

MAJOR SOURCES OF COLLOCATIONAL ERRORS MADE BY EFL LEARNERS

AT KOYA UNIVERSITY

A Master's Thesis

by

HAWRAZ Q. HAMA

The Department of

Teaching English as a Foreign Language

Bilkent University

Ankara

June 2010

*All that I am or ever hope to be, I owe to her;*

*my beloved mother*

MAJOR SOURCES OF COLLOCATIONAL ERRORS MADE BY EFL  
LEARNERS AT KOYA UNIVERSITY

The Graduate School of Education

of

Bilkent University

by

HAWRAZ Q. HAMA

In Partial Fulfillment of the Requirements for the Degree of

MASTERS OF ARTS

in

The Department of

Teaching English as a Foreign Language

Bilkent University

Ankara

June 2010

BİLKENT UNIVERSITY  
GRADUATE SCHOOL OF EDUCATION  
MA THESIS EXAMINATION RESULT FORM

June 29, 2010

The examining committee appointed by the Graduate School of Education  
for the thesis examination of the MA TEFL student

Hawraz Q. Hama

has read the thesis of the student.

The committee has decided that the thesis of the student is satisfactory.

Thesis title: Major Sources of Collocational Errors Made by  
EFL Learners at Koya University

Thesis Advisor: Asst. Prof. Dr. Philip Lee Durrant  
Bilkent University, MA TEFL Program

Committee Members: Asst. Prof. Dr. JoDee Walters  
Bilkent University, MA TEFL Program

Dr. İsmail Boztaş  
Bilkent University, Translation and Interpretation

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Teaching English as a Second Language.

---

(Asst. Prof. Dr. Philip Lee Durrant)  
Supervisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Teaching English as a Second Language.

---

(Asst. Prof. Dr. JoDee Walters)  
Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Teaching English as a Second Language.

---

(Dr. İsmail Boztaş)  
Examining Committee Member

Approval of the Graduate School of Education

---

(Vis. Prof. Dr. Margaret Sands)  
Director

## ABSTRACT

MAJOR SOURCES OF COLLOCATIONAL ERRORS MADE BY EFL  
LEARNERS AT KOYA UNIVERSITY

Hama, Hawraz Q.

M.A., Department of Teaching English as a Foreign Language

Supervisor: Asst. Prof. Dr. Philip Lee Durrant

June 2010

The aim of the present study was to explore the main sources of collocational errors made by learners of English as Foreign Language (EFL). To address this issue, 40 Kurdish seniors studying EFL at Koya University's College of Languages located in Northern Iraq participated in this study. Quantitative data were obtained from the collocation completion test used to explore the main sources of collocational errors made by the participants. Qualitative data were obtained from think-aloud protocols aimed to find out possible main source(s) of collocational errors.

The results showed that the participants' collocational errors resulted from two major sources, namely, low frequency of collocations and the influence of L1. Factors such as the frequency of collocation components and Mutual Information (MI) were found to be ineffective in the production of correct collocations because these factors did not cause errors in collocations.

Finally, implications of these results for teaching are discussed. Additionally, suggestions were made for ways in which researchers and materials designers could provide better language teaching materials with respect to collocations taking into account major factors that often cause difficulty in collocations.

Key terms: Collocation, Error, Frequency, Mutual Information, variables, and English as a Foreign Language (EFL).

## ÖZET

### KOYA ÜNİVERSİTESİ'NDE YABANCI DİL OLARAK İNGİLİZCE ÖĞRENEN ÖĞRENCİLER TARAFINDAN YAPILAN EŞDİZİMLİLİK HATALARININ TEMEL KAYNAKLARI

Hama, Hawraz Q.

Yüksek Lisans, Yabancı Dil Olarak İngilizce Öğretimi Bölümü

Tez Yöneticisi: Yrd. Doç. Dr. Philip Lee Durrant

Haziran 2010

Bu çalışmanın amacı, İngilizceyi Yabancı Dil olarak öğrenenler tarafından yapılan eşdizimlilik hatalarının ana kaynaklarını araştırmaktır. Bu konuyu değerlendirmek amacıyla, Kuzey Irak'ta bulunan Koya Üniversitesi Yabancı Diller Okulu, İngilizce Bölümünden 40 Kürt son sınıf öğrencisi söz konusu araştırmaya katılmıştır. Eşdizimlilik tamamlama testinden elde edilen sayısal veriler katılımcılar tarafından yapılan eşdizimlilik hatalarının ana kaynaklarını araştırmak için kullanılmıştır. Sesli-düşünme tutanaklarından elde edilen nitel veriler eşdizimlilik hatalarının ana kaynaklarını ortaya çıkarmayı amaçlamaktadır.

Sonuçların gösterdiğine göre, katılımcıların eşdizimlilik hataları tam olarak iki nedenden kaynaklanmaktadır; eşdizimliliğin düşük tekrarı ve ana dilin etkisi. Eşdizimliliğin az tekrar etmesinin bileşenleri ve Karşılıklı Bilgi Edinme gibi sebepler



dođru eřdizimliliđin oluřumu sırasında etkisiz olarak grld nk bu faktrler eřdizimlilikte herhangi bir hataya sebep olmadılar.

Neticede, bu sonuların đrenim aısından olan anlamları tartıřıldı. Buna ek olarak, eřdizimlilikte sıka glklere sebep olan ana faktrleri gznne alarak , arařtırmacılar ve materyal geliřtiricilere daha iyi dil đretim materyallerin hazırlanmasını sađlayacak neriler yapıldı.

Anahtar Kelimeler: Eřdizimlilik, Hata, Sıklık, Karřılıklı Bilgi Edinme (MI), ve Yabancı Dil olarak İngilizce.

## ACKNOWLEDGEMENTS

I owe several debts of gratitude to sainted Professor “Ihsan Dođramaci”, God rest his soul, for giving me this opportunity to study in Bilkent University. I would like also to assert my appreciation to Dr. Rasim Özyürek for his continuous support.

I would like to express my profound gratitude to my thesis advisor, Asst. Prof. Dr. Philip Lee Durrant, whose rigorous academic coaching and endless assistance enabled me to complete this thesis successfully. Without his invaluable instructions and information this thesis would have been far weaker.

My genuine thanks are also extended to the head of the MA-TEFL Department, Asst. Prof. Dr. Julie Mathews-Aydınlı, who accepted me as an MA-TEFL student. I also owe thanks to Asst. Prof. Dr. JoDee Walters and Prof. Dr. Kim Trimble for their continuous academic instructions.

My appreciation goes to the head of the Department of English Language and Literature in Koya University, Dr. Hoshang Farooq, for his inspiration and constant assistance. I would like also to express my special thanks to the participant students of this study for their willingness to help me with my research.

Many thanks also go to my classmates for their warm friendship and assistance during the years of my studying at Bilkent University.

Last but not least, I would like to acknowledge my family specifically my parents for their generous encouragement and sponsorship.

