



**T.C.**  
**YEDİTEPE UNIVERSITY**  
**INSTITUTE OF EDUCATIONAL SCIENCES**

**FACTORS INFLUENCING TURKISH EFL LEARNERS'  
COLLOCATION PRODUCTION**

**by**

**Sinan NİŞANCI**

**Submitted to the Institute of Educational Sciences  
in partial fulfillment of the requirements for the degree of  
Master of Arts  
in English Language Education**

**ISTANBUL, 2014**




YEDİTEPE ÜNİVERSİTESİ  
EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ

TEZ TESLİM ve ONAY TUTANAĞI

KONU: Factors Influencing Turkish EFL Learners' Collocation Production

ONAY:

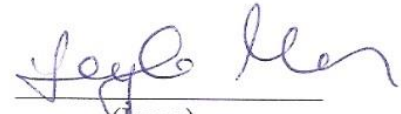
Yrd. Doç Dr. Erkan Karabacak  
(Danışman)

  
(İmza)

Öğr. Gör. Hossein Farhady  
(Üye)

  
(İmza)

Doç. Dr. Leyla Martı  
(Üye)

  
(İmza)

TESLİM EDEN : Sinan Nişancı  
TEZ SAVUNMA TARİHİ : 5.06.2014  
TEZ ONAY TARİHİ : 5.06.2014

## TABLE OF CONTENTS

<b>TABLE OF CONTENTS</b> .....	<b>ii</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>iv</b>
<b>LIST OF TABLES</b> .....	<b>v</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>vi</b>
<b>ABSTRACT</b> .....	<b>vii</b>
<b>ÖZET</b> .....	<b>viii</b>
<b>I. INTRODUCTION</b> .....	<b>1</b>
1.1. Purpose.....	1
1.2. Significance.....	1
1.3. Research Questions .....	4
1.4. Limitations and Delimitations.....	5
<b>II. LITERARY REVIEW</b> .....	<b>7</b>
2.1. Definitions of Collocations in the Literature .....	7
2.2. Approaches to Collocation Studies in SLA Research.....	11
2.3. A Review of Learner Corpus Research into Collocations .....	12
2.4. The Origins of Collocational Errors .....	16
<b>III. METHODOLOGY</b> .....	<b>22</b>
3.1. Participants.....	22
3.2. Data .....	23
3.3. Procedures.....	23
<b>IV. RESULTS</b> .....	<b>28</b>
4.1. The Degree of Restriction.....	29
4.2. Semantic Transparency .....	30

4.3. L1 Influence in Collocational Production.....	30
<b>V. DISCUSSION .....</b>	<b>35</b>
<b>VI. CONCLUSIONS.....</b>	<b>39</b>
6.1. Implications and Applications .....	39
6.2. Suggestions for Further Research .....	42
<b>REFERENCES.....</b>	<b>44</b>
<b>APPENDIX: LIST OF INCORRECT COLLOCATIONS.....</b>	<b>48</b>

## LIST OF ABBREVIATIONS

BNC	British National Corpus
COCA	Corpus of Contemporary American English
EFL	English as a Foreign Language
ICLE	International Corpus of Learner English
L1	First Language
L2	Second Language
LLs	Language Learners
NL	Native Language
SLA	Second Language Acquisition
TL	Target Language

## LIST OF TABLES

Table 2.1. Summary of the findings of restrictedness .....	15
Table 3.1. Details of the corpus .....	23
Table 4.1. A summary of the collocations extracted .....	28
Table 4.2. Restricted collocations .....	29
Table 4.3. Semantic transparency of the collocations.....	30
Table 4.4. Errors with L1 origin .....	31
Table 4.5. Examples of synonymy as a learner strategy.....	34

## **ACKNOWLEDGEMENTS**

In completion of this paper, I owe special gratitude to my research supervisor Assist. Prof. Erkan KARABACAK, who, with his in-depth expertise in corpus linguistics, offered me invaluable guidance in how to organize and analyze my data and training in using corpus tools.

Besides, I would like to thank my wife and two kids, Orhan and Oğuzhan, as they voluntarily allowed me to focus on my work.

## ABSTRACT

This study aims to detect the factors influencing Turkish learners' production of English collocations. Working specifically on *verb + object noun* combinations, the effect of three important factors were analyzed: restriction on the combinability of words, semantic transparency and L1 transfer. 76 writing exam papers by 11th graders at a public school were collected and built into a learner corpus. The corpus was tagged with Stanford English Tagger (3.2.0), and *verb + noun* collocations were automatically extracted using the AntConc 3.2.4w concordancer. Manually analyzing the data, a total of 111 collocations were listed, and 30 of them were marked as incorrect-use. The study found that students had slightly poorer performance in collocations with more restricted and less transparent constituents. In addition, L1 proved to have a minimum effect on incorrect collocation production. Learners seem unable to recognize that collocations are not free combinations but fixed phrases which need to be stored as single items.

**Key Words:** Collocation production, restriction on combinability, semantic transparency, corpus, L1 influence.



## ÖZET

Bu çalışma, Türk öğrencilerin İngilizce sözcük öbekleri üretimlerine etki eden faktörleri tespit etmeyi amaçlamaktadır. Sadece *fiil + nesne* öbekleri üzerinde duran araştırmada şu üç faktörün etkisi incelenmiştir: Sözcük öbeği oluşumundaki sınırlama prensibi, anlam şeffaflığı ve ana dilin etkisi. Bir devlet okulunun 11'inci sınıfındaki öğrencilerin 76 kompozisyon sınavı kâğıtları toplanarak bir öğrenci derlemi oluşturulmuştur. Stanford English Tagger (3.2.0) adlı yazılım ile derlemdeki kelime grupları belirlenmiş ve AntConc 3.2.4w adlı derlem tarama yazılımı kullanılarak *fiil + nesne* sözcük öbekleri otomatik olarak tespit edilmiştir. Verinin elle analiz edilmesiyle, toplam 111 adet sözcük öbeği elde edilmiştir ve 30 tanesi yanlış kullanım olarak işaretlenmiştir. Araştırma, öğrencilerin daha fazla sınırlamalı ve anlam olarak daha az şeffaf olan sözcüklerden oluşan öbekleri kullanmada az da olsa daha başarısız olduklarını bulmuştur. Ayrıca, yanlış sözcük öbeği üretiminde en az etkinin ana dilden geldiği sonucuna varılmıştır. Araştırma bulgularına göre, öğrencilerin sözcük öbeklerinin serbestçe bir araya gelen tamlamalar değil, ayrı birer kelime/kalıp olarak öğrenilmesi gereken sabit ifadeler olduğunu anlamakta başarısız oldukları görülmektedir.

**Anahtar Sözcükler:** Sözcük öbeği üretimi, sözcük öbeği oluşumundaki sınırlamalar, anlam şeffaflığı, derlem, ana dilin etkisi.

## I. INTRODUCTION

Throughout my 15 years of experience in teaching I have come to the conclusion that most learners with a wide range of vocabulary storage and an excellent control of grammar may have problems with fluency in writing and speaking. Besides, when listening to something, they have a tendency to try to catch every word they hear, whereas expert speakers of English can understand utterances even in a noisy environment without hearing every word because they can predict and anticipate all the possible alternatives that may come after a certain word. Learners, especially after reaching “advanced level”, are expected to be able to produce near native-like and natural utterances and understand long stretches of discourse. However, many learners fail to achieve this level of proficiency. The difference in the way expert speakers of English and EFL/ESL learners process lexical and syntactical structures deserves investigation.

### 1.1. Purpose

This study was designed to determine the factors influencing the collocation production of EFL learners who are native speakers of Turkish. The collocability principles of restriction and semantic transparency on *verb + object noun* combinations were analyzed in order to investigate their effect on collocational performance. The study also aimed to determine to what extent the first language transfer plays a role in collocation production.

### 1.2. Significance

Lewis (1997) discusses that vocabulary is stored in chunks not as individual words. These language units can be retrieved from memory as a whole, and this reduces the processing burden on the brain. With less time and effort spent on processing, “expert” users of English (McCarty & Carter, 2001, p. 69) retain word combinations and

produce fluent and natural utterances both in speaking and writing. This also makes it easier for them to follow a speech and fully comprehend it. On the other hand, learners who only learn individual words have to cope with greater processing drudgery and will need a lot more time and effort to express themselves. Their utterances may at times not sound natural, as some sequences of words they use may not appear together in a native speaker's lexicon. This explains why native speakers are fluent in speech and writing and can have a superior comprehension whereas many learners cannot. Consequently, learners need to beware of the chunks in the target language. They should be provided with opportunities to identify, organize and record words that go together, and thus build an understanding of collocational learning, which may then be helpful in developing the expert use of collocations in the target language.

Collocational competence has been gaining popularity in the world, but to my present knowledge and experience in Turkey it has not attracted the attention it deserves. Although many students are incredibly good at producing grammatically complex sentences, their language does not sound natural presumably because they have awkward word combinations. Can collocational competence truly heal the Turkish EFL learners' problems and enhance their proficiency? Delving into this question, the present study attempts to find an answer to why a learner, even at advanced levels, fails to bring together words that sound natural. Some of the literature on collocations suggests that lack of collocational competence stems from L1 influence, yet there is also research ascribing this to such factors other than L1 interference, as paucity of collocational input and ineffective approaches to teaching and learning vocabulary. Therefore, this study aims to detect areas of difficulty in Turkish EFL learners' collocation production.

The idea that languages are made up of frozen, formulaic phrases, has recently occupied the central focus of current approaches to language teaching, yet how such formulaic language is learned has not satisfactorily been comprehended. The attempts to answer this question fall into two distinct categories. While some researchers have the opinion that L2 learners acquire lexical items in chunks, just as the native speakers do

(Ellis, 2003), others believe that, L2 learners, unlike L1 speakers, perceive and learn language input as isolated, individual words (Wray, 2002).

