yes _____ No _____

Date: _____

Name _____

Signature _____

Appendix 2:

Trabzon University 2018-2019 Proficiency Exam

A) CHOOSE THE CORRECT OPTIONS IN QUESTIONS 1-28.

1. Some teachers are usually _____ with the performance of their students since they don't reflect their knowledge and intelligence as much as they can.
a. emotional b. responsible c. dissatisfied d. worried e. happy
2. Vincent Van Gogh was born in Brabant in the Netherlands in 1853. As a young man he worked as an art dealer in London and Paris. He was _____ from this job because he had argued with customers about art.
a. recognized b. dismissed c. employed d. tried e. quarrelled
3. Some authorities claim that the _____ approach of today's European life style to broken families thrived in the society notably after the World War II.
a. rewarding b. productive c. aggravating d. spawning e. indifferent
4. As some Third World governments in the continent Africa refuse in each time to _____ the aid offer each time, the WHO has correspondingly had to cancel the aid agenda now.
a. let down b. pass out c. fall in with d. round up e. ring up
5. Training in an English speaking country _____ your life in a positive way. First _____ and foremost, you can find a job easily and understand the culture of people.
a. effect b. effectively c. affect d. affectionate e. effective
6. In their _____ to overcome the anti-social effects of modern architecture, architects have directed their attention to more informal settlements.
a. avoidance b. condition c. attempt d. involvement e. development
7. They argue that today's armed forces should be prepared for multiple but smallish expeditionary operations in remote and disparate places, _____ for just one big conflagration.
a. as soon as b. than ever c. most of all d. more often e. rather than
8. With the help of a fifth of the white voters Mr Jackson was elected mayor of Atlanta in 1973, and _____ became the first black mayor of a major southern city.
a. nevertheless b. even so c. otherwise d. thus e. yet
9. After _____ 2,200 years, Latin _____ a slow decline around 1600s, making its use rare and at last extinct.
**a. lasting/ had begun b. lasted/ began c. having lasted/ began
d. being lasted/ began e. lasting/ began**
10. If my sister _____ us her plans at the beginning, these problems _____ us now.
**a. showed/ would not have been worrying
b. have shown/ would not be worrying
c. had shown/ will not be worrying
d. show/ will not be worrying
e. had shown/ would not be worrying**
11. Although Greece and Turkey are now working _____ mutual understanding, many observers feel that no real progress will be made _____ the situation of Cyprus still persisting.
a. at/in b. within/ since c. towards/with d. about/ before e. into/ at
12. The Atatürk Dam, which is _____ the largest dams in the world, is capable of generating 8.9 billion kilowatt-hours of electricity annually from the run-off the vast lake _____ its construction created.
**a. such/ of which b. one of/ that c. more than/ where
d. other than /when e. on/ which**
13. The British Government _____ details of the new stage of its genetically modified crop field trials, thus _____ off a now familiar cycle of debate and demonstration.
**a. has released/setting b. will release/ to set c. had released/ setting
d. is releasing/ to have set e. released/ to be set**
14. A curriculum that teaches one side _____ the story risks creating a nationalistic generation with little interest _____ international issues
a. of / in b. for/at c. for /to d. on/of e. of /on

15. Although a previous form of battery _____ in antiquity, the development of modern batteries _____ with the Voltaic pile invented by the Italian physicist Alessandro Volta in 1800.

- a. should be used/ will have started
- b. may have been used/ started
- c. might have been used/ have started
- d. would have been used/ had started
- e. can be used/ to start

16. People _____ as socially or politically aware since the early seventies; as the green movement _____ momentum in the late seventies.

- a. had been described/ gain
- b. has described/ has gained
- c. were described/ had gained
- d. have been described/ gained
- e. have described/ would gain

17. It _____ that the first use of the term "privatization" _____ in the 1930s with The Economist in covering the economic policy in Germany.

- a. has been claimed/ evolved
- b. is claimed/ was evolving
- c. would be claimed/ has evolved
- d. was claimed/ had evolved
- e. had been claimed/evolve

18. Metaphor is known for usage in literature, especially in poetry, where with _____ words, emotions and associations from one context _____ associated with objects and entities in a different context.