## TABLE OF CONTENTS

ABSTRACT .....	iv
ÖZET.....	vi
ACKNOWLEDGEMENTS .....	viii
TABLE OF CONTENTS .....	ix
LIST OF TABLES .....	xii
LIST OF FIGURES.....	xiii
CHAPTER I: INTRODUCTION .....	1
Introduction .....	1
Background of the Study.....	2
Statement of the Problem .....	4
Research Question.....	5
Significance of the Study .....	5
Conclusion.....	6
CHAPTER II: LITERATURE REVIEW .....	7
Introduction .....	7
The Notion of Collocation.....	7
The Importance of Collocation in EFL/ESL Education.....	13
Studies Exploring the Causes of Collocational Errors .....	17
Conclusion.....	30

CHAPTER III: METHODOLOGY .....	32
Introduction .....	32
The Instructional Setting and Participants .....	32
Instruments .....	34
Collocation Completion Test .....	34
Retrospective Think-aloud Protocol .....	40
Data Collection Procedure .....	41
Data Analysis .....	41
Conclusion.....	42
CHAPTER IV: DATA ANALYSIS .....	43
Introduction .....	43
The Results of Quantitative Data Analysis .....	43
The Results of Qualitative Data Analysis .....	47
Conclusion.....	50
CHAPTER V: DISCUSSION AND CONCLUSION.....	51
Introduction .....	51
Discussion of the Findings .....	51
Pedagogical Implications .....	58
Limitations of the Study .....	61
Conclusion.....	63
REFERENCES.....	65
APPENDICES.....	69

Appendix A: Collocation Completion Test.....	69
Appendix B: Transcript of a Sample Student’s Think-aloud Protocol in English .....	78
Appendix C: Transcript of a Sample Student’s Think-aloud Protocol in Kurdish .....	80

## LIST OF TABLES

Table 2.1: The most important and probable collocations according to Hill & Lewis (1997) .....	12
Table 3.1: Number of L1 equivalent and non-equivalent collocations in each collocation type .....	38
Table 3.2: Frequency/MI data for L1 equivalent and non-L1 equivalent collocations....	38
Table 3.3: Frequency/MI data for collocations with different parts of speech .....	39
Table 4.1: The results of non-parametric correlations .....	45

## LIST OF FIGURES

Figure 1: Median of the students who correctly answered the collocation types.....	44
Figure 2: Median of the students who correctly answered the collocations which are similar and dissimilar in L1 .....	47

## CHAPTER I: INTRODUCTION

### Introduction

Vocabulary and grammar are known as inseparable parts of language. However, in the field of language education, vocabulary should be at the center of language teaching and should be prioritized more than grammar, because “a language consists of grammaticalized lexis, not lexicalized grammar” (Lewis, 1993: 89). Since vocabulary is vital for language education, collocation, which is integral to vocabulary knowledge, needs undivided attention, because collocation constitutes a considerable amount of language (Hill, 2000). Generally, collocation is defined as “the way in which some words are often used together, or a particular combination of words used in this way” (Longman Dictionary of Contemporary English, 2003: 294). According to Bahns (1993), a particular feature of vocabulary that deserves more attention than it has received so far is the problem of word combination, because one of the major difficulties of EFL/ESL learners is that they do not know the possible collocations of many words.

Many scholars have acknowledged the importance of collocation, because many studies have confirmed that collocation enables EFL/ESL learners to speak more fluently, to improve their reading speed and listening comprehension, and to write in a more native like way (Brown, 1974; Pawley & Syder, 1983; Hill, 2000). However, research has constantly shown that EFL/ESL learners from different proficiency levels have problems with using L2 collocations (Biskup, 1992; Bahns & Eldaw, 1993; Farghal & Obiedat, 1995; Lennon, 1996; Park, 2003); the learners, as a result, make many collocational errors. In the literature, there are only a handful studies that have



been conducted on investigating learners' collocational errors. Many of these studies failed to give detailed information on showing the learners' major sources of collocational errors in L2. Therefore, the major sources of EFL learners' collocational errors have been little explored. For this reason, much research should be conducted in order to provide further information about what mainly leads EFL/ESL learners' to make collocational errors.

### Background of the Study

One of the considerable phenomena in the vocabulary education is the importance of prefabricated expressions (or prefabs). According to Bolinger's view "language does not expect us to build everything starting with lumber, nails and blueprint, but provides us with an incredibly large number of prefabs" (Bolinger, 1976:1 cited in Fan, 2009: 110). In the field of language teaching, prefabs refer to language units such as collocations, idioms and free combinations. Some scholars claim that among the prefabs, the main learning load for all language users is collocations, because collocations constitute a considerable amount of what native speakers say and write (Howarth, 1998; Conzett, 2000; Hill, 2000).

Collocations such as *strong tea*, *commit murder*, and *insist upon* have been defined in various theoretical frameworks. Some scholars define collocations as the co-occurrence of lexical items (e.g. Halliday and Kjellmer), co-occurrence of two or more words (e.g. Sinclair), and a type of word combination that has a syntactic function (e.g. Cowie, Mel'cuk and Howarth). The only consensus among the scholars is that collocations refer to "some kind of syntagmatic relation of words" (Nesselhauf, 2005:11).

The importance of collocational knowledge in second language (L2) competence has been widely accepted, because collocations form a major part of native speakers' language competence. Moreover, collocations enable the language learners to speak more fluently, to improve their listening comprehension and reading speed, and to write or sound more native like (Pawley & Syder, 1983; Brown, 1974; Hill, 2000; Lewis, 1997, 2000). However, many researchers have repeatedly reported that EFL/ESL learners produce many collocational errors while speaking and writing, and much research has been conducted on exploring the causes of these errors. Researchers have used students' writings (Nesselhauf, 2003; Jing, 2008; Mahmoud, 2005; Zinkgraf, 2008), translation tasks (Bahns & Eldaw, 1993; Boonyasquan, 2005; Farghal & Obiedat, 1995) and collocation completion tests (Koosha & Jafarpour, 2006; Huang, 2001) to investigate EFL learners' collocational use. The results of these studies show that EFL learners make many collocational errors. These are often seen in the students' use of paraphrasing, avoidance, synonymy, and analogy. The most-often cited causes of collocation errors are L1 interference and a lack of cultural awareness. However, serious limitations in the instruments used in these studies have meant that we still do not have a clear picture of the sources of collocation errors.

Since there is insufficient information about main sources of collocational errors in the literature, the current study contributes to the literature by providing further information about the major sources of EFL learners' collocational errors. The features that differentiate the current study from the previous ones are threefold. Firstly, the participants are Kurdish EFL learners, who are different from those people who participated in the previous studies in terms of cultural background. This is important,

because according to Baker (1992) and Huang (2001), producing L2 collocations can be affected by learners' cultural background. In addition, Kurdish language structures have their own characteristics, which are different from the majority of the languages of the subjects who participated in the previous studies. This is important, too, because native language structures have effects on producing the target language. For instance, in English you *smoke cigarette*, in Turkish you *drink cigarette*, in Kurdish you *pull cigarette*; in English you *lie in the sun*, in German you *lie on the sun*, and in Kurdish you *lie in front of the sun*. Secondly, the instruments are different from others used in the previous studies in terms of quality and quantity. The instruments contain a large number of items and different types of collocations. Finally, think-aloud protocols, which help researchers to get explicit data from what is implicitly present in students' minds (Jaaskelainen, 2002), will be used as another means of collecting data in which students' responses to these protocols will be analyzed to explore possible major source(s) of collocational errors.

#### Statement of the Problem

Language teachers accept that EFL learners make many collocational errors while producing language, whether it is spoken or written. Researchers, too, have conducted research to address this issue (Nesselhauf, 2003; Jing, 2008; Mahmoud, 2005; Zinkgraf, 2008; Koosha & Jafarpour, 2006; Huang, 2001). However, previous studies have failed to provide detailed information about EFL learners' main sources of errors in collocations.

At Koya University, which is situated in Northern Iraq, instructors in the English Department claim that the majority of EFL learners make many collocational errors,

which cause their English language production to be far from native like. This results possibly in part from some factors such as students' ignorance of the importance of English collocations, the materials designers' negligence of prioritizing collocations in English language materials, and students' and some teachers' unconsciousness of the sources of collocational errors. Thus, producing correct collocations is a major problem for EFL learners at Koya University.

#### Research Question

The present study is aimed to address the following research question:

- What are major sources of collocational errors among Kurdish EFL learners at Koya University?

#### Significance of the Study

It is widely accepted that incorrect collocations are a serious problem for EFL students. Therefore, one of the major responsibilities of language teachers is to deal with students' collocational problems (Lewis, 1997). Although many teachers who are aware of this issue devote much time to teaching collocations, students inevitably make collocational errors in their writing or speaking performance. Therefore, exploring the major sources of collocational errors is one of the major factors for reducing the rate of students' collocational errors.