In one of the studies reviewed for this paper Durrant and Schmitt (2010) claim that second language learners do not, as Wray (2002) claims, totally rely on individual words, but retain memory of words that often appear together. “Any deficit in learners’ knowledge of collocation” according to their findings stems more likely from “insufficient exposure to the language”, than from the different ways learners are taught collocations (Durrant & Schmitt, 2010, p. 182).

On the other hand, sticking to Ellis’ model, Keshavarz and Salimi (2007) report that vocabulary teaching should capitalize not on “semantic isolates”, but on “semi-preconstructed phrases”. This standpoint implicitly blames learners’ poor collocational competence on ineffective teaching approaches. Jean (2007), having found that university students have more serious problems in productive collocational competence, states as an implication of the study that this can be overcome by an efficient pedagogical intervention program.

According to some other research, some creative cognitive processes, which mostly result from L1 transfer, are held liable for incompetence in collocations even at advanced levels (Zughoul & Abdul-Fettah, 2001). The study on Arabic learners of English concludes that even advanced learners produce collocations inadequately on account of direct translation from native language (NL) to target language (TL). Huang (2001) and Nesselhauf (2003), too, emphasize that L1 transfer has a more profound effect on collocational errors.

The literature agrees on the significance of collocations for both L1 speakers and L2 learners. Hill (2000), for instance, refers to lexis, and mentions the countless number of collocations that native speakers have at their disposal and how people state their opinions. He then goes on to deduce a list of nine reasons accounting for the properties that bring collocations to the fore. Duran and Schmitt (2010) also enumerate a couple of reasons as to why collocations occupy a central position in language research. Some of

these are their pervasiveness, salience and the already developed quantitative methods to study them. It is possible, therefore, to strongly assert that collocational knowledge is the core of language knowledge, collocational knowledge enables learners to be both fluent and accurate, and only if learners know the collocates of a word does it mean that they know the word (Nation, 2001).

Today, as a corollary to the increasing interest in research into collocations, it is widely accepted, that collocational knowledge occupies a crucial place in SLA. The large body of research carried out in the field agrees that collocational competence is one of the foremost problems for L2 learners irrespective of their length or method of instruction, command of the target language or L1 background (Nation, 2001; Wray, 2002; Schmitt, 2004; Boers et al., 2006). L2 learners' lack of collocational knowledge results in awkward utterances in their writing and speech. That's why "teaching collocations should be a top priority in every language course" (Lewis, 2000, p. 8).

As a result, collocations do occupy a central place in second language acquisition process. To date, no research studies have been reported on the factors influencing Turkish EFL learners' collocational competence. Considering the significance of the topic, Turkish learners, teachers and researchers desperately need to be familiarized with the concept of collocations. The present study attempts to fill in the gap addressing to the areas of difficulty Turkish EFL learners have in their collocation production.

### **1.3. Research Questions**

The study attempts to answer the following questions:

RQ 1: What is the relationship between restriction on collocability and Turkish EFL learners' collocation production?

RQ 2: What is the relationship between semantic transparency of the constituents of a collocation and Turkish EFL learners' collocation production?

RQ 3: What is the role of native language on collocation production?

#### **1.4. Limitations and Delimitations**

This study covers only the *verb + object noun* collocations as they were found to be one of the most troublesome combinations for language learners (Källkvist, 1998; Bonelli, 2000; Liu, 2000; Nesselhauf, 2005). In addition, since the construct of collocations is used as an umbrella term for many multi-word units including grammatical combinations as well, all the prefabricated utterances could not be covered in a thesis paper. Therefore, this paper only deals with those lexical collocations excluding grammatical phrases, idioms and free combinations. Another reason for this preference is that errors in lexical choices impair comprehensibility of an utterance whereas even with grammatical mistakes a statement proves to be understandable. It is lexical errors – and particularly those relating to lexical collocations – that mark non-nativeness (Waller, 1993; Källkvist, 1998). Lexical collocations, thus, appeal more to researchers.

A problem that showed up in the course of data collection process was some students missing exams. Due to the tight schedule at school program during the exam week, some of the students who were on sports teams were out of school, so they couldn't take the exam. Although the school gave them makeup exams, because of time constraints, their papers could not be included in the study. Therefore, the 21.031-token corpus used in this study might have actually been larger if all the students could have been given the writing exam as planned.

Most natural writing occurs outside the classroom setting, and students have at their disposal references and material to proofread and edit their writing. Since the corpus was built out of students' exam papers, participants did not have any reference resource to consult to while writing their essays. Additionally, sitting an exam causes anxiety in students regardless of their grade or level of English. An indispensable result of these two

factors was poor spelling. Students had spelling mistakes a lot more than they would normally do. Since spell check on word processor cannot always catch all the erroneous spelling, each paper was carefully proofread while they were being typed in computer.

Another possible problem was the tagging software used in the analysis process. Although Stanford Tagger (3.2.0) is a commonly used computerized tagging tool which proved to have 97.5% per-token accuracy (Manning, 2011), certain problems were detected. During detailed manual analysis of the concordance lines, it appeared that some present and past participles in the corpus data were identified as adjectives. To compensate for such deficiencies that might arise due to the inefficiency of the tagging tool, a syntax was formed to extract both *adjective + noun* and *verb + noun* collocations. Manually analyzing concordance lines, only the real *verb + noun* combinations were identified and added into the target collocations list. For instance, *captivating animals* appeared in *adjective + noun* combinations, yet as the following concordance line shows, it should be classified in the *verb + noun* collocations.

*I think captivating animals is not panacea for our world.*

## II. LITERATURE REVIEW

Learner corpora have been used to study the collocational performance of L2 learners from different perspectives. Some of the issues examined are finding the most frequent collocations (Shin & Nation, 2008), different ways of extracting and detecting collocational errors (Li, 2005; Futagia et al., 2008; Saif & Aziz, 2011; Orenha-Ottaiano, 2012; Pecina, 2010), L2 learners' choice of collocation patterns and its effect on collocation learning (Martelli, 2007; Webb & Kagimoto, 2010; Laufer & Waldman, 2011), and collocations and vocabulary learning (Wei, 1999; Durrant & Schmitt, 2010). This pervasive interest in collocations attests that acquiring these prefabricated combinations helps learners improve their language fluency and approximate native-like use of L2. Therefore, the following section reports on the different aspects and principles governing collocational performance. The literature is reported in four respects: definitions of the construct, different trends in collocation research, review of some relevant studies, and the origins of collocational errors.

### 2.1. Definitions of Collocations in the Literature

The concept of collocations was pioneered in 1957 by the British linguist J. R. Firth. When he first suggested that a word acquires its meaning partly through other words in its close vicinity, he was indeed questioning the traditional perception of word classes and introducing a novel conception of language as a continuum, grammar at one end and lexis at the other (Firth, 1957 as cited in Martelli, 2007). Since then the term has been tailored by a number of researchers making many additions and adaptations. Each attempt to define collocations gave birth to a new conception, so today the literature embodies fifty distinct yet mostly synonymous terms (Wray, 2000).

Duran and Schmitt (2010) admit that there is not a universally accepted definition of collocations. Referring to different researchers, they mention such terms as “semi-preconstructed phrases”, “linguistic chunking” and “psychological association between



words” (p. 164). Their definition depends upon two broader terms of “formulaic language” and “construction”, classifying collocations under the former, which is defined, by Wray (2002, p. 9), as “a sequence, continuous or discontinuous of words or other elements which is [...] stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar”.

Keshavarz and Salimi (2007) also contend that “we still lack a clear, non-controversial and all-embracing definition of collocation” (p. 85). Stating various researchers’ dissimilar terms or definitions, they adopt a “categorization model of collocation”, since it is more thorough in its exposition and unequivocal with its easy-to-follow examples. This continuum based model consists of “free combinations, restricted collocations, figurative idioms, and pure idioms” (p. 85).

In fact, there is a substantial number of researchers using the concept of collocations as an all-inclusive term covering most multiword expressions. Nation (2001), for instance, takes collocations as an umbrella term putting under it idiomatic as well as fixed expressions. This tendency seems to exert far-reaching impact on the digestion of the term, which comprises all the multiword units. In this viewpoint, collocations are just one type of multiword expressions. Others are “fillers (e.g., *sort of*), functional expressions (e.g., *excuse me*), idioms (e.g., *back to square one*), proverbs (e.g., *let’s make hay while the sun shines*), and standardized phrases (e.g., *There’s a growing body of evidence that*)” (Boers et al., 2006, p. 246).

Another comprehensive approach to the concept of collocations is the general inclination to classify collocations in two broad categories. Multiword expressions fall into lexical collocations. The second class is grammatical collocations or *colligations*, “combinations of words in which some sort of syntactic dependency is involved, such as the relation between a verb and its complementation patterns” (Martelli, 2007, p. 23). Roughly speaking, the distinction between the two is based on word classes. While lexical collocations are word combinations from open classes – such as an adjective and a noun – grammatical collocations are composed of a combination of elements from open

class and closed class – such as an adjective from the former (afraid) and a preposition from the latter (of), hence *afraid of*.

Huang (2001), accepting the diversity in definitions, places collocations on a continuum, too, ranging from free combinations to pure idioms. He further touches on a few criteria for this classification which include “semantic transparency, degree of substitutability and degree of productivity.” In this approach, free combinations are the most transparent and have the highest degree of substitutability and productivity whereas idioms are the most semantically opaque with the lowest degree of productivity and substitutability (p. 114).