- a. some /is
- b. a lot /are
- c. a little/are
- d. few /are
- e. a good amount/ is

19. Health professionals warn that doing physical exercise could cause irreversible damage on tissues _____ it is performed in a form consistent with the body's physical features.

- a. now that
- b. however
- c. unless
- d. on condition that
- e. even though

20. Scientists are not sure why that civilization _____ 500 years ago. They _____ enough food, or perhaps their enemies attacked them.

- a. have disappeared / might not have had
- b. disappeared/ couldn't have
- c. disappeared/ cannot have had
- d. had disappeared/ shouldn't have had
- e. disappeared/ may not have had

B) FILL IN THE BLANKS.

The Spanish are famous for their natural sociability and appetite for life. They (23 _____) to put as much energy into (24 _____) life as they do into their work. The typical lazy Spaniard is a myth, but many people fit their work to the demands (25) _____ their social lives, rather than be ruled by the clock. The day is quite long in Spain, and the Spanish have a word, Madrugada, for the time between midnight and dawn, (26 _____) city streets are often still full of people enjoying (27 _____)

- 21. a. knew b. are known c. have known d. were knowing e. are to be known
- 22. a. enjoying b. to enjoy c. enjoy d. to enjoying e. to have enjoyed
- 23. a. towards b. at c. of d. into e. through
- 24. a. where b. which c. that d. when e. whose
- 25. a. their own b. theirs c. the other's d. one another's e. themselves

C) COMPLETE THE SENTENCES.

26. The election for President of the United States is a process of indirect elections _____

- a. that the presidential election is separate and simultaneous state elections.
- b. because Washington D.C is granted electors in the same number as the smallest state.
- c. although U.S territories cannot be represented in the cabinet system.
- d. until democrats include more extensive groups of delegates in Senate.
- e. that is regulated by a number of federal and state laws.

27. The reason why we have only one fiction written by Emily Bronte was not that she was an unproductive writer _____

- a. and the flair in her was instantly discovered among the family members.
- b. after she attended a private college with her sister Charlotte.
- c. but that there has been only one work of hers which has survived today.
- d. but Emily actually released a joint Literary work of poems with her sister in 846.
- e. so she preferred to use a male nickname to be free from the social prejudice in the 1850s.

28. Whereas Amazon.com, an American electronic commerce site, actually dates back to the early 1990s _____

- a. the web site is always powered and hosted by Amazon itself.
- b. the customers' reviews of the products are perpetually monitored by the executive board.
- c. the domain had attracted at least 600 million visitors annually in 2004.
- d. it gained popularity only in 2005 by selling products under its own label.
- e. the headquarter of the site is located in Pac Med building Seattle.

29. Even though there were a great many applicants for the vacancy _____

- a. not one of them had the qualifications required.
- b. actually the salary is expected to improve.
- c. the new recruits will be put on a three-week social training programme.
- d. unfortunately some of the staff had already resigned.
- e. the personnel department cope with the paper work.

30. Unless the central European countries can keep their labour costs at this level, _____

- a. exports are already starting to drop.
- b. they haven't been able to compete with western Europe.
- c. unrest in these countries would have been unavoidable.
- d. the finished products will have to go up in price.
- e. a firm control of distribution and sales has not proved adequate.

D) ANSWER THE QUESTIONS 31-38 ACCORDING TO THE TEXTS.

A population is a group of individual organisms of the same kind that are limited to some particular space. The most familiar example is the human population, but there are also populations of animals and plants everywhere on Earth. In fact, scientists regard a population as a biological unit that has both structure and function. The parts of a population are its individual members. The functions of a population are similar to those of other biological units: growth, development, and self-maintenance in a changing environment. Individuals enter a population by birth and by moving in, that is, by immigration. Individuals leave a population by death and by moving out, that is by emigration. If the environment of a population remains the same, loss and replacement of members in balance. The population will be able to survive in the particular environment. If the environment changes, however, loss or addition of members increases or decreases the size of the population.

31. It is pointed out in the passage that the changes that occur in the environment of a population _____

- a. have an impact, negative or positive ,on the members of that population.
- b. speed up the process of replacement of the members of the population.
- c. always contribute greatly to the survival off all the members of that population.
- d. are mostly caused by the uncontrollable size of that population.
- e. can be reduced through an increase in the size of the population.