At the local level, this study is possibly beneficial for the Department of English Language and Literature in Koya University to take practical steps to prioritize teaching collocations and to enhance teachers' skills in teaching lexis. In addition, this study can help EFL instructors to become more conscious about various sources of collocational errors. With this knowledge, teachers can in turn promote their students' collocational

awareness through using effective activities and remedial tasks relating to collocations. The present study can also help EFL students, especially those at Koya University, to be aware of the sources of collocational errors and practice more collocations so that they avoid collocational errors. Moreover, the results of this study can provide information for English curriculum and course planners, specifically those in Northern Iraq to design appropriate lexical materials and activities concerning EFL learners' problems with collocations.

### Conclusion

This chapter discussed the rationale for the present study. In the first part, the key points of the study were brought into focus. Following this, the background of the study was presented. In addition, the problems in both the literature and intended local institution regarding the sources of collocational errors were shown. In the final part, the importance of conducting the study for both the literature and local institution were explained. In the following chapter the previous literature about collocations will be reviewed.

## CHAPTER II: LITERATURE REVIEW

### Introduction

This chapter will review the previous literature on collocation. The first part introduces the notion of collocation, in which definitions of collocation and types of word combination identified in different theoretical approaches will be presented. The second part will focus on the importance of collocation in EFL/ESL contexts and some key features noted within different theoretical frameworks will be shown to indicate why collocations are thought to be important in English language education. In the final section, previous empirical studies investigating collocational errors will be summarized and discussed.

### The Notion of Collocation

The term “collocation” was first discussed with reference to language learning by Palmer (1933), and later introduced by Firth (1957) to the field of theoretical linguistics (cited in Hsu, 2007). Since that time, collocation has been defined within different theoretical frameworks; therefore, it is challenging to form a precise definition of collocation. One of the basic reasons that contributed to the variation in the use of collocation is that it is used by researchers working in many different fields, and its definition is usually adapted to the different aims and methods of researchers’ investigations (Nesselhauf, 2004).

Collocation has commonly been approached from two different ways. One is the “frequency-based”, or Firthian, approach in which a collocation is considered as the co-occurrence of words within a certain distance of each other in spoken or written

discourse. According to this approach, collocations are co-occurrences of words that frequently appear in a language. Firth, who is widely known as the first researcher to explicitly introduce the notion of collocation, defines collocation with reference to this approach as words with habitual company (Firth, 1957 cited in Mahmoud, 2005). This notion has inspired many researchers, such as Halliday, Kjellmer, and Sinclair in the field of vocabulary. Halliday (1961 cited in Nesselhauf, 2004) claims that co-occurrences of all probabilities of lexical items are collocations. In accordance with this view, words that are semantically related to each other occur in close distance together in a text. For instance, some words like *play*, *laugh* and *knife* frequently appear with *tennis*, *joke* and *sharp*, respectively, in context because they are semantically related to each other.

Sinclair, who is another representative of the Firthian approach, defines collocation as “the occurrences of two or more words within a short space of each other in a text” (1991:170). Sinclair (1991) states that there are three useful technical terms to describe a collocation. Firstly, *node*- the word that is under investigation. Secondly, *collocates*- the words that occur to the left and right of the *node*. Finally, *span*-the number of words on either side of the *node*. For instance, in the following sentence, the word *cinema* is analyzed; *they go to the cinema every weekend*. Words such as *they*, *go*, *to*, *the*, *every*, and *weekend* are all collocates of the node *cinema*, and a span of -4, +2 means that there are four words on the left side of the node, and two words on the right. Thus, according to the “frequency-based approach”, words that frequently co-occur are considered to be collocations.

However, many scholars believe that not all words that frequently co-occur can be considered as collocations. On this view, combinations such as *open door* and *eat food* for instance, may frequently co-occur in context, but they are not counted as collocations, because these combinations are combined due to having semantic or syntactic relations. Collocations are therefore seen as words that frequently appear with each other and whose high frequency of co-occurrence cannot be explained by semantic or grammatical relations. For instance, in the collocations *strong tea*, *heavy smoker* and *pay visit*, words such as *strong*, *heavy* and *pay* do not have any necessary semantic and syntactic relationship with *tea*, *smoker* and *visit*, respectively. Thus, words that frequently co-occur cannot always be collocations and collocations whose high frequency is a result of semantic or grammatical relations can be very misleading.

The second approach to collocation is known as the “phraseological” approach, and is strongly influenced by Russian phraseology. Typically, researchers who adopt this approach consider collocation as one particular type of phraseological unit, and see collocation as partly fixed and one type of word combination (Nesselhauf, 2004). Cowie, Mel’cuk, Benson, Benson and Ilson, and Howarth are typical representatives of this approach. Cowie (1994 cited in Nesselhauf, 2005) defines collocation on the basis of transparency and commutability (or substitutability). Transparency refers to whether the elements of the combination and the combination itself have a literal or non-literal meaning, and commutability refers to whether and to what extent the substitution of the elements of the combination is restricted. For instance, in a collocation such as *heavy smoker* the elements of the collocation have their own literal meaning; however, the combination has a non-literal meaning because the meaning of the combination does not



reflect the meaning of the component words (i.e. *heavy* and *smoker*). In addition, we cannot use *\*weighty smoker* instead of *heavy smoker*, since *smoker* is restricted to *heavy* not to *weighty*.

Mel'cuk (1998), another representative of the phraseological tradition defines collocation as “a subclass of what are known as *set phrases*; they, therefore, have to be defined in terms of their *differentiae specificae* with respect to set phrases that are not collocations” (p. 24). To Mel'cuk, collocations consist of two elements A+B, where A is freely chosen on the basis of its meaning, while the selection of B depends on A. In other words, the choice of B is restricted by A. For instance, in *do a favor* and *heavy rain* the choice of the verb *do* and the adjective *heavy* are determined by the nouns *a favor* and *rain*, respectively (since, *\*make/give a favor* or *weighty/strong rain* are not possible) (Nesselhauf, 2004).

Benson, Benson and Ilson (1986a), other representatives of the phraseological tradition, categorize word combinations into five types from most to least fixed: compounds (e.g. *floppy disk*), idioms (e.g. *be on cloud nine*), transitional combinations (e.g. *for old time's sake*), collocations (e.g. *to commit crime*) and free combinations (e.g. *to analyze, to report, to investigate a murder*). However, Hill (2000) believes that word combinations can be categorized into three parts, which are idioms such as *put the cat among the pigeons*, phrasal verbs, such as *make up* a story, and collocations, like *make a choice*. According to Hill (2000), all phrasal verbs and idioms are collocations or contain collocations. One of the major points that differentiate collocation from other types of word combinations is the frequency of collocation. In other words, in any spoken or written context, collocation appears more frequently than the other word

combinations. Among these different types of word combinations, collocation has been acknowledged as the main learning load for all language users, because it constitutes a large amount of what native speakers say and write (see Howarth, 1998; Conzett, 2000; Hill, 2000).

In the present study, following the phraseological approach, collocation is considered as the combination of two words where one of the elements is freely chosen on the basis of meaning and the other is lexically restricted to some words. This entails that collocation has two elements: one of them is free, which is a “base”, and the other is lexically determined, which is a “collocate”. The free element in a collocation retains its literal meaning, and the “collocate” often contributes a meaning element that it does not have on its own. For instance, in the collocation *pay a visit*, the word *pay* has a different meaning in isolation (*pay = to give someone money for something you buy or for a service*). However, when it collocates with *visit* (= to go and spend time in a place), its meaning changes (*pay a visit = to visit a person or place*).