Jean (2007), also conceding the diversity of approaches to the definition of collocations, taps on two characteristics collocations should have. To him, collocations have formal and functional features. The formal features relate to the constituents of a collocation: a base and a collocate, which have dissimilar linguistic properties, as “the base is semantically autonomous whereas the collocate is determined and somehow selected by the base” (p. 129). Among the functional features, two are worth mentioning. First, the frequent cooccurrence of words are salient in languages. Certain frozen words relatively frequently exist together. Second, collocations are arbitrarily existing word combinations. According to this feature, “although the expression *to finish a war* is acceptable as grammatically and semantically correct, a proficient user of English would usually say *to end a war*” (p. 130), and this has no rule-governed explanation.

Nesselhauf (2003) is concerned with delimiting the collocations from other types of word combinations and asserts that these criteria do not help remove these concerns. In order to overcome this problem, she develops a notion, called “restricted sense”, on which she bases her classification of word combinations and definition of collocations. On the basis of this notion, she mentions three categories of word combinations, “free combinations, collocations and idioms” (p. 225).

All this diversity lucidly causes some confusion. To get rid of this miscellany, Nesselhauf (2005) ascertains two basic approaches: the frequency-based approach and

the phraseological approach. The frequency-based approach derives its data from corpus linguistics. Word combinations are considered collocations under the condition that their cooccurrence have a statistically significant frequency. In the phraseological approach, however, two concepts are of paramount importance in terms of delimiting collocations from other multiword units. First, the selection of the constituents is arbitrary, not semantically accounted for. This distinguishes collocations from free combinations. The second is semantic transparency, which suggests that “at least one of the elements has a literal meaning” (Martelli, 2007, p. 17). Thus, collocations are separated from idioms, whose meaning is non-transparent.

The present study takes collocations as combinations of elements from open classes. Therefore, the first category excluded is grammatical combinations. Complementation patterns (e.g. gerund/infinitive complementation), phrasal verbs (e.g. *make out, take after*) or any combinations of open class + closed class units (e.g. *afraid of, focus on*) are not considered as collocations.

A lexical combination whose frequent co-occurrence has some restriction – either semantic or arbitrary – will be taken as a collocation. Thus, a word combination with at least one restricted element will be counted as a collocation, which excludes free combinations. For instance, *take* and *a book* come together without any restriction as in the sentence “He took the book and ...” thereby considered a free combination, while *take a precaution* is counted as a collocation due to the restriction on the constituting elements. As regards the distance between a verb and its collocating noun, the finding that “the usual measure of proximity is a maximum of four words intervening” (Sinclair, 1991, p. 170 as cited in Martelli, 2007) was accepted as a determining factor. This four-letter span enabled to catch combinations even when a collocating noun was not adjacent to a verb.

Finally, the last criterion is semantic transparency. Combinations which have at least one element with literal meaning will be counted as collocations. The degree of transparency may vary – as a matter of fact, it is one of the factors studied in this paper -;

however, combinations with both constituents being totally opaque will not be analyzed. This excludes the idiomatic expressions. In *pay attention*, for example, *pay* is used in figurative sense, yet it contributes to the overall meaning of the collocation of *pay attention*. However, both of the elements in *pay the piper* are opaque and do not contribute to the overall meaning of the phrase. Therefore, any combination of such will be ignored as they are idioms.

In sum, this study defines collocations through the perspective of phraseological approach. Therefore, “a combination of items belonging to open classes which seem to ‘attract’ each other and establish a relation of co-occurrence along the syntagmatic axis, whose constituents are limited in their commutability, subject to some syntactic variation and semantically transparent” (Martelli, 2007, p. 31) will be considered as a collocation.

## **2.2. Approaches to Collocation Studies in SLA Research**

Research into collocations may follow several different paths, yet the data on how L2 learners retain and produce prefabricated utterances generally comes from two sources. While some researchers favor elicitation techniques, especially while working on some specific collocations, some others prefer to use natural language through compilation of learner corpora (Leśniewska, 2006). While the former utilizes certain tests to observe learners’ performance of some pre-selected word combinations, the latter analyzes collocations naturally produced by learners in written or spoken texts.

Using both these sources of data has their advantages and disadvantages. When the decisions learners make about the selection of collocates of a word are elicited via a test, researcher’s load is lessened since the focus is on limited items. Using gap-filling activities, multiple choice test items, cloze exercises or acceptability judgments, data governing collocational performance is obtained and compared to native speaker production. Much of the existing research into collocations adopt elicitation techniques (Lozano & Mendikoetxea, 2013).

The literature offers some rationale behind preferring elicited data over natural one. One obvious reason is that the lexical combination under scrutiny may not appear in the natural data. The absence of a particular item does not necessarily imply that learners are not aware of its existence. What's more, it seems hard with natural data to control the variables. Another important factor is researchers' level of training. Most researchers are already well-trained in data elicitation techniques having far more experience than they have in corpus tools as they are only recently used in SLA research (Myles, 2005).

Corpus linguistics as an SLA instrument, on the other hand, presents somewhat limitless data for researchers which help conceptualize learner production. It is possible to investigate what learners can naturally produce. Analysis of such data can offer many insights into nonnatives' use of the target language. Data gathered in this way can enable a researcher to reveal what lexical or grammatical structure learners with distinct linguistic backgrounds would need in order to improve their writing skills, for instance (Orenha-Ottaiano, 2012).

The seeming difficulty in coping with vast amount of data in a corpus can easily be removed with the use of automated techniques like concordancers. Researchers can automatically build and analyze a corpus, and then compare the findings with those of other learner or native speaker corpora. A corpus also allows for repetitive use. As the need arises researchers can resort to the corpus they developed over and over to extract further results. Finally, with its applicability and large body of data it has, a learner corpus promises more reliable findings (Futagia et al., 2008).

### **2.3. A Review of Learner Corpus Research into Collocations**

The most comprehensive research into collocations is perhaps Nesselhauf's (2005), in which she delved into German learners' collocational performance. She extracted *verb + noun* collocations from the German component of the International Corpus of Learner English (ICLE). Admitting to the fact that even advanced learners

have difficulty in producing proper collocations, she attempted to bring to the light the factors leading to problems in L2 learners' production of collocations. Having studied the incorrect instances of *verb + noun* collocations, she concluded that learners' L1 was responsible for most of the poorly produced collocations. She also found out that collocations with a medium degree of restriction were the most troublesome combinations for language learners.

Liu (2000) also studied the collocational patterns that L2 learners have difficulty processing. He concluded that *verb + noun* lexical combinations are the most troublesome for language learners. There is a host of research delving more into *verb + noun* collocations where the verb is delexicalized; that is, when it "loses its original lexical value and often acquires other meanings and other functions within a larger unit" (Bonelli, 2000, p. 229), as *take* in *take a nap*. All such research suggests that collocations with delexicalized verbs pose specific difficulty for ESL/EFL learners (Bonelli, 2000; Keshavarz & Salimi, 2007). They attest that the problem with the delexical verbs stem from their lack of lexical transparency and less restriction in collocability, and learners resort to direct translations from their mother tongue in such cases.

Actually, interest in restricted combinability of collocations started much earlier. Källkvist (1998) dealt with the issue of restriction on the verbs in *verb + noun* combinations. In her analysis of the error types in learners' collocational performance, she found that collocational errors increase when such common verbs as *make, do, have, take* are used. Since, being the most frequent verbs, they are among the words learned at the very early stages of the L2 acquisition process, learners have a tendency to overuse them. Källkvist ascribes this to the polysemous nature of such verbs. Due to overuse, learners fail to realize subtle differences in the use of these verbs. In other words, they cannot notice the restrictions on such high-frequency words. The conclusion to be drawn from this study is that L2 learners' collocational performance becomes poorer as less restrictions apply on the verbs.

Similar conclusions were reported by Howarth (1998), who studied *verb + direct object* collocations in a learner corpus composed of advanced learners' academic writing. He claims that collocations which are neither completely free combinations nor fixed as idioms are the most problematic ones for language learners. This, he speculates, should be because fixed combinations are processed as one lexical item whereas free combinations are learned separately. Granger's (1998) findings with the *verb + adverb intensifier* collocations extracted from the French part of ICLE are of the same kind. L2 learners' collocational competence approximate to that of native speakers when the adverb intensifiers in the collocations have fewer restrictions.

Leśniewska (2006) discusses the possible reasons as to what accounts for the extra difficulty that L2 learners have learning some lexical combinations. Referring to research done in psycholinguistics, Leśniewska, too, brings forth the concepts of semantic transparency and restrictedness. Semantically nontransparent collocations seem to be easier for learners to process than transparent ones. Due to their saliency, semantically opaque combinations like idioms attract more attention and are noticed much earlier by language learners. Initially considering them as difficult items to be learned, LLs concentrate more on them. Combinations which are totally free and fixed are not perceived to be difficult by learners. The same principle applies to the issue of restrictedness. As the number of nouns a verb can collocate with increases, i.e. less restriction applies, so does the number of incorrect productions.

Not all studies, however, reach the same conclusion. Huang (2001), for instance, studied Taiwanese EFL students' collocational performance analyzing their error types. Classifying collocations into four types (free combinations, restricted collocations, figurative idioms, pure idioms), the researcher concludes that free combinations and pure idioms cause the greatest difficulty for language learners. Subjects did slightly better on free combinations and figurative idioms, which was attributed to their flexibility and positive L1 transfer. Therefore, different from other research, Huang asserts that when a word has a greater range of collocates, that is less restriction, learners' collocational performance increases.

In a more recent study, Martelli (2007) used the Italian sub-corpus of ICLE to analyze the erroneous combinations produced by L2 learners for all types of collocations. Out of the 143 manually extracted collocational errors, verb + noun combinations proved to be the most problematic for learners. As for semantic transparency and restriction in collocability, the results reveal some similarities and differences compared to other studies. Most of the Italian learners' erroneous collocations are those the constituents of which are fully transparent. This is similar to earlier findings, which suggest that totally opaque combinations like idioms are learned better as they are processed as single vocabulary items. The issue of degree of restriction, however, bears contradictory results. Learners face greater problems when producing collocations that have a high degree of restriction in combinability. The contradiction in the findings of studies based on restriction principle is demonstrated in the Table 2.1 below.