32. According to the passage, what is called a "population" in biology _____

- a. can be defined as any group of organisms that is not subject to loss and replacement.
- b. is a biological unit that has only the function of growth.
- c. is a group of animals and plants that can survive all kinds of environmental changes.
- d. solely refers to any human group that lives in a specific region on Earth.
- e. is a unit that consists of the same kind of individual organisms living in a particular area.

33. It is clear from the passage that, so long as a population lives in a constant environment _____

- a. the growth, development, and self-maintenance of its members can be fully controlled.
- b. its size remains more or less stable.

- c. it usually undergoes a rapid structural change, which considerably affects its size.
- d. the replacement of its members is relatively slow, compared with other populations in different environments.
- e. its survival becomes difficult owing to the uncontrollable increase in its size.

34. As it is indicated in the passage, if the addition of new members to a population exceeds loss _____

- a. this can have a restrictive effect on emigrations from the population.
- b. this has no effect whatsoever on the environment in which the population lives.
- c. the survival of the population can be maintained in a balanced way.
- d. the size of the population shows a growing pattern.
- e. new measures must be introduced to prevent environmental changes.

Scottish philosopher and historian David Hume emerged as an economist also with the publication of his Political Discourses. The famous Adam Smith was a friend of his and may have been influenced by Hume: they had similar principles, and both were very good at illustrating and supporting these form history. Although Hume did not formulate a complete system of economic theory, as dis Smith in his Wealth of Nations, he introduced several of the new ideas around which the “the classical economics” of the 18th century was built. His economic philosophy can be understood from his main arguments: that wealth consists not of money but of commodities, that the amount of money in circulation should be kept related to the amount of goods in the market, and that poor nations impoverish the rest because they do not produce enough to be able to take much part in trade. Beyond this, he urged society to welcome the shift from an agricultural to an industrial economy, without which civilization could not be achieved.

35. According to Hume _____

- a. rich nations should produce more to be able to feed their citizens.
- b. rich nations should provide financial help to poorer nations.
- c. poor nations have a negative effect on richer nations.
- d. poverty can be overcome by increasing the production capacity of rich nations.
- e. poor nations can take part in international trade only when rich nations are impoverished.

36. According to text, Adam Smith _____

- a. was very much under the influence of Hume.
- b. formulated a complete system of economic theory.
- c. had ideas that conflicted with Hume’s.
- d. was uncertain about Hume’s principles.
- e. had a great effect on Hume.

37. According to the text, Hume _____

- a. was against the ideas on which the “ classical economics” of the 18th century was based.
- b. misunderstood the principles that his friend Adam Smith believed in.
- c. was not the only one who excelled at illustrating and supporting his principles from the past.
- d. argued that money in circulation had to be barely related to the amount of goods in the market.
- e. stated that civilization required advances in both agricultural and industrial production.

38. Hume’s belief was that poverty was mainly caused by _____

- a. the amount of money in circulation
- b. the scarcity of produced goods
- c. the abuse of poor nations by rich ones
- d. the lack of a complete economic theory
- e. Adam Smith’s poor grasp of economics.

E) CHOOSE THE EQUIVALENT SENTENCES IN MEANING.

39. In extremely cold regions, many animals hibernate, that is, they sleep through the winter, because in this way they can survive.

- a. Many animals sleep through the winter, that is, they hibernate, in really cold parts of the world, as this makes it possible for them to survive
- b. Hibernation, or the practice of sleeping right through the winter, is a survival technique favoured by many animals in very cold regions.

- c. Many of the animals that hibernate, that is, sleep through the long, cold winters, do so from choice, not necessity.
- d. When the winters are particularly cold, many animals choose to hibernate, that is, sleep through the long dark days and increase their chances of survival.
- e. In order to survive in the coldest parts of the world, many animals are forced to hibernate, or sleep through at least a part of the winter.

40. When UNICEF was founded in 1946, its main aim was to provide help for the many children in need as a result of World War II.