In English, as in other languages, collocations are too numerous to list. Therefore, many scholars have grouped collocations into *Grammatical collocations* and *Lexical collocations* (see Benson, Benson and Ilson, 1997; Hill, 2000; Lewis, 2000; Conzett, 2000). According to Benson, Benson and Ilson, 1997, grammatical collocation such as *rely on* and *in advance* is “a phrase consisting of a dominant word (i.e. verb, noun, or adjective) and a preposition or grammatical structure such as an infinitive or clause” (p.1). Lexical collocation, in contrast, does not include prepositions, infinitives, or clauses; typical lexical collocations consist of noun, verb, adjective, and adverb (Benson, Benson and Ilson, 1997); typical instances are *hopelessly addicted*, *compose*

*music* and *break a code*. Apparently, all types of collocations are important for producing native-like language. However, some of them are more frequent and probable than others. Hill and Lewis (1997) in the *dictionary of selected collocations* listed the most important and most probable collocations (see Table 2.1). They believe that storing these selected collocations in your memory is one of the most important ways to build an effective vocabulary and to make your English sound natural. The focus of the current study is on three types of collocations, namely, verb + noun, adjective + noun and verb + preposition. The reason for choosing verb + noun and adjective + noun collocations is their high frequency in language production (Lewis & Hill, 1997). Verb + preposition collocations were chosen because Kurdish EFL learners' have particular problems with this type, as well as with the other two types.

**Table 2.1: The most important and probable collocations according to Hill & Lewis (1997)**

Collocation Type	Example
Adjective + Noun	fatal accident
Verb + Noun	accept responsibility
Noun + Verb	bombs explode
Adverb + Adjective	highly desirable
Verb + Adverb	discuss calmly

In conclusion, from the appearance of the concept of collocation, some researchers have oriented themselves to one specific definition or categorization of collocations and word combinations, whereas some others have mixed different types of

definitions and categorizations or even have come up with new ones. Generally, scholars have defined collocation with respect to two different approaches: the “frequency-based” approach and the “phraseological” approach. In the current study, collocation is defined in accordance with the “phraseological” approach, in which collocation is considered as a type of word combination. In the following section, the reasons collocation deserves more attention in EFL/ESL education will be clarified.

### The Importance of Collocation in EFL/ESL Education

Since the middle of the 20<sup>th</sup> century, the power of syntactic rules has been one of the captivations of many scholars, especially those following the Chomskyan approach. It has been accepted that one of the main parts of the language learners’ tasks is to learn structures of rules that form infinite set of sentences in the language, and to distinguish those infinite sentences from ungrammatical structures (Pawley & Syder, 1983). Recently, many scholars have come to consensus that teaching vocabulary is as important as, or even sometimes more important than, teaching grammatical structures (Lewis, 1993, 1997; Pawley & Syder, 1983; Hill, 2000). In recent years, many scholars have argued that some conventional ways of teaching vocabulary such as teaching words in isolation and memorizing bilingual vocabulary lists is less helpful than teaching words in phrases and chunks (Nation, 2001; Woolard, 2000; Howarth, 1998; Lewis, 1993, 1997, 2000; Conzett, 2000; Hill, 2000).

The importance of prefabricated units in the learners’ languages has led many teachers to shift their attention towards prioritizing word combinations, especially collocations, in EFL/ESL education. Scholars of second language vocabulary acquisition

(Bahns & Eldaw, 1993; Howarth, 1998; McCarthy, 1990), EFL/ESL materials and curriculum design (Coady & Huckin, 1997; Richards & Rogers, 2001), pedagogy (Ellis, 2001; Nation, 2001), and lexicography (Benson, Benson & Ilson, 1997) have also acknowledged the necessity of studying English collocations as an integral part of language teaching. This specifically has pushed materials designers to take this phenomenon into consideration while designing language teaching/learning materials. Howarth (1998) states that recent EFL course books show that teachers and materials writers pay considerable attention to the need for learners to acquire collocational knowledge (e.g. *Teaching collocations* by Lewis (Ed.), 2000 and *English Collocations in Use* by McCarthy & O'Dell, 2005).

A number of researchers have claimed that prefabricated units, including collocations, play a part in language learning and language fluency (Nation, 2001; Nattinger & DeCarrico, 1992; Wray, 1999 cited in Nesselhauf, 2004). Brown (1974), for example, suggested the incorporation of collocations in the EFL/ESL classrooms. She claims that learning collocations not only increases EFL/ESL learners' collocational competence, but also improves their oral fluency, listening comprehension, and reading speed. It has also been suggested that one of the basic reasons that EFL learners often find listening and reading difficult is due to the density of collocations (Hill, 2000). According to Pawley & Syder (1983), one of the major secrets behind the fluency of native speakers' language is the ready-made prefabricated units in their minds. Lewis (1997), who is another scholar prioritizing collocations in language teaching, also supports this claim by stating "fluency is based on the acquisition of a large store of fixed or semi-fixed prefabricated items" (1997:15). According to him, "fixed or semi-

fixed prefabricated items” which include collocation, are the basis for the foundation of any linguistic novelty and creativity. Thus, prefabs, including collocations are essential for fluency in both oral and written production. Further, Kjellmer (1990) ascribes the distinction between native speakers and language learners largely to the difference in the automation of collocations. According to him, native speakers have already acquired the collocations, and in producing utterances, natives benefit from those ready-made prefabricated units. The learners, on the other hand, have few ready-made collocations in their mental lexicon; therefore, language learners tend to use long sentences or inappropriate phrases while expressing their ideas.

Another claim is made by Carter and McCarthy (1988), who state that “students do not have to reconstruct the language each time when they want to say something; instead, they can use these collocations as pre-packaged building blocks” (p. 75). Sometimes students, who are insufficient in collocational knowledge, stop in the middle of conversation, because they cannot find suitable phrases for conveying their messages. This is also acknowledged by Hill (2000), who claims that collocations make thinking easier, because they allow us to “identify and produce complex ideas without using all our brain space to focus on the form of the words” (p.55). Moreover, Hill (1999) in his article states that “students with good ideas often lose grades because they do not know the four or five most important collocations of a key word that is central to what they are writing [or speaking] about” (p. 5). Therefore, collocations always can be used as ready-made phrases for expressing various ideas.

The knowledge and the capability of using collocations are essential for language learners and for naturalness of language. Unfortunately, however, language learners,

even advanced ones also face considerable difficulties in using collocations correctly.

Quotations similar to the following abound in the literature:

*Language learners often stumble across co-occurrence relations.*

*(Smadja, 1989:164)*

*Any analysis of students' speech or writing shows a lack of [...] collocational competence.*

*(Hill, 2000:49)*

*Knowing which subset of grammatically possible utterances is actually commonly used by native speakers is an immense problem for even the most proficient of non-natives.*

*(Wray, 1999:468 cited in Nesselhauf, 2005:3)*

Learners who are deficient in collocational competence or do not have ready-made chunks in their mental lexicon, which help them to precisely express their ideas, tend to generate utterances on the basis of grammatical rules that leads to numerous collocational errors.

There is a wide agreement that collocations have to be taught (Nation, 2001; McCarthy, 1990; Hill, 2000), because when we look at the error types of EFL/ESL students, we accept that collocations play a major role in EFL/ESL contexts, since many of the errors are in collocations (Meara, 1984). However, many types of prefabricated units, including collocations are still not considered adequately in English language teaching today (Nesselhauf, 2004). By the same token, many teachers and researchers (e.g. Boonyasquan, 2006; Lewis, 2000; Conzet, 2000) suggest that collocations should be covered in every single stage of a learner's academic path, and should be highlighted when teaching any English language skill such as listening, speaking, reading, writing and translating because one of the most essential phenomena to improve students' fluency and accuracy is to enhance their mental lexicon by providing quality

collocational input. The next section will describe some empirical studies which have aimed to investigate what leads EFL learners to make collocational errors.