**Table 2.1. Summary of the findings of restrictedness**

Researcher	Conclusion
Granger (1998) Howarth (1998) Källkvist (1998) Bonelli (2000) Nesselhauf (2005) Leśniewska (2006)	Less restriction causes poorer collocation production.
Huang (2001) Martelli (2007)	More restriction causes poorer collocation production.

To summarize the learner corpus research into collocations, the most conspicuous result is the unanimous agreement on the *verb + noun* phrases being the most challenging for L2 learners. That explains why this specific type is the focus of this study. Besides, despite the general consensus that learners are more successful in producing word combinations with more restricted and less transparent collocates, as they are considered to be salient, so attract more attention, there is also support for the opposite.



## 2.4. The Origins of Collocational Errors

Different researchers may define collocations in distinct terms, and put forward diverse reasons for learners' incompetence, yet there seems to be a unanimous agreement on the fact that L2 learners are not adequately proficient in their collocational competence. All the literature reviewed in this study concludes that learners, even at advanced levels, produce awkward utterances due to nonstandard use of collocations. Lack of collocational competence may, as some of the literature asserts, stem from interlingual transfer, while in some other research this incompetence is ascribed to such factors, other than L1 interference, as paucity of collocational input and ineffective approaches to teaching lexis, which result in learners' wrong conceptualization of the construct.

The origins of collocational errors are best revealed through a careful look at how collocations are learnt. Ellis (2003) contends that native speakers learn, store and retain formulaic language as single items through a process called chunking. The basic principle in this model is the "law of contiguity", which states that "objects once experienced together tend to become associated in the imagination, so that when any one of them is thought of, the others are likely to be thought of also" (p. 43). Frequent co-occurrence of words triggers associative learning in long term memory. All this chunking process, Ellis claims, takes place implicitly, that is without learners' conscious attention.

This model of collocation learning developed for L1 learners, according to Ellis (2003), works also for adult L2 learners. Wray (2002), however, claims that adult learners follow a non-formulaic approach to language learning, learning vocabulary as separate items and not retaining information about what words appear together. According to this view, when adult L2 learners are exposed to language input, they primarily notice not chunks but individual words. Although Wray also states that native speakers would note a string of formulaic utterances as a single item, second language learners, different from what Ellis believes, break it down and store the words separately without paying any attention to the fact that they appear together (p. 206).

Depending on which edge of this continuum they are, different researchers contend different reasons for the origins of collocational errors. Durant and Schmitt (2010), for example, believe that adult second language learners do retain language input as items that frequently go together. In their study, they selected some adjective-noun combinations from British National Corpus (BNC) and used them in three training sessions with 84 non-native speakers of English at the University of Nottingham. Participants in the first session were exposed to the target collocations for only one time, with those in the second session '4-3-2 minute talk' repetition strategy was used, and in the third one participants again repeated the collocations but each time in a different context. The training session was then followed by a test.

According to the test results, all the participants, even those who saw adjective-noun pairs only once in the training session, retained some memory of collocations. Therefore, they concluded that adult L2 learners "do, in contrast to Wray's claims, retain some memory of which words go together in the language they meet", a process which takes place implicitly without learners' paying attention to word pairs consciously (Durrant & Schmitt, 2010, p. 179). Learners should, then, pick up collocations as they are exposed to language input regardless of specific teaching or learning techniques. This, they contend, suggests that any deficiency in non-natives' collocational competence stems from lack of exposure to language rather than instructional approaches to vocabulary learning.

Keshavarz & Salimi (2007) looked into the relationship between collocational competence and cloze test performance, the latter being a measure of learners' overall proficiency level. With a test of 50 items designed to measure their participants' knowledge of collocations, they found that "learners' collocational competence and proficiency level are closely and positively associated", which, according to them, suggests that "proficient language users know a large number of collocational patterns" (p. 88). Their participants, Iranian university students majoring in English Language and Literature or English Translation categorized as being at the intermediate level based on a

TOEFL test administered as part of the study, were found to have insufficient command of English collocations.

The authors, relying on Iranian EFL teachers' experience, commented that this should be due to instructional approaches adopted. They concluded that "vocabulary teaching needs to be seen as being concerned not only with the meaning of words as semantic isolates, but as elements in semi-preconstructed phrases" (Keshavarz & Salimi, 2007, p. 89). The study, though implicitly, ascribes the poor collocational performance to how vocabulary is handled in classrooms, suggesting that words that go together should be taught together.

Another investigation with similar findings in that learners, regardless of how much progress they may have achieved in their linguistic proficiency, fail to reach a satisfactory collocational competence due to ineffective approaches to vocabulary teaching is a corpus-based study by Jean (2007), who reports an assessment of the collocational proficiency of students of English Linguistics at the University of Granada. Relying on three English corpora – the Bank of English, the BNC and the Longman Corpus Network, the study aims at measuring the collocational competence of 63 Spanish learners of English in 80 adjective-noun combinations. Jean found significant differences among participants' scores on receptive and productive collocational tests.

These findings also "indicate that students may fall short in the social and academic demands made on their command of L2", and this can be overcome through a pedagogical intervention program (Jean, 2007, p. 143). The researcher does not give details of the program, yet relates the deficiencies in learning formulaic language to instructional approaches, so input or L1 influence alone cannot account for poor collocational competence.

Zughoul & Abdul-Fettah (2001) studied the collocational competence of Arabic EFL learners studying at English department at a university. They used an Arabic word "kasara" – broke – to measure participants' receptive and productive skills in two individual tasks. One task depended on participants' success at recognizing the correct

English collocations equivalent to those of the verb “kasara”, and the other task was based on translation of the same collocations to measure subjects’ productive proficiency in collocations. The data analysis showed that Arabic learners’ overall proficiency in collocations is far from being satisfactory, the productive one being even worse. In their discussion on what might this poor performance originate from, the researchers came up with 11 distinct strategies mostly based on translation from NL to TL. Consequently they maintain that “on the whole, the study has subscribed to the role of NL in the FL acquisition”, indicating that it is the L1 transfer that accounts for the insufficient collocational competence even at advanced levels (p. 14).

Huang (2001) looked into 60 Taiwanese EFL students’ knowledge of collocations and their collocational errors using a Simple Completion Task which contained specific food and animal collocations or idioms. The scores were analyzed quantitatively in order to see the relative difficulty of different lexical categories, and qualitatively with a view to revealing “which words caused confusion in terms of collocability” and which collocations were the most challenging for subjects (p. 120). The test bore the most correct answers in free combinations and the least in pure idioms categories of collocations. A qualitative analysis of the test scores in terms of detecting the origins of the participants’ collocation errors indicated the influence of native language. For instance, “the subjects chose *eat* to collocate with *a bite*”, which is a direct translation from Chinese (p. 123). The participants also transferred cultural stereotypes to replace missing items in idioms. Where learners could not transfer negatively or positively, they provided their own alternatives, failing to recognize the idea of fixed expressions and collocational restrictions.

As a result of findings, Huang suggests that teachers, with a variety of examples, compare and contrast similar collocations in L1 and L2, and in this way “learners attend to the lexico-semantic distinctions between the two languages and reduce errors caused by L1 interference” (p. 125). The study concludes that EFL learners’ phraseological competence can be increased only by incorporating collocations into vocabulary teaching.

Nesselhauf (2003) investigated advanced learners's difficulty in collocations and the factors which might contribute to these difficulties relying on some verb-noun combinations selected from the German subcorpus of ICLE. Analyzing the incorrect combinations, Nesselhauf discusses whether it is the degree of collocational restriction that causes nonstandard use. The analysis revealed that it is not the degree of restriction but the role of L1 that leads to errors. Interference, she believes, plays a much more critical role in learners' insufficient production of collocations.

Indicating the maximum degree of first language effect on collocation production, Nesselhauf's (2003) study concludes that "collocations do deserve a place in language teaching" and that although rote learning turned into a passing fad with the fall of behaviorism, "it seems indispensable that a number of collocations be taught and learnt explicitly" through rote learning (p. 238). Considering the substantial influence of L1 on collocational competence, the study suggests that collocations to be taught be selected on the basis of native language, and this selection be taught with reference to L1.

The variation in the degree of L1 influence on incorrect collocation production was best put forward in a comparative study by Biskup (1992). Comparing the collocational performance of 28 German and 34 Polish learners of English, the study investigated to what extent L1 transfer was responsible for deviant collocations. Due to the perceived distance between Polish and English, Polish learners' errors could mostly be traced back to their first language, but for German students negative transfer failed to account for incorrect production. It is, then, rational to conclude that L1 does definitely have some effect on collocation production with varying degrees depending on each unique language, but cannot fully account for most of the errors in collocational performance.

To summarize, the actual differences in collocation studies relate to the origins of poor collocational competence of learners. The findings fall into three categories in this respect. While some researchers assert that this deficiency stems from ineffective instructional approaches to vocabulary teaching and lack of exposure to input, many

others assertively put the blame on negative transfer from L1. Consequently, wherever the collocational errors might be rooted, they certainly deserve more place in our language teaching programs, especially if we want our learners to produce more natural and fluent utterances in both speech and writing.

### III. METHODOLOGY

The present study based on natural occurring data from Turkish EFL learners probes into collocational competence using the procedures in Corpus Linguistics. Since research on collocations has proved that *verb + noun* type is the most troublesome structure for SLA learners, the study focuses on these lexical combinations produced by 11th graders at a public school. Collocations were extracted from students' writing exam papers that they took at the end of the academic year 2012-2013. Looking into semantic transparency, restriction on collocability and the L1 influence, the study attempts to contribute to the discussion on the difficulties EFL learners have in the production of lexical collocations.