- a. When it was founded in 1946, UNICEF took a number of steps to improve the conditions of children that survived World War II.
- b. In 1946, after the end of World War II, UNICEF was set up to assist children throughout the world.
- c. Because so many children had suffered during World War II, UNICEF was set up in 1946 to give them a better education.
- d. In 1946 UNICEF was set up primarily because, as a result of World War II, large numbers of children were in need of help
- e. Following World War II, UNICEF was set up in 1946 and has since

41. If the meeting has to be on Monday, I can probably manage to come; but I'd much prefer Tuesday.

- a. Tuesday would suit me much better than Monday, but if the meeting's got to be on Monday I'll do my best to come.
- b. I won't be able to come to the Monday meeting, but I'll come to the Tuesday meeting.
- c. It would suit me better if the meeting was held on Tuesday as usual, but I suppose I could manage Monday.
- d. If the Tuesday meeting is put back to Monday, I don't think I will be able to come.
- e. I can't come to a meeting on Monday, but I can on Tuesday; would that be suitable?

42. Obviously concessions will have to be made on both sides if an agreement is to be reached

- a. Allowances will have to be mad on both sides if they fail to reach an agreement.
- b. If either side shows positive signs of giving way, the resulting agreement will not be in their favour.
- c. Both sides have admitted that they are willing to make any concessions.
- d. Before coming to an agreement, both sides will have to discuss any concessions they might consider making.
- e. Clearly there can only be an agreement so long as both sides are willing to give way over something.

F) COMPLETE THE PARAGRAPHS WITH THE MOST SUITABLE SENTENCE.

43. Paris, which is the capital of France, is situated on the Seine. It is a beautiful and historic city and has, therefore, become one of the world's main tourist centres. _____ These might include things as diverse as a visit to the Louvre and to the Euro Disney Theme Park.

- a. It has for several centuries been a centre of fashion.
- b. Paris is especially famous for its museums.
- c. There are a great many things for a visitor to do there.
- d. Many of the world's luxury goods are produced in Paris.
- e. The Palace of Versailles is just 23 kilometres south west of Paris.

44. _____ Furthermore, it shares the title of being the largest predator with the Kodiak bear which is actually an omnivore living in Alaska. Interestingly, an adult male is likely to weigh 650 kg and measure 3 meters. As for the females, they are some half the size and weight of males with roughly 250 kg and 2 meters. However, a female polar bear can weigh as much as 500 kg during pregnancy.

- a. The polar bear is equipped with astonishing furred feet allowing it good traction on ice.
- b. Polar bears are of a distinctive rank with its unique features for Arctic life.
- c. The largest carnivore on land is the polar bear which is twice as big as the Siberian tiger.
- d. The brown polar bear is, compared with its closest relatives, has a less elongated body and shorter skull.
- e. The polar bear, often considered a marine mammal, spend much of his time at sea.

45. _____ Of all movies, the most popular ones are: "Dracula" in 1931, "The horror of Dracula" in 1958 and Bram Stoker's "Dracula" in 1992. On the other hand, many other films have used this character as

a villain in addition to the ones naming him in their titles, such as “Dracula’s Daughter” or Besides of Dracula. an estimated 160 films have features Dracula in a major role.

- a. Jonathan Harker’s character largely indicates the matters of dwelling in a rationally modern world.**
- b. Most film adaptations of the character Count Dracula do not bear all the major characters in the novel.**
- c. The characters in the film “Count Dracula” largely utilise modern technology.**
- d. The character Count Dracula has never lost its popularity in film industry over the years.**
- e. The climaxes of Dracula films are traditionally the scenes in marble town of an old graveyard.**

46. Pope Benedict XVI is known to be technologically conservative and his teachings and writings always defended traditional Catholic values and doctrine. In that way, he always emphasizes what he sees as a need for Europe to return to fundamental values against the increasing secularisation and de-Christianisation in most of developed countries _____

- a. For this reason, he frequently stresses about relativism’s denial of objective truth.**
- b. Benedict XVI, the oldest person to have been elected Pope, was elected at the age 78.**
- c. Moreover, he serves as a cardinal of all Pope since Pope Clement XII.**
- d. He has also become the pioneer of a charitable organisation called Ratzinger Foundation.**
- e. Even well prior to becoming Pope, Ratzinger was already one of the most influential figures in Roman Curia.**

G) CHOOSE THE IRRELEVANT SENTENCES

47. (I) The first systematic survey on gorillas could only be conducted in the 1920s. (II) It was when a scientist, Carla Akeley, from the American Natural History Museum travelled to Africa. (III) The following observations on them were performed in 2005 by a team of Thomas Brever and Linda Bentley (IV) Her initial aim was hunting for an animal to be shot and stuffed. (V) Then, she became a devotee advocate for the conservation of gorillas and wrote several research books on them

- a. I b. II c. III d. IV e. V**

48. (I) One of the most frequently depicted themes in fiction is” person vs. society”. (II) it is often concerned with a main character’s or group of main characters’ conflict with social traditions or concept. (III) In this sense, there exist two parties: the protagonist and the society in which the protagonist is included. (IV) the most striking example in literature in this sense is Wuthering Heights written by Emily Bronte. (V) “The person vs. Supernatural” is another literary theme placing a character against supernatural forces.