### Studies Exploring the Causes of Collocational Errors

Previous empirical studies have concluded that EFL learners make many collocational errors while producing language. Furthermore, there has been a great concern among researchers about the reasons why EFL/ESL students frequently commit collocational errors in their writing and speaking. Researchers who have tried to investigate collocational errors have used different instruments as a means of data collection. EFL/ESL learners' writing, discrete point tests of collocation including translation tasks, cloze tests, interviews, vocabulary test and proficiency tests, and collocation completion tests have all been used as instruments for addressing why learners make collocational errors. However, these researchers have failed to find the major sources of errors in collocations.

To begin with those researchers who investigated EFL students' writing, Nesselhauf (2003) examined 32 essays written by German speaking learners of English to explore the use of verb + noun collocations in their free written production. She conducted the methodology in three different steps. Firstly, she extracted the verb + object + noun combinations from the essays; then she classified the combinations according to their degree of restriction (i.e. idioms, collocations, and free combinations). Finally, the combinations were evaluated as to their acceptability in English. She found many errors in collocations, free combinations and idioms. She claims that of all types of verb + object + noun combination errors "the one occurring most frequently is the



wrong choice of the verb” (2003:231). Indeed, this is not surprising, because according to her definition of collocation the verb in a collocation has a limited sense, which leads the students to face difficulties in using the verb correctly. According to the results of her study, the percentage of errors in free combinations and collocations are very close to each other; therefore, she claims that “the degree of restriction does not have a major influence on the types and amount of mistakes that learners make” (p.234). Nesselhauf (2003) also finds that L1 has considerable influence on all types of word combinations, including idioms and free combinations. However, the influence of L1 is greatest in collocations.

In order to examine how collocations are handled by Chinese EFL learners, Jing (2008) examined the most common types of collocational errors in Chinese EFL learners’ compositions and tried to explore the possible causes of these errors. The data for this study came from the one-million word CLEC-Chinese Learner English Corpus (1997). According to this study, Chinese EFL learners tended to make errors which are caused by language transfer, such as using synonyms (e.g. *\*large improvement* and *\*develops very much*) and words with overlapping meanings (e.g. *\*reasons cause* and *\*works a job*). Based on the analysis of this corpus, it is claimed that the extracted collocational errors resulted from forming hypotheses of semantic equivalents between English and their native language. In other words, Chinese EFL learners are apt to make word-for-word translations in their writings; as a result, they make collocational errors. Major drawbacks of this study are that the number of participants is not mentioned and the number of the extracted collocational errors is unknown. These are important, because the number of the participants and investigated collocations affect the results of

the study, especially when these numbers are small. Additionally, using synonym is not a cause of error; rather it is a type of error because it doesn't explain why the error happened, it just tells us what the error was. Therefore, the conclusions of the study could be questionable.

In another study, Mahmoud (2005) investigated 42 essays written by Arabic-speaking university students majoring in English to explore their collocation error types and the causes of these errors. The essays were written as a homework assignment, in which the students were free to write about any social issue of their choice.

Additionally, the students were unaware that their usage of collocations would be examined. In the students' essays, many lexical and grammatical collocational errors were extracted, and they were given to native-speaking university teachers to check whether they were correct. Mahmoud (2005) concludes that two thirds of the extracted collocations were incorrect and the majority of them were lexical collocations. These lexical collocational errors resulted from incorrect word choice such as *\*make the homework* (= do homework) and *\*hurts the mind* (= harms the brain), incorrect word form such as *\*wants to get marriage* and *\*famous musician band* and contextual errors (i.e. linguistically correct but contextually incorrect) like *\*bring a boy* (the correct form is *give birth to a boy*) and *\*carrying her baby* (the correct is *pregnant with her baby*). The results show that students in their writing relied mostly on their native language, since they possibly believed it would be easy to find the EFL equivalents in their native language. According to Mahmoud (2005), the students produced some lexical errors due to having problems within their first language. For instance, the error *\*gain language* could be due to the students' inability to see the difference between *yaksab* (= gain/win)

and *yaktasab* (= acquire) in their first language. These errors, however, may not be due to negative transfer of the students' first language, but could be due to their insufficiency in their first language, because if they knew the difference between *yaksab* and *yaktasab*, positive transfer would occur; as a result, the students would be able to produce correct collocations.

Some researchers who investigated students' writings have concluded that besides L1 interference, there could be other possible sources that contribute to learners' collocational errors. Zinkgraf (2008) analyzed verb + noun collocations in the written production of 102 Spanish-speaking university students of English as a foreign language taking English courses of teacher and translator training programs. The data were collected from 13 different assignments including comprehension tasks, essays and reviews that students completed during the courses. According to the results of data analysis, the frequent atypical combinations were those collocations that included the most frequent delexical verbs such as *do*, *make*, *take* and *have*. What is striking regarding these verbs is that they are simple, they are learnt at the early stages of the acquisition process, and most of them belong to the 1000 most frequent words in English. However, these very frequent words in English appeared in the incorrect collocations produced by the students who are advanced learners of EFL (Zinkgraf, 2008). The results of the study show that the extracted miscollocations were attributable to the wrong choice of both nouns and verbs in atypical collocations, since the students used the verbs with many nouns that do not collocate (i.e. overgeneralization). In addition, the influence of the learners' mother tongue and semantic overlap between appropriate form and possible synonyms of either the base or the collocate were also

other causes of producing incorrect collocations. Since, the focus of this study is on verb+noun collocations in the students 13 compositions, it is hard to draw a generalizeable conclusion on the basis of 13 compositions and one type of collocation, because there are many types of collocations and most of them should be considered during investigating learners' collocational knowledge. Thus, the results of the study may not show the learners' actual knowledge of collocations.

The studies described so far used students' writings as a means of collecting data. The majority of the studies stressed that L1 has a vital role in producing incorrect collocations. The results of these studies can be used as evidence to support Baker's (1992) statements, in which she claims that many learners or translators often face difficulties in using the second language correctly because in their first language, these people cannot find some collocations that carry similar meaning (Baker, 1992). Consequently, the learners try to make word-for-word translations, which make their language incorrect. For instance, *play the piano* is an unacceptable collocation in Kurdish, where the usual expression is *\*hit the piano*, which is quite unnatural in English. Therefore, if learners could not find the equivalent collocation in the target language, they tend to translate the phrase word by word, which sometimes causes errors. Moreover, some researchers, who investigated students' essays, report that apart from L1, substituting synonyms and overgeneralization could also be other possible causes of making collocational errors. It is worth stating however that overgeneralization and using synonym are not causes of collocational error, but they are types of error.

In addition to investigating students' writings, researchers have used discrete point tests of collocation to examine students' patterns of collocational errors, and to explore the possible causes of these errors. Bahns and Eldaw (1993) gave a translation task and a cloze task to 58 German EFL students enrolled in their first to third year at Keil University. In the translation task, 15 English verb + noun collocations were selected, translated into German and were set into 15 German sentences. The participants were asked to translate the sentences into English; the ideal aim of this translation was to see the selected English collocations in students' translated sentences. They assumed that if the students did not know a collocation, they would try to paraphrase it. For the cloze task, the selected collocations were set into similar English sentences in which the students had to provide the missing verbal collocates to the given noun node. The tasks were distributed to the participants during their regular classes, and the informants did not have access to any reference books. The items in both tasks were rated as acceptable if semantically accurate and idiomatically correct and unacceptable if semantically inaccurate and idiomatically incorrect. The collected data were then evaluated by three native English speakers. According to the results of this study, EFL learners' competence in general vocabulary does not expand in parallel with their knowledge of collocations, because "learners are more than twice as likely to select an unacceptable collocates as they are to select an unacceptable general lexical word" (1993:108). Additionally, in many items of the translation task the students successfully paraphrased the collocations; therefore, the collocations which are easy to paraphrase, were avoided by the informants by replacing them with alternate but correct forms, while those collocations that are difficult to paraphrase were produced incorrectly. For