Although, frequency, too, influences collocation use, it is deliberately left untouched in this study as there is a body of research concluding that collocations with such frequent verbs as *do, make, take, give* are among the most problematic for learners. That is mostly ascribed to their delexical characteristics. Namely, their original lexical value fades away and are used in other meanings and functions (Källkvist, 1998; Bonelli, 2000; Liu, 2000; Keshavarz & Salimi, 2007). This brings to the fore their nontransparent and unrestricted nature in collocability. Therefore, semantic transparency and restriction principles must apply to both more frequent and less frequent verbs. That is why the focus of this study is the three principles governing collocation production; restriction on collocability, semantic transparency and L1 influence.

#### 3.1. Participants

The 61 participants of the study are 11th graders at a selective state high school at 2012-2013 academic year. They have been studying English for 4 years including the preparatory class. They take three monthly exams and a writing exam every semester. The papers collected for this study are from the writing exam in the second semester. To test students' written performance they are given 40-minute writing exams in which they

are offered several topics to choose from. The topics are always related to issues covered in their course book, *Mission II* (Evans & Dooley, 2001). The course book has some reference to collocational learning, so students in general are aware of the importance of learning collocations.

### 3.2. Data

As mentioned earlier, the data for the study come from students' writing exam papers. Students' writing exam papers were built into a learner corpus. There are 76 discursive essays on three different topics which students have covered throughout the academic year. The essay topic students chose from 5 alternatives and the number of papers on each topic is as follows: 30 papers on "keeping animals in captivity", 30 on "the effects of video games on teenagers" and 16 on "whether every student should learn a foreign language at school". The corpus has 2212 word types and 21031 word tokens. The details about the size of the corpus are shown in Table 3.1 below.

**Table 3.1. Details of the corpus**

Topics	Number of essays	Types	Tokens	Average word count per essay
Animals in captivity	30	1228	8935	298
Video games	30	1238	7394	247
Learning a foreign language	16	837	4702	294

### 3.3. Procedures

As part of the writing exams applied at the school, 11<sup>th</sup> graders were given 5 topics related to the issues discussed in lessons and were asked to write an essay on one



within a class period of 40 minutes. They were not allowed to use their dictionaries or any other supplementary material during the exam. Below are the topics they were given:

1. People are increasingly becoming aware of endangered species. Authorities now accept that some animals are in danger of extinction. Some people argue that leaving such animals in their natural habitat will cause them to die out, hence be placed in zoos. However, there are also people who strongly claim that keeping them in zoos is not a fair treatment to animals. Discuss whether keeping animals in captivity is right or not.
2. Doing sports is associated with many benefits. One benefit, some people believe, is that participation in sports can keep teens out of trouble and help lower the increasing crime rate among the youth. However, there are also people who claim that it is just another factor that drag the young to violence because of the competition factor involved in most sports. Discuss whether sports can keep teenagers out of troubles or cause more troubles.
3. Homework is something almost all the teachers like to give and nearly all the students hate to do. Teachers believe it helps students practice what they have learned, and students think it does not help them learn but prevents them from doing extra studies. Do you believe homework is harmful or helpful?
4. Recently there is a growing tendency to learn a foreign language. Parents try to send their children to high quality schools where they can learn a foreign language effectively believing that this is an indispensable requirement for every individual. Do you think every student should learn a foreign language at school?
5. People are increasingly interested in the influence of video games on teenagers. Some parents are concerned that such games promote violence along with much other harm. However, there are also people who claim that video games offer a lot of learning opportunities for young people. Discuss whether video games help young people learn or lead to criminal acts.

The exam was applied by school teachers. The papers were computerized and saved as text documents. The texts were tagged for part of speech using Stanford Tagger (3.2.0), an automatic tool for assigning parts of speech for each word in a text. Designed by Stanford University Natural Language Processing Group, Stanford Tagger (3.2.0) proved to have 97.5% per-token accuracy (Manning, 2011). The tagged corpus was then uploaded in the AntConc 3.2.4w for analysis. Total number of tokens and word types were identified through the Word List tab in AntConc to find out the corpus size. The following regular expression syntax was used to extract *verb* + *noun* collocations with a maximum space of four words between the verb and its collocating noun.

$$\backslash w+_V\{Lu\}+\backslash s(\backslash w+_w+\backslash s)\{0,4\}\backslash w+_N\{Lu\}+$$

Besides, with another regular expression syntax ( $\backslash w+_J\{Lu\}+\backslash s\backslash w+_N\{Lu\}+$ ) *adjective* + *noun* combinations were also extracted as some present and past participles could appear as adjectives before nouns. The collocations were then classified according to restriction and transparency principles. They were grouped as either “Restricted” (R) if the noun collocated with 1-3 verbs or “Unrestricted” (UR) if it collocated with more than three verbs and “Transparent” (T) if both the verb and the noun were used in literal sense or “Semi-transparent” (ST) if one of the constituents was used in figurative sense. The collocation of *pay attention*, for example, was marked as R since *attention* collocates with a limited number of verbs to mean *give one’s attention to something*, as found out in the Corpus of Contemporary American English (COCA), and ST because one of its constituents *pay* is semantically non-transparent.

The reference corpus of COCA was used in deciding restrictedness of a combination. This has been much systematized in collocation studies dealing with the restrictedness issue (Martelli, 2007; Nesselhauf, 2005). However, the decision process of semantic transparency is a problematic area. Related studies do not report on how they classify word combinations as transparent or opaque. The general tendency is that a lexical combination is thought to be transparent if both constituents directly contribute to the overall meaning of the phrase. Frequency is generally accepted as the criterion when

deciding whether a constituent in a phrase directly contributes to the overall meaning or not (Bartsch, 2004). If a word has a sense different from its most frequent meanings, it is considered to be figurative. This is done through checking the dictionary entities of a word.

The next process was deciding whether the combinations extracted from the corpus were acceptable or not in terms of combinability with the intended meaning. First, each collocation was checked on the concordance plot to make up for the errors that may have been caused by the tagging process. The concordance lines also helped to decide what was intended by a specific phrase. Without knowing the intended meaning, it would not be possible to classify word combinations as acceptable (A) and unacceptable (UA). For instance, the phrase *to produce milk*, which normally exists in English as meaningful collocation, was considered unacceptable because the intended meaning was *to milk the cow*.

Several sources of reference were utilized in this decision process. The initial step was to check the word combinations in the electronic version of the Longman Dictionary of Contemporary English, which has a very rich phrase bank. Those combinations which did not appear in the dictionary were then searched in the Corpus of Contemporary American English (COCA) and British National Corpus (BNC). A minimum frequency criterion of 5 was set during the search procedure. If two words, out of millions of words, come together only 1-5 times, that should be considered as mere coincidence not as an acceptable collocation. For instance, *give* and *damage*, according to COCA search results, collocated only once. Therefore, such combinations that were used less than 5 times in the selected corpora were also left to the next stage. The final step was to refer to a native speaker. When a phrase did not appear in any of these three references, it was asked to an American man who had been employed as a native teacher at the school where the study was carried out.

The number of acceptable and unacceptable collocations, and relevant percentages were calculated. Then, wrong combinations were analyzed to answer the

research questions 1 and 2. The purpose here was to find out whether there were more collocational errors in R or UR, and T or ST type of collocations. This analysis bore out the areas of difficulty Turkish learners of English experienced in their collocational production. Finally, a qualitative analysis was performed to find an answer to the third research question. From concordance lines the sentences with erroneous collocations were taken to delve into the sources of errors. Students were asked to clarify their intention with these wrong collocations. Then, referring also to a teacher of Turkish Language and Literature employed at the school, Turkish equivalents of these combinations were found. Thus, the role of native language on collocational errors was identified.

## IV. RESULTS

The study aims to cast light on the areas of difficulty for L2 learners in their production of lexical collocations, basic determinants of a learner’s proximity to target language norms. In so doing, *verb + noun* collocations were extracted from the learner corpus developed out of Turkish EFL learners’ writing exam papers. The erroneous collocations were identified and then analyzed with respect to degrees of restriction on collocability of the constituents and semantic transparency. A further analysis was performed to find out to what extent the non-standard combinations stemmed from L1 influence.

The regular expression syntax typed in the concordance plot in AntConc 3.2.4w bore out 1915 hits. All the concordance lines were carefully studied and a total of 111 *verb + noun* combinations fitting into the collocation criteria defined for this research were found. A large proportion of these lexical bundles, that is 72.97%, proved to exist in the target language as acceptable collocations while the remaining 27.03% came out to not comply with the native speakers’ lexicon. 17.12% of the total collocations consisted of relatively less restricted constituents and 82.88% had more restriction on the combinability of their lexical items. Finally, while in 69.37% of the collocations both collocates were semantically transparent, in 30.63% one of the constituents was used in figurative sense.

**Table 4.1. A summary of the collocations extracted**

Acceptable		Unacceptable		Total
81		30		
Restricted	Unrestricted	Restricted	Unrestricted	
67	14	25	5	111
Transparent	Semi-transparent	Transparent	Semi-transparent	
58	23	19	11	

#### 4.1. The Degree of Restriction

Since the purpose is to detect the difficulties language learners have in their production of collocations, the focus of the analysis was on the erroneous utterances. In answering the first research question of the relationship between collocational production and the degree of restriction on the combinability of the constituting elements, the wrong combinations were classified either as restricted or as unrestricted.

In 16.77% of the 30 unacceptable collocations produced by learners the verbs can collocate with more than three verbs without a change in meaning. For each collocation, this information was always verified using COCA as the reference corpus. An example of this case from the corpus is *make empathy*. The noun *empathy* can combine with *develop*, *display*, *have* or *show*, but not with *make*. Therefore, it has a smaller degree of restriction.