- a. I b. II c. III d. IV e. V**

49. (I) Marine pollution is often a genetic term referring to the penetration of some harmful particles and chemicals into the sea. (II) However, marine debris is used to describe human-created waste. (III) The major culprits are undoubtedly rivers emptying into the sea with many chemicals such as pesticides. (IV) The irrevocable result of such a process is “hypoxia (V) In scientific sense, this case is defined as the decreased of oxygen depleting chemicals into the water.

- a. I b. II c. III d. IV e. IV**

50. (I) Polar bears rarely live more than 25 years. (II) One of the oldest wild bears recorded was 32 years old when he died. (III) Besides the oldest captivated one was a female at 43 years old who died in 1991. (IV) They are also less affected by infectious diseases and parasites than many other terrestrial mammals. (V) As for the oldest living polar bear, Debby who was born in 1966 is now leading a contented life in Assiniboine Park Zoo.

- a. I b. II c. III d. IV e. V**

Appendix 3:

Trabzon University 2018-2019 Proficiency Exam, Listening Test

A) Fill in the blanks with what you hear. (3 POINTS EACH)

1. I usually have meat or (1)..... Usually (2).....or something as a starter and then maybe (3)..... for the main course.
2. I quite often have (4)..... vegetable soups that you only have to heat up-in fact they're the only vegetables I ever eat! And I usually have a couple of frozen pizzas in the freezer for (5)..... I don't really order (6)when I'm on my own, but if I'm with friends in the evening, we sometimes order Chinese food for dinner.
3. Eggs and Coke. I have eggs for breakfast at least twice a week and I drink a couple of cans of coke every day.
4. If I'm feeling down, chicken soup with nice big (7)..... of chicken in it. It's warm and (8)..... I usually have a banana before going to the gym. If I know I'm going to have a really long meeting, I usually have a coffee and a cake because I think it will keep me (9).....and give me energy.
5. Fruit-cherries, strawberries, (10).....and apples. Vegetables-peppers, tomatoes, and cucumbers. The only I don't like is beetroot. I can't even stand the smell of it.

B) Answer the questions according to the track. (8 POINTS EACH)

1. What did Steve like eating when he was six?
.....
2. Why did he decide to go Spain?
.....
3. How was Steve's restaurant different from typical Spanish restaurants?
.....
4. What does not he like cooking?
.....
5. What does he like cooking?
.....

C. Write True or False. If there is a false sentence, write the correct answer. (3 POINTS EACH)

1. Jane went to Malta with her family.
2. She writes thrillers.
3. She is married with three kids.
4. They were travelling in a taxi when it broke down.
5. The children who came near Jane spoke English.
6. Jane was shocked by the good condition of the school.
7. She taught the students "Heads, shoulders, knees and toes" song.
8. The school was for kids of rich families.
9. The headmaster said they needed a brand-new building.
10. When she turned back, she forgot about the school and the students.

Good Luck!

CURRICULUM VITAE

Handan İLYAS KARATAŞ, born in 1984 and in Trabzon, completed her whole education life in Trabzon. She graduated from Tefvik Serdar Anatolian High School in 2002, and right after that she started her university education in Karadeniz Technical University. She graduated from the Department of English Language and Literature in 2007 with the degree of the second place. She started her master's degree in 2007; however, due to some problematic issues, she could not have the chance of completing her post-graduate education in the expected period. She worked at Rize University for three years, then she began working at Karadeniz Technical University. While working at KTU, she had the chance of visiting Germany via Erasmus exchange program for 2 months. She served her internship in Chemnitz. After working for 7 years at KTU, she was transferred to Trabzon University which is the second state university in Trabzon. She currently works there at the Department of English Language Teaching.

She is married and the mother of two daughters.