this reason, Bahns and Eldaw (1993) suggest that “we should concentrate [...] on those collocations which cannot at all or easily be paraphrased” (p.109). It is worth mentioning that the number of collocations used in this study is rather small; therefore, it is hard to believe that using only 15 collocations can be considered to measure students’ knowledge of collocations, since there is a huge amount of collocations in language. Moreover, there is an imbalance between the selected collocations in terms of frequency, in which some of them such as *arouse perfection* (Freq. = 8 per 400 million words according to Corpus of Contemporary American English), *refuse admission* (Freq. = 29) and *pay compliments* (Freq. = 31) (p.111) have rather lower frequency than *whip cream* (Freq. = 710), *do damage* (Freq. = 3366) and *serve sentence* (Freq. = 726). This imbalance of frequency of the collocations is actually a serious problem that affects the results of the study, because if the collocations are not at the same level of frequency or at least if their frequencies are not close in number, it would not be obvious whether some incorrect collocations result from the students’ generally insufficient knowledge of collocations or from the infrequency of these particular collocations. Additionally, no information is given about the frequency of the elements of the collocations. It may therefore be that students may not know a particular collocation just because the component words of the collocations are infrequent. Thus, these drawbacks of this study make its results to be questionable.

To show the effect of L1 on grammatical collocations, Koosha and Jafarpour (2006) conducted a study to establish to what extent presenting materials relating to collocations including prepositions through data-driven learning (DDL) has any effect on the teaching/learning of these collocations and to determine the extent to which

Iranian EFL learners' knowledge of collocation of prepositions is affected by their L1. The participants were 200 Iranian EFL seniors who were selected randomly from three different universities in Iran. The informants were given the Michigan Test of English Language Proficiency to identify their proficiency levels, and were divided into low, mid and high levels of proficiency according to their scores. Additionally, the participants in each low, mid and high group were randomly assigned to experimental and control groups. Completion test of collocations of prepositions as a pre-test including 60 items was distributed to the students to determine their knowledge of the collocations. Both groups received 30-hour sessions on collocation of prepositions; but the experimental group received the sessions through data driven-based approach and the control group received the sessions through conventional approaches. After receiving the sessions, another completion test on these collocations including 60 items as a post test was given to the participants in order to identify the effect of the instruction. To find out the extent to which the learners' knowledge of these collocations is affected by their L1, a translation task including 60 fill-in the blank items on the collocation of prepositions was also used. The study showed a significant difference between the performance of the participants in the DDL group and control group, suggesting that presenting materials through data-driven learning (DDL) is highly effective in the teaching and learning collocation of prepositions.

Regarding the effect of the participants' L1, it is concluded that 68.4% of the extracted errors of collocation of prepositions are due to the interference of the students' L1, and 31.6% were attributable to intralingual transfer. Therefore, the impact of L1 on the use of collocations of prepositions seems to be highly significant. Koosha and

Jafarpour (2006) suggest that such collocations should be taught both in context and with reference to L1. One of the prominent characteristics of this study is that a larger number of participants and items were used, and in both the completion test on collocation of prepositions and the translation task many types of collocation of prepositions were considered. Therefore, this can be considered as one of the better studies investigating collocations of prepositions.

In another study, Boonyasaquan (2005) analyzed collocational violations in a translation task. The participants were 32 fourth-year English majors in a university in Thailand. The instrument of this study was a business news article translated from Thai into English, and the translated article was parsed into 30 meaningful parts on the basis of the Thai version. The parsed parts were listed and rated by two English native speakers. The focus of the study was on nine types of collocations: adjective + noun, verb + noun, noun + noun, verb + adverb, noun +verb, adverb + verb, verb + prepositions, prepositions + noun, adverb + adjective. According to the results of the study, adjective + noun pairs had the highest percentage of collocational violations (21.31 %), and preposition + noun pairs had the lowest (4.91 %). After analyzing the patterns of collocational violations, the possible sources of violations including over-literal translation, paraphrasing, using synonymy, L1 transfer and avoidance were explored. According to the study, over-literal translation (32.76%) was the most frequent strategy that the participants used during the translation task. A major limitation of this study is that the frequency of the collocations was unknown. It was not mentioned whether the selected collocations are frequent or infrequent in English



language, because generally infrequent collocations are naturally difficult for learners. Therefore, it is hard to decide whether the study is valid.

Farghal and Obiedat (1995) used an Arabic translation task and an English blank filling task involving 22 collocations relating to core topics such as food, color and weather. The English blank filling task was administered to 34 junior and senior English major students at Yarmouk University, and the Arabic translation task was given to 23 English teachers who had had a minimum of five to ten years' experience in teaching English. In the fill-in-the-blank task, one of the elements of the collocation is given and the other had to be provided by the informants. Additionally, in the translation task the subjects had to provide English equivalents to the given Arabic collocations. According to the results of the study, the participants were seriously deficient in collocations, as many collocational errors were detected in their tasks. Since the informants were very deficient in collocations, they relied heavily on lexical simplifications such as substituting synonyms, paraphrasing, avoidance, and L1 transfer. In accordance with this study, among these four strategies, using synonyms was the most frequent and reliance on L1 was the least frequent strategies adopted by the participants. It is worth saying that this study has many serious drawbacks. One of which is the quantity of the items. In this study, only 11 items were used to measure the students' collocational knowledge and to explore the causes of the errors. It is hard to decide that the participants were deficient in collocations on the basis of only 11 collocations, since there are numerous collocations in English. Another major drawback is the quality of the selected collocations. Some of the required collocations such as *lenient rules* (Freq. = 4/400 million words) and *weak tea* (Freq. = 43/400 million words)

are very infrequent. Further, in the item “*To many people, cold food is better than **hot food.***” (p.330), *hot food* is not a collocation, rather, it is free combination; since the adjective *hot* can be used with numerous nouns such as *hot bath, hot chocolate, hot air etc.*(see Longman Dictionary of Contemporary English, 2003). Moreover, in one of the items, instead of measuring the knowledge of collocations, the knowledge of general vocabulary was measured. For instance, in “*There is a lot of pepper in this kebab. It is too **hot** for me.*” (p.330) the students were required to provide the adjective *hot* because it has a semantic relation with *pepper*; and in this item *hot* is not as an element of a collocation, rather it is an element of the sentence. Further, the collocate *hot* is not a part of the node’s (i.e. *pepper*) sentence, rather the collocate is in another sentence. Therefore, these types of combinations cannot be counted as collocations, because in collocations there is, at least, one restricted element or at least the elements of a collocation should co-occur; however, in some of the combinations that Farghal and Obiedat (1995:330) considered as collocations, both elements are free (e.g. *hot food* and *rich food*) and the elements do not co-occur (e.g. *pepper.....hot* in item No.4, p.330). Based on these drawbacks, it can be concluded that the results of this study are highly questionable.