A quick conclusion to be drawn from this finding is that Turkish EFL learners have a poorer collocational performance in restricted collocations where the combinability of the constituents alternate between 1 and 3, since the number of errors in restricted collocations – 83.33% - is much higher. However, this proportion is a direct result of the learners' general collocational performance as a large proportion of their lexical combinations fall into the restricted category. While 27.17% of the times the production in restricted combinations was erroneous, the ratio in non-standard production in unrestricted type was 26.32%, which is almost the same. As a result, restricted combinations seem to be little more difficult for L2 learners. The Table 4.2 below, taken out from Table 4.1, summarizes the results for restrictedness.

**Table 4.2. Restricted collocations**

	Acceptable	Unacceptable	Total
Restricted	67 (%72.83)	25 (%27.17)	92
Unrestricted	14 (%73.68)	5 (%26.32)	19

## 4.2. Semantic Transparency

When the unacceptable combinations were analyzed, it appeared that 36.77% of the time the meaning was semi-transparent, that is one of the constituents was used in figurative sense. For instance, *attend language life* in the corpus was actually used to mean *pay attention ...*, in which *pay*, despite contributing to the overall meaning of the collocation, does not have its literal sense. The data included 19 transparent combinations – 63.33% out of 30 erroneous collocations. These results seem to imply that learners produced less non-standard lexical bundles in transparent type of collocations. Table 4.3 below, extracted from Table 4.1 shows the results for semantic transparency.

**Table 4.3. Semantic transparency of the collocations**

	Acceptable	Unacceptable	Total
Transparent	58 (%75.32)	19 (%24.68)	77
Semi-transparent	23 (%67.65)	11 (%32.35)	34

A further analysis, however, is required for verification. A scrutiny of acceptability in proportion to semantic transparency should give more dependable results. Out of the semantically transparent combinations 24.68% is unacceptable and 75.32% is acceptable. The rates for semi-transparent combinations are 32.35% and 67.65% respectively. These findings suggest that the possibility of L2 learners' making an error is higher when one of the constituting elements in a collocation is semantically opaque. Namely, semi-transparent combinations pose greater difficulty for language learners.

## 4.3. L1 Influence in Collocational Production

In answering the third research question, which is the role of mother tongue on learners' production of lexical combinations, the wrong collocations were analyzed to find out the source of error in each. The following table shows the non-standard

utterances that have some kind of L1 origin. Out of the total 30 erroneous combinations, 23.33% seemed to have been caused by the first language. Therefore, it sounds rational to conclude that L1 does not play a very critical role in learners' erroneous production, since more than 75% of the times the non-standard form was influenced by some intralingual strategies other than L1 transfer.

**Table 4.4. Errors with L1 origin**

	<b>Used</b>	<b>Intended</b>
a	display foreign language	to teach a foreign language
b	follow technology	to catch/keep up with technology
	follow world	to catch/keep up with the world
c	give damage	to do / cause damage
d	hold animals (in zoo)	to keep animals (in zoo)
e	make addiction	to cause addiction
f	turn animals (to money)	to make money (out of animals)

Following are detailed analysis of the examples in the table as they are originally used in the learner corpus.

(a) *Firstly, we should **display foreign language** to every child ...*

As the context implies, student's intention here is *to teach a foreign language*. Though rare, it is possible to hear in Turkish the verb *göstermek* (*display*) used to mean *to teach a subject*.

(b.1) *People have to get information about technology from these countries to **follow the technology**.*

(b.2) *At last the need for foreign language is rising every day and children should learn it at very young age in order to improve their language and **follow the world**.*



The acceptable collocation in the above examples should be *to catch/keep up with technology/world*. The errors in these statements reflect a very common usage in Turkish. The verb *takip etmek* (*follow*) often collocates with *technology* or *the world*, so the learners here apparently transferred the verb *takip etmek* (*follow*) to mean *keep/catch up with technology/the world*.

(c) *But people use them for their purposes and **give damage** to them (them refers to endangered species).*

This is one of the most ostensible instances of interlingual errors in collocational production. Turkish abounds in cases where the noun *zarar* (*damage*) comes together with the verb *vermek* (*give*) to produce the phrase of *zarar vermek* (*give damage*). Yet, in English, the original collocate of *damage* should be *cause* or *do*, hence the combination of *cause/do damage*.

(d) *Keeping animals in captivity I think is true because if we **hold animals in zoo**, ...*

The verb *tutmak* (*hold*) in Turkish can also be used in sentences where the meaning is *to keep*. As a result of negative transfer from Turkish *tutmak* (*hold*), the learner produced the wrong combination of *hold animals in zoo*.

(e) *On the other hand, some type of games **make addiction** so ...*

The mistake in the statement above was caused by the collocation of *bağımlılık yapmak* in Turkish. The noun *bağımlılık* (*addiction*) and the verb *yapmak* (*make*) go together in cases when something such as video games *causes addiction*.

(f) *Finally, people **turn animals to money**.*

The incorrect collocation above was used to suggest that by keeping animals in zoos, people try to *make money out of animals*. The phrase of *hayvanları paraya dönüştürmek* (*turn animals to money*) is not an acceptable pre-fabricated utterance in English.

There are some other erroneous combinations which prove that L1 transfer does not play a critical role in collocational production. One such example is the wrong collocation of *build commerce* as used in the statement below.

*For instance, if an entrepreneur speaks Chinese, he/she will **build commerce** between China and his country.*

In Turkish the noun *ticaret* (*commerce*) collocates with *kurmak* (*establish*) to form the phrase *ticaret kurmak* (*establish commerce*). If the mother tongue had been in effect, the student could have produced the acceptable collocation of *to establish commerce*. The following sentence is another representation of such a case.

*This is very bad action. We should **make empathy**.*

In Turkish the correct collocate of *empati* (*empathy*) is *göstermek* (*show*) not *yapmak* (*make*). Transferring from L1 then, the learner should have produced the correct utterance of *show empathy*. In this incorrect combination, therefore, an interlingual problem exists. There is this tendency to overuse such common verbs as *do* and *make* in collocations. The same problem exists in the following example, as well.

*Firstly, people **did fails** but now people can help them.*

In the above sentence, the student is trying to bring to light how attempts to protect endangered species by keeping them in zoos *ended in failure*. Lack of knowledge of this collocation ended in using a common strategy, overuse of *do*. In addition to overuse, synonymy is another strategy employed by L2 learners. The table below shows examples where the verbs and, in one case the noun, are substituted with their synonyms, resulting in incorrect combination.

**Table 4.5. Examples of synonymy as a learner strategy**

Used	Intended
get amusement	get fun
grow children	raise children
make hunting	do hunting
obtain balance	maintain balance

Following statement exemplifies another strategy, circumlocution, adopted by L2 learners. The student seemingly does not know, or more probably cannot remember, the word *favor*. In order to make up for this language barrier, he tries to describe what he intends to say.

*... and you **do a nice action** for world.*

As a result, these findings suggest that there are different strategies, besides transfer, involved in the process of collocational production. Considering all such strategies as overuse of generic verbs, synonymy and circumlocution, L1 is not the only factor to be blamed for second language learners' incorrect production of collocations.

## V. DISCUSSION

The present study suggests that combinations with a higher degree of restriction on the constituting elements seem to appear more frequently in the learner corpus built of Turkish EFL learners' writing exam papers. While 72.97% of the total 111 collocations extracted from the corpus proved to exist in the references selected for the study, 27.03% of the times students produced incorrect combinations. Collocations with less restriction and semantically more transparent meaning were used more often than unrestricted and semi-transparent ones. The possibility of producing erroneous combinations turned out to be almost the same for restricted and unrestricted type of collocations, and higher for semi-transparent collocations than combinations with both constituents being totally transparent. As for the L1 influence, the data demonstrated that interlingual transfer does not play a very big role in learners' incorrect production of collocations.

In collocational studies, it is unfortunately hard to make comparisons since the methodologies employed, sizes of corpora, and procedures followed often show great distinctions. As summarized in literary review section, different studies delve into different aspects of collocations. While some studies are interested in on certain types of collocations such as *verb + noun* combinations, some others analyze all the pre-fabricated utterances including the grammatical combinations, as well. There are differences in terms of methodology, too. Some studies depend on learners' errors and some on correct use to reach a conclusion on collocational performance.

This difficulty is particularly pervasive in terms of the restriction rule on collocability and semantic transparency principle. There are two very comprehensive studies in the literature that offer some comparability as the methodology adopted in this study share some similarity. In one of these studies Nesselhauf (2005) studied the German part of ICLE and, with respect to the restriction principle, concluded that higher degree of restriction facilitate L2 learners' collocational production.

The difficulty with lower degree of restriction on combinability is believed to be caused by the fact that restricted collocations are perceived and retained as single items whereas unrestricted ones are learned separately (Howarth, 1998; Nesselhauf, 2005). This is speculated to originate from the polysemous nature of certain verbs like *make*, *do* and *get* (Källkvist, 1998). One final explanation is that learners pay more attention to set phrases more than those that combine more freely (Nesselhauf, 2005).

In the present research, however, there appeared to be more errors and slightly higher possibility of producing incorrect collocations in restricted type of collocations. Therefore, the above explanations fail to account for the findings in this paper, although they still may shed light on the origins of the incorrect production learners made. This difference may be due to the data analysis techniques and the size of the data. In Nesselhauf's (2005) study, for instance, the unrestricted collocations outnumber the restricted ones – 1496 and 322 respectively, which increases the likelihood of erroneous production in unrestricted combinations. While 82.29% of all the collocations fall into the unrestricted type in Nesselhauf's research, this is just the opposite – 82.88% belong to the restricted category – in this study.

The other all-inclusive study comparable to this one in terms of methodology (Martelli, 2007), reports similar findings. Martelli extracted all the incorrect combinations from the Italian part of ICLE. The study concludes that Italian learners have more difficulty in highly restricted combinations. This finding also verifies the previous speculation that the details of the data may affect the results. In Martelli's data, the number of restricted type of combinations is way over the unrestricted type. This may lead to more cases of incorrect production in phrases where a higher degree of restriction applies.

As regards the relationship between collocational performance and semantic transparency principle, the present study found that semi-transparent collocations pose greater difficulty for L2 learners, which contradicts to what much of the literature reported earlier. Both Nesselhauf (2005) and Martelli (2007) assert that when the

meaning is transparent, learners produce more erroneous phrases. They attribute this to the fact that transparent items receive less attention from the learners, because they are perceived to be easier to learn.

There are also studies which concluded, similar to the present paper, that transparent meaning facilitates learners' collocational production. Bonelli (2000) and Keshavarz and Salimi (2007), for instance, also claim that a dearth of lexical transparency causes greater difficulty in learner production. The general belief is that collocations with a figurative meaning are processed as single items and learned as whole. Thus, they contribute more to correct production. The contradiction in this study should relate to the instructional methodology employed in the course book used. Students' course book, Mission 2 (Evans & Dooley, 2001), abounds in activities dedicated to teaching collocations. Since the writing topics used in the study are from the lessons in the book, students must have grasped relevant collocations, whether transparent or semi-transparent, and correctly used many of them in their essays.

Differences in findings in respect to restriction and transparency may also stem from differences in L1 backgrounds of the subjects. The existence of one-to-one corresponding collocations may affect performance with regard to collocability principles. If the directly translated equivalent of a semi-transparent collocation in L1 transfers to L2 in its figurative meaning, this will again contribute to better collocational performance. For instance, *kill some time* was used correctly in the corpus probably because *öldürmek* (*kill*) is used in the same figurative sense in Turkish, too. The number of such occurrences in different learner corpus will definitely influence the results.

When it comes to L1 effect on collocation production, this paper suggest with 23.33% of the errors reflecting some interlingual interference that the role of L1 on incorrect collocations is not very critical. This ratio is 14.7% in Martelli's (2007) corpus of Italian learners, indicating a minimum L1 interference, and 53% in Nesselhauf's (2005) study with German learners of English, implying a maximum L1 interference. Therefore, as regards the effect of mother tongue, findings are contradictory.

Since each language is unique, their influence, too, on a second language learning process should be unique. This is most apparent in a much earlier study by Biskup (1992). In this comparative study with 28 German and 38 Polish students, it was found that while 21% of the time German impaired collocation production in English, Polish had a 46% influence on incorrect combinations. Another factor that affects the varying degrees of L1 interference is the diversity in the participants' level of the target language. At lower levels, students may resort more often to their mother tongue, and as the amount of input they receive increases, reference to L1 may wane. It can be concluded, for this study, that the role Turkish plays on collocation production is weak.

Considering the ultimate goal of language teaching programs or of personal attempts to learn a second language is to approach native-like fluency and accuracy, it is clear that without a comprehensive knowledge of collocations, this goal is unattainable. Actually, this is not a recently discovered fact. Palmer (1933, as cited in Wua et al., 2010) accepted collocations as a prerequisite for reaching a satisfactory linguistic competence. Brown (1974, as cited in Wua et al., 2010), early in seventies, strongly asserted that in addition to vocabulary development, collocational performance can boost listening comprehension and reading speed. This explains why non-natives try to catch every single words they hear while listening to something, but the natives, as they have piles of word combinations at their disposal, anticipate what should come after a certain word (Lewis, 2000). Collocational performance, therefore, proves to be an area that needs attention of learners, teachers and SLA researchers.

## VI. CONCLUSIONS

The present study concludes that there are three basic factors that may cause problems in LLS' collocational performance, which are restriction on commutability, semantic transparency and L1 interference. Restriction principle, according to the findings of this study, does not seem to play a major part in Turkish learners' incorrect collocation production. Yet, they have more problems in producing correct collocations when the combination has a semi-transparent lexical item. Besides, Turkish has a minor role in erroneous collocation production.

### 6.1. Implications and Applications

The factors influencing collocational competence, instructional methodologies, problems learners face in learning collocations or the origins of collocational errors may have been expressed in variable conceptualization and wording by different researchers, yet the fact that learning collocations is significant and that the basic difference between native and nonnative speakers of English is their control of multi-word units is invariable. Therefore, it is of particular importance that SLA learners need to develop competence in pre-fabricated utterances if they want to reach a native-like level in English. Indeed, Wray (2002) and Nesselhauf (2003) allegedly suggest that knowledge of collocations enable language learners to be far more accurate and fluent.

One trouble in collocation teaching is the limitless number of word combinations. Which collocations to teach is a difficult question to answer. Researchers prioritize different aspects of collocations in answering this question. However, it appears that collocations are hard to deal with in educational settings. First of all, linguistics wise, the structure of collocations is one obvious reason. As each word combines with another arbitrarily, SLA learners fail to recognize restriction principle. This research presents instances where students use synonymous verbs to collocate with a noun. It is, therefore, of critical significance to raise awareness about this issue. Considering the impracticality



of teaching almost limitless number of word combinations, it is rational to teach the principles of restriction and semantic transparency, which supposedly will push learners closer to more natural collocation production.

Frequency can be taken into consideration to overcome the problem of what collocations to teach out of the countless number of multiword units. The most frequent verbs, for instance, can be extracted from different corpora and vocabulary instruction can be based on teaching the noun that collocate with these verbs. However, there are studies asserting that learners have poorer incorrect collocation production with the most frequent verbs due mainly to the delexicalized nature of such verbs losing their transparency and learners' failure to notice the restrictions in combinability (Källkvist, 1998; Bonelli, 2000; Liu, 2000; Keshavarz & Salimi, 2007). Therefore, it seems more plausible to teach the principles governing the collocation use.

From a more pedagogical perspective, the problem with collocations relates to learners' dependence on their mother tongue in language production. Although this research does not directly blame the incorrect collocation production on L1 transfer, it does have some effect to some extent as found in this study, and to a large extent as found in the most comprehensive collocation research (Nesselhauf, 2005). It is again a more practical option to make learners aware of the fact that word combinations do not necessarily have one-to-one correspondence across languages. Once students realize the arbitrary nature of commutability, they will improve in collocation production.

One final explanation for the apparent difficulty of collocations comes from psycholinguistics. Expert speakers of English and SLA learners sharply differ in the way they organize and retain vocabulary. Although natives do this in chunks, non-natives store vocabulary as individual items (Lewis, 2000; Wray, 2002). Thus, they tend to create their own combinations haphazardly bringing together different words that might actually not coexist in the target language. Psycholinguistic perspective of the issue again refer to the principles of collocations. Hence, learners need to be enlightened to the fact that

certain restriction rules apply in combinability of different words, and that this restriction is most of the time not semantic but arbitrary.

Apparently, although the situation has been changing recently, collocation teaching has long been ignored in language classrooms due most probably to the reasons mentioned above. There have been two opposing opinions on how to deal with collocations in classrooms. The implicit – explicit teaching dichotomy applies also to collocation teaching (Wray, 2002; Ellis, 2003). Whatever side of this argument may be supported, the truth is that teaching collocations complicate and increase already heavy teachers' work load simply because of practicality. Nonetheless, there are still things to be done by teachers. The most important thing, as it appears in this research and some other relevant work, learners lack an understanding of the linguistic nature of collocations. They should be enlightened about the restrictions governing the co-occurrence of words in English.

One solution is the use of collocation dictionaries that are increasingly becoming popular. There are very comprehensive dictionaries in which each word combination is presented as separate entities. Lien (2003) studied such dictionaries and demonstrated how they can not only increase learners' awareness of collocations but also lead to correct collocation production. Encouraging the use of such dictionaries, learners can develop a true understanding of the construct of collocations, and be guided through the right pedagogical paths. Namely, they can eventually be convinced that due mainly to their arbitrary and sometimes nontransparent nature, collocations should be processed and stored as single vocabulary items. Laufer (2011) elucidated how dictionary use can contribute to successful collocation production by language learners.

Another assistance can come from computational linguistics, which has recently, been gaining popularity as an instructional tool in language teaching. Although primarily designed for researchers, concordancers can help reduce teachers' burden. With thousands of texts brought together concordancers could offer fun-to-study data for learners especially at higher levels. It is widely attested today that “corpus descriptions

have also enhanced our understandings of units of fixed phrasing, collocation, and more extended language patterns” (O’Keeffe et al., 2007, p. 23).

Thanks to online concordancers simple enough for language learners to work on, students can “get hands-on experience of using a corpus through guided tasks or through materials based on corpus evidence, such as concordance lines on handouts” (O’Keeffe et al., 2007, p. 24). Referred to, in literature, as “data-driven learning”, this new approach has the capacity to “make the learner a linguistic researcher” (Johns, 2002, p. 108). It is again teachers’ responsibility to familiarize learners with such tools by carefully selecting or designing concordance-based exercises.

This careful selection might again be a daunting task for teachers. A digital library, “collection of texts (although it can also contain other resources including images, sound files, etc.), and can function as a searchable corpus” can eliminate this burden by providing “authentic, focused material that is carefully selected and organized, exposing learners to contemporary language usage” (Wua et al., 2010, p. 91). Wua et al. describe four different activities based on a digital library automatically extracting lexical collocations. For those interested, Witten et al. (2010) meticulously described the steps to build a digital library.

## **6.2. Suggestions for Further Research**

Collocation studies follow two different approaches. Using elicitation techniques, researchers can contain the collocations to work on, such as the ones with the most frequent verbs, and carry out a more to-the-point study. This methodology would be weak in generalization of the results. With corpora, however, naturally occurring data can be studied to suggest more generalizable findings. Since corpus linguistics tools have recently been used in SLA studies, more corpus research into collocations is needed.