Another researcher who has used collocation completion tests as a means of investigating EFL learners’ collocational errors is Huang (2001). He investigated Taiwanese EFL learners’ knowledge of collocations and the collocational errors they committed. He gave a self-designed collocation completion test to 60 Taiwanese EFL learners to measure their knowledge of four types of lexical collocations: free combinations, restricted collocations, figurative idioms and pure idioms. The test

included a total of 40 items in the form of free-responses, with ten items for each collocation type. The items were presented in the form of sentence contexts, in which the students had to provide the missing parts of speech. The grammatical errors were not counted, since the focus was on choosing the correct collocates. The data were analyzed using both qualitative and quantitative paradigms. The results indicate that due to their insufficient collocational knowledge, the participants were unsatisfactory in producing correct English collocations. Free combinations were the easiest for the students; this perhaps is because free combinations can be easily paraphrased without causing any lexical errors. However, students had the most difficulty in producing pure idioms; according to the researcher, this is due to “their lack of cultural awareness” (Huang, 2001:126). Additionally, both restricted collocations and figurative idioms were at the same level of difficulty, and errors in both of them were attributed to the influence of students’ L1. For instance, the participants chose *eat* to collocate with *bite*, *grow* with *fruit*, *pure* with *coffee*, which are direct translations from Chinese (Huang, 2001:123). Additionally, in some instances the participants adopted strategies such as avoidance and analogy. According to Huang (2001), to enhance learners’ lexicon, they need to learn words’ cultural connotations, semantic fields and collocational restrictions, because through this, learners can improve their lexical competence. One of the aspects that Huang (2001) stresses regarding teaching idioms and collocations is the cultural connotation of these combinations. This claim is quite convincing, because some idioms and collocations give offensive meaning in some languages; therefore, these culturally specific concepts confuse EFL learners, and they often use these concepts incorrectly. For instance, the Russian collocations on emotions are connected with local images of

nature; for this reason, these collocations are culturally marked (Huang, 2001). Thus, teaching word combinations through cultural perspectives may promote the processing and retention of these combinations of words, whether they are idioms or collocations.

The studies discussed so far have been conducted to examine EFL learners' collocational knowledge by using various instruments such as writing, translation task, cloze task, blank filling task and collocation completion test. The previous literature has confirmed that EFL learners are insufficient in producing correct collocations, and most EFL learners adopt various strategies, which lead to certain types of collocational errors. Previous empirical studies on analyzing collocational errors have concluded that L1 interference, using synonyms, paraphrasing, avoidance, analogy and lack of cultural awareness are causes of collocational errors. However, apart from the influence of L1 and lack of cultural awareness, these are not causes of collocational errors, rather they are types of errors. Generally, the previous studies that used elicitation tests such as translation and cloze tasks, blank filling tests, and collocation completion tests have two major limitations. Firstly, the items were used in these tests were generally small in number. Therefore, the results of these studies failed to show the actual knowledge of the learners in L2 collocational use. Secondly, the investigation was often narrowed down to a particular collocation type. For instance, Nesselhauf (2003), Zinkgraf (2008) and Bahns and Eldaw (1993) examined only verb + noun collocations, and Farghal and Obiedat (1995) investigated only adjective + noun collocations. These studies therefore do not provide a deeper understanding of the L2 learners' collocational use, since the focus was on the specific type of collocations. Thus, further research should be conducted to get information about L2 learners' treatment of other types of collocations.

The results of the studies discussed above also confirm the importance of conducting the current study, because in the current study the participants are Kurdish EFL learners, who have different cultural background from those people who participated in the previous studies. Moreover, the semantic and grammatical structures of Kurdish language are different from the native language of those participants of the previous studies; this phenomenon may be helpful to explore major sources of collocational errors. Another reason for conducting the present study is that larger numbers of items and more types of collocations will be included in the intended collocation completion test. This will help to assess the actual knowledge and get a deeper understanding of EFL learners' L2 collocational use. Additionally, another means of collecting data will be think-aloud protocols, which have not been used in any of the previous studies. This will also be helpful for exploring possible main sources of collocational errors.

### Conclusion

The major concern of this chapter was to review the literature on collocations. This was presented in three sections: the notion of collocation, in which definitions and types of collocation were showed, the importance of collocation in EFL education, in which the need for studying collocations in EFL classrooms was reconfirmed, and empirical studies about analyzing EFL learners' collocational errors were reviewed. Although there is a considerable amount of literature on investigating collocational errors, there is still a need to conduct further research to obtain information about major sources of collocational errors. Therefore, the aim of the present study is to explore

Kurdish EFL learners' major sources of collocational errors by using a collocation completion test and think-aloud protocols.

## CHAPTER III: METHODOLOGY

### Introduction

This experimental study aimed to probe into sources of collocational errors made by Kurdish EFL learners at Koya University. The data were collected through a collocation completion test and think-aloud protocols and were analyzed to answer the following research question:

- What are major sources of collocational errors among Kurdish EFL learners at Koya University?

This chapter includes information about the instructional setting and participants, instruments, data collection procedures, and data analyses processes.

### The Instructional Setting and Participants

The present study was conducted at Koya University in Northern Iraq. The instructional setting was the fourth year class in the department of English Language and Literature, which is a faculty of the College of Languages. The participants were 40 Kurdish college seniors (24 male and 16 female) studying English language and literature, and their level of English proficiency was expected to be between upper-intermediate and advanced. The reason for choosing those participants is due to their problems with collocations. English major seniors at Koya University at this level still make many collocational errors while speaking and writing. Since they will soon become English teachers, it is important that these problems should be overcome and students' awareness of collocations increased. Therefore, this research was conducted in order to explore these students' main sources of collocational errors.

English major students at Koya University have to study for four academic years and have to pass the faculty's required examinations in order to receive a Bachelor's certificate in English Language and Literature. In this department, students study at different proficiency levels as they progress through the academic years. They study beginner and elementary levels in first year, intermediate level in second year, upper-intermediate in the third and advanced level in the last year. During these four years, students study different subjects. Students are taught subjects relating to reading, speaking, writing, vocabulary and grammar skills of English language and some literary subjects such as short story, poetry and drama. Classes concerning listening skill are very rarely given to the students.

The course-books, which are at the same time the faculty's syllabi, consist of linguistic and literary subjects. Linguistic subjects, on one hand, include books relating to grammar, syntax, semantics, linguistics, vocabulary, speaking, and writing. However, collocation, which is an important part of vocabulary, is not given serious consideration; rather it is treated as a subsidiary part of vocabulary. This is possibly due to students' and some teachers' unawareness of the importance of collocation in English language. Literary subjects, on the other hand, are books including short stories, drama, poetry, novel, and criticism. It is worth saying that these different subjects are arranged according to the students' proficiency levels in each year. Language tests and examinations cover the topics in the course-books and what has been studied.

In general, students participate in classroom discussions; do their homework assignments and some other language-related activities. Additionally, students in the 4<sup>th</sup> year write a graduation research paper about a literary or linguistic topic.



## Instruments

In the present study two different instruments were used to gather the intended data to answer the research question. The first instrument was a collocation completion test, the second was a retrospective think-aloud protocol.

### *Collocation Completion Test*

This instrument was a multiple-choice collocation completion test designed by the researcher, which included 75 items and covered three types of collocations: Verb + Noun, Adjective + Noun, and Verb + Preposition. 25 items were included for each collocation type. These collocations were presented in sentence contexts in which one of the elements of the collocations (i.e. verb in V+N, adjective in ADJ+N, and preposition in V+P) was deleted. The participants were required to choose the best among the given options to complete the sentences.

The test aimed to measure the importance of a number of different variables in determining how difficult collocations are for learners. The variables included are the part of speech of the collocation, the frequency of the collocation and its component words, the mutual information score of the collocation, and whether the collocation matches a collocation used in the L1.

Accordingly, the test was prepared through a number of different steps. First of all, the focus of the researcher was on three types of collocations (i.e. V+N, Adj+N and V+Prep); the reason for choosing these types of collocations is their high frequency in English language production. Of course, all types of collocations are important for producing native-like language, but some of them are more frequent and probable than others. Two dictionaries, which were *Longman Dictionary of Contemporary English*

(2003) and *Oxford Collocations Dictionary for Students of English* (2002), were used as sources of extracting the intended collocations. Collocations were intended to be extracted on the basis of particular criteria which are frequency and Mutual Information (MI). The MI compares “the probability of two words occurring together through intention with the probability of the two words occurring together by chance” (Lee & Liu, 2009:208). This means that MI shows the extent to which a strong relationship exists between the components of a collocation. High MI score indicates a strong relationship between the components of collocations. For instance, the components of the collocation *ground pepper* (MI = 11.73) have a stronger association than components such as *face problem* (MI = 4.25), since the former ones have higher MI score.

According to the criteria, collocations, in order to be selected, should have a frequency of at least once per million words, and a Mutual Information (MI) of at least 4.00.

Concerning the extraction of V+N collocations, a list of target verbs to be searched for was created. The target verbs were taken randomly from two dictionaries, namely, *dictionary of selected collocations* (1997) and *Oxford collocations dictionary for students of English* (2002). For each collocation, the *Longman Dictionary of Contemporary English* (2003) and *Oxford Collocations Dictionary for Students of English* (2002) dictionaries were consulted for nouns that can be used with each target verb. These candidate collocates were then checked in the Corpus of Contemporary American English (COCA) website until a collocation meeting the criteria was found. This strategy meant that sometimes only one noun was checked in COCA; sometimes

two or three nouns were checked before a collocation meeting the requirements was found. The process of V+N collocations extraction was repeated for extracting the Adj+N and V+Prep collocations. Before the extraction, lists of target nouns and verbs for each Adj+N and V+Prep collocations, respectively, were created. These target nouns and verbs were taken at random from both *dictionary of selected collocations* (1997) and *Oxford collocations dictionary for students of English* (2002). For each Adj+N and V+Prep collocation, many adjectives and prepositions were checked in the *Longman Dictionary of Contemporary English* (2003) and *Oxford Collocations Dictionary for Students of English* (2002) dictionaries and these adjectives and prepositions together with their candidate collocates (i.e. nouns and verbs, respectively) were checked in COCA to find out whether those collocations met the criteria. Among those possible adjectives and prepositions, each one of them was checked in COCA until the required Adj+N and V+Prep collocations were found. In this way, a list of 150 collocations (i.e. 50 collocations for each collocation type) that met the criteria was compiled.

Secondly, the frequency of each element that constitutes the extracted collocations was also checked in the COCA website. This is important, because students may not know a particular collocation just because they do not know one of its constituent parts because they are infrequent in English language. Therefore, collocations in which both elements are infrequent were not included in the collocation completion test. Each element in the collocations in order to be counted as a frequent element had to have a frequency of at least 25 per million words. As a result,

collocations in which the frequency of both elements was less than 25/million words were not included in the collocation test.

Thirdly, the literal meaning of these collocations in the students' first language (i.e. Kurdish) were also considered. Therefore, the researcher with a lecturer teaching English at Koya University checked the collocations to indentify whether these collocations are similar or dissimilar in the participants' L1.

As a result of these processes, 75 collocations were selected and included in the collocation test. These collocations were selected on the basis of parts of speech (i.e. collocation types), frequency, MI, and similar and dissimilar in the students' L1. These selected collocations included equal numbers of the three parts of speech (i.e. 25 collocations for each collocation type), and roughly equal numbers of higher and lower frequency and MI of the collocations, and both L1 equivalent and non equivalent collocations.

Since the effects of each collocation type - the frequency and MI of the collocations, the frequency of constituent words, and the relationship to L1 on producing correct collocations - were going to be investigated separately, it is important that the relationships between these factors be understood in advance.

First, it should be noted that approximately equivalent numbers of L1 equivalent and non-L1 equivalent collocations were included under each part of speech (see table 3.1). A chi-square test confirmed that number of L1 equivalent collocations did not differ across part of speech ( $\chi^2(2) = 18.03, p > .05$ ). The influence of these factors can therefore be evaluated entirely independently of each other.

**Table 3.1: Number of L1 equivalent and non-equivalent collocations in each collocation type**

Collocation types	Number of L1 equivalents	Number of non-L1 equivalents
Verb + Noun	12	13
Adjective + Noun	12	13
Verb + Preposition	13	12

Similarly, the frequency and mutual information of the collocations and the frequencies of their component words did not differ significantly between collocations which were L1 equivalent and non-L1 equivalent (see Table 3.2).

**Table 3.2: Frequency/MI data for L1 equivalent and non-L1 equivalent collocations**

	L1 equivalent collocations	Non-L1 equivalent collocations	Mann-Whitney tests' results
<i>Mdn</i> Collocation frequency	1,136	1,144	$U = 691.5, p > .05$
<i>Mdn</i> MI	5.56	5.36	$U = 699.5, p > .05$
<i>Mdn</i> Word 1 frequency	29,441	28,632	$U = 649.0, p > .05$
<i>Mdn</i> Word 2 frequency	109,229	102,708	$U = 614.0, p > .05$

However, it is important to note that, since collocations of different parts of speech inevitably occur at different levels of frequency, it was not possible to keep frequency and MI equal across the different part of speech categories (See table 3.3). According to my corpus searches, V + Prep collocations were the most frequent, followed by V + N. Adj + N collocations were the least frequent. For MI, this

relationship was reversed, with V + Prep having the lowest, and Adj + N the highest scores. V + N had the highest frequency first words, and V + Prep the lowest, while V + Prep had the highest and Adj + N the lowest frequency second words. These interactions will need to be taken into account when the effects of parts of speech and of frequency are considered.

**Table 3.3: Frequency/MI data for collocations with different parts of speech**

	V + N	Adj + N	V + Prep	Friedman's ANOVA
<i>Mdn</i> Collocation frequency	1,347	693	4,383	$\chi^2(2) = 9.15, p < .01$
<i>Mdn</i> MI	6.18	6.27	4.41	$\chi^2(2) = 20.12, p < .001$
<i>Mdn</i> component 1 frequency	46,524	31,526	11,722	$\chi^2(2) = 20.04, p < .001$
<i>Mdn</i> component 2 frequency	47,589	43,174	1,643,271	$\chi^2(2) = 41.55, p < .001$

To create the test items, sentences including the target collocations were taken from the British National Corpus (BNC) and *Longman Dictionary of Contemporary English* (2003). In each item, one part of the collocation was deleted (i.e. nouns, adjectives and prepositions were deleted in V + N, Adj + N and V + Prep collocations, respectively). Four options were then provided: one being the correct collocate and three distracters.

Distracters were selected based on two aspects. One of the distracters was chosen on the basis of L1. The other two distracters were selected based on synonym. This denotes that one of the distracters was the L1 equivalent of the correct answer, whereas the other two distracters were synonyms of the correct answer. It is worth noting that

when the correct answer was the L1 equivalent distracter, in these cases the third distracter would also be a synonym of the correct answer.

After designing the collocation test, it was piloted on 4 native speakers of English at Bilkent University. The items which were not answered as predicted by all native speakers were modified.

#### *Retrospective Think-aloud Protocol*

Retrospective think-aloud protocol was the second means of collecting data. When asked to “think-aloud,” it means that the participants are generally asked to express aloud the thoughts running through their heads while completing a task provided by the researcher. In this process, the participants are asked to say whatever they think of, whether related to the task or not, and the transcripts of these spoken records of mental process are called protocols, which are analyzed for patterns and these patterns can form the basis for generalizations and further research (Rankin, 1988).

For these protocols, 6 students were chosen based on their scores (two top scores, two average scores, and two poor scores) in the collocation completion test. The selected students were separately interviewed in their first language. These interviews were conducted two days after distributing the collocation completion test. This is important, because this process is related to something done in the past; if the process was carried out late, the interviewees possibly would forget many things; as a result better results could not be achieved.

In this process, the interviewees were asked to give reasons for choosing a particular option when completing 15 selected collocations in the collocation completion test. The items selected were the same for all interviewees. These selected collocations

included equal numbers of each collocation type (5 collocations for each). The majority of these selected collocations were those answered incorrectly by the participants. However, a few of these collocations were those answered correctly by the students. The reason for choosing these correct items was to understand whether they really knew that choosing this particular distracter is the correct choice. The answers of these students were audio taped and their reasons for choosing particular options were translated into English.

#### Data Collection Procedure

The data were collected in two different phases. Firstly, the collocation completion test was administered as an in-class activity to 40 seniors majoring in English. Additionally, the students were not allowed to use any language sources such as dictionaries and