Furthermore, the already existing work on collocations deal with different aspects of word combinations, leaving the restriction principle of collocability and semantic

transparency open for further studies. Therefore, this gap offers promising opportunities for SLA researchers. There seems to be some focus on erroneous collocation production, so factors leading up to correct production of collocations can be analyzed. Overspread of online concordancers also pave the way for action research. Practitioners in the field can readily determine a set of collocations tailored to the needs of their specific students and design motivating materials dedicated to teaching collocations. It is noteworthy here that both researchers and teachers who are especially new to such issues should develop some competence in corpus linguistics tools like concordancers before they delve into the field.

Corpus-based collocation research is a much deeper gap in Turkish EFL setting. As this study is limited to an approximately 21.000-word corpus, more dependable and generalizable results can be obtained with a larger size corpus. This opens up another deep gap, which is the need to develop a learner corpus of Turkish students. As each language is unique, so is their influence on second language learning processes. It is, therefore, highly advisable to detect the difficulties in collocation production specific to Turkish learners.

Future studies may address to the problem of insufficient comparable data in the related literature. It has been observed that research into collocations does not offer comparable data as the methodologies employed or the corpus sizes developed in each study differ. Those well-established concordance-based studies in the field need to be replicated using subjects from different linguistic backgrounds so that the findings can be more generalizable.

To conclude, collocations do deserve attention of SLA researchers, language teachers and L2 learners, as well. It is true that a language learner cannot possibly store as many prefabricated utterances as a proficient user of English. Nevertheless, once learners' awareness of the principles governing collocation use is increased, they will hopefully begin to produce more standard lexical combinations.

## REFERENCES

- Biskup, D. (1992). L1 influence on learners' renderings of English collocations: A Polish/German empirical study. In P. Arnaud, & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 85-93). London: Macmillan.
- Bonelli, E. T. (2000). Corpus Classroom Currency. *Darbai ir Dienos*, 205-243.
- Brown, D. F. (1974). Advanced vocabulary teaching: The problem of collocation. *RELC Journal*, 2(5), pp. 1-11.
- Durrant, P. & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, 26(2), 163-188.
- Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty, & M. Long (Eds.), *The Handbook of Second Language Acquisition*. Oxford: Blackwell.
- Evans, V. & Dooley, J. (2001). *Mission 2*. Berkshire: Express Publishing.
- Firth, J. R. (1957). *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Futagia, Y., Deane, P., Chodorow, M. & Tetreault, J. (2008, October). A Computational Approach to Detecting Collocation Errors in the Writing of Non-Native Speakers of English. *Computer Assisted Language Learning*, 21(4), pp. 353-367.
- Granger, S. (1998). Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae. In *Phraseology: Theory, Analysis and Applications* (pp. 145-160). Oxford: Oxford University Press.
- Hill, J. (2000). Revising Priorities: From Grammatical Failure to Collocational Success. In M. Lewis (Ed.), *Teaching Collocation: Further Developments in the Lexical Approach* (pp. 49-50). London: Language Teaching Publications.
- Howarth, P. (1998). Phraseology and Second Language Proficiency. *Applied Linguistics*, 19, pp. 24-44.
- Huang, L.S. (2001). Knowledge of English collocations: An analysis of Taiwanese EFL learners. *Texas Papers in Foreign Language Education*, 114-132.

- Jean, M. (2007). A corpus driven design of a test for assessing the ESL collocational competence of University Students. *International Journal of English Studies*, 7(2), 127-147.
- Johns, T. (2002). Data-driven learning: The perpetual challenge. In B. Ketteman, & G. Marko (Eds.), *Teaching and Learning by Doing Corpus Linguistics* (pp. 107-117). Amsterdam: Rodopi.
- Källkvist, M. (1998). Lexical Infelicity in English: The Case of Nouns and Verbs. In K. Haastrup, & A. Viberg (Eds.), *Perspectives on Lexical Acquisition in a Second Language* (pp. 149-174). Lund University Press.
- Keshavarz, M. & Salimi, H. (2007). Collocational competence and cloze test performance: A study of Iranian ELF learners. *International Journal of Applied Linguistics*, 17(1), 81-92.
- Laufer, B. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, 24(1), pp. 29-49.
- Laufer, B. & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning*, pp. 647-672.
- Leśniewska, J. (2006). Collocations and Second Language Use. *Studia Linguistica*, 95-105.
- Lewis, M. (1997). *Implementing the Lexical Approach: Putting Theory into Practice*. Hove: Language Teaching Publications.
- Lewis, M. (2000). *Teaching Collocation: Further Developments in the Lexical Approach*. (M. Lewis, Ed.) Hove, Language Teaching Publications: Language Teaching Publications.
- Li, C.C. (2005). *A Study of Collocational Error Types in EFL/ESL College Learners' Writing*. Retrieved December 22, 2012, from [ethesis.lib.mcu.edu.tw/ETD.../getfile?](http://ethesis.lib.mcu.edu.tw/ETD.../getfile?)
- Lien, H. Y. (2003). The effects of collocation instruction on the reading comprehension of Taiwanese college students. *Doctoral Dissertation*. Indiana University of Pennsylvania.

- Liu, C. P. (2000). An Empirical Study of Collocation Teaching. *The Proceedings of the Seventeenth Conference on English Teaching and Learning in the Republic of China* (pp. 165-178). Taipei: Crane.
- Lozano, C. & Mendikoetxea, A. (2013). Learner corpora and second language acquisition: The design and collection of CODEL2. Retrieved April 7, 2013, from <http://wdb.ugr.es/~cristobalozano/wp-content/uploads/Lozano-and-Mendikoetxea-2012-CEDEL2-versi%C3%B3n-p%C3%A1gina-web.pdf>
- Manning, C. D. (2011). Part-of-speech tagging from 97% to 100%: is it time for some linguistics? *CICLing'11 Proceedings of the 12th International Conference on Computational Linguistics and Intelligent Text Processing*, (pp. 171-189).
- Martelli, A. (2007). *Lexical Collocations in Learner English: A Corpus-Based Approach*. Edizioni dell'Orso.
- McCarty, M. & Carter, R. (2001). Ten criteria for a spoken grammar. In E. Hinkel, & S. Fotos (Eds.), *New Perspectives on Grammar Teaching in Second Language Classroom* (pp. 51-75). Mahwah, NJ: Lawrence Erlbaum Associates.
- Myles, F. (2005). Interlanguage corpora and second language acquisition research. *Second Language Research*, 4(21), pp. 373-391.
- Nation, I. S. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.
- Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Philadelphia: John Benjamins Publishing Co.
- O'Keeffe, A. McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- Orenha-Ottaiano, A. (2012). *English Collocations Extracted from a Corpus of University Learners and Its Contribution to a Language Teaching Pedagogy*. Retrieved March 23, 2013, from [www.periodicos.uem.br/ojs/index.php/ActaSciLangCult/article/view/17130/pdf](http://www.periodicos.uem.br/ojs/index.php/ActaSciLangCult/article/view/17130/pdf)
- Palmer, H. E. (1933). *Second Interm Report in English Collocations*. Tokyo: Kaitakusha.

- Pecina, P. (2010). Lexical Association Measures and Collocation Extraction. *Language Resources & Evaluation*, pp. 137-158.
- Saif, A. M. & Aziz, M. J. (2011). An Automated Collocation Extraction from Arabic Corpus. *Journal of Computer Science*, 7(1), pp. 6-11.
- Schmitt, N. (Ed.). (2004). *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins.
- Shin, D. & Nation, P. (2008, October). Beyond Single Words: The Most Frequent Collocations in Spoken English. *ELT Journal*, 62(4), pp. 339-348.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Waller, T. (1993). Characteristics of near-native proficiency in writing. In H. Ringbom (Ed.), *Near-Native Proficiency in English* (pp. 183-293). Abo Academy University: English Department Publications.
- Webb, S. & Kagimoto, E. (2010). Learning Collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics*, 32(3), pp. 259-276.
- Wei, Y. (1999). *Teaching Collocations for Productive Vocabulary Development*. Retrieved April 17, 2012, from ERIC: [eric.ed.gov/?id=ED457690](http://eric.ed.gov/?id=ED457690)
- Witten, I. H., Bainbridge, D., & Nichols, D. M. (2010). *How to Build a Digital Library*. Burlington, MA: Morgan Kaufmann.
- Wray, A. (2000). Formulaic Sequences in Second Language Teaching: Principle and Practice. *Applied Linguistics*, 21(4), 463-489.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wua, S., Franken, M., & Witten, I. H. (2010). Supporting collocation learning with a digital library. *Computer Assisted Language Learning*, 23(1), pp. 87-110.
- Zughoul, M. & Abdul-Fettah, H. (2001). Collocational Competence of Arabic Speaking Learners of English: A study in Lexical Semantics. *Journal of the College of Arts*, 1-19.



## APPENDIX: LIST OF INCORRECT COLLOCATIONS

Used	Intended
attend language life	to pay attention to the language they use
build commerce	to establish/create commerce
catch the world	to catch up with the world
catch the improvement of the world	to catch up with the improvement
did fails	to end in failure
display benefits of learning a foreign language	to show/demonstrate the benefits
display foreign language to every child	to teach a foreign language
do a nice action for world	to do a favor
follow technology	to catch/keep up with technology
follow world	to catch/keep up with the world
get character	to get into a character
get amusement	to get fun/to have fun
get addiction	to develop an addiction
give damage	to do / cause damage
grow children	to raise children
have a life	lead a life
hold animals in zoo	to keep animals in zoo
hold the question with different aspects	to handle the problem
law their opinion	their opinion can be made a law
listen a language	hear a language
(games) make addiction	to cause addiction
make empathy	to develop/ display/ have/show empathy
make hunting	to do hunting
obtain ecological balance	to maintain balance
play a lot of time in the computer	to spend time
produce milk	to get milk
produce reasons	to give the reason
publish a zoo	to establish a zoo
publish a safari park	to establish a park
turn animals to money	to make money out of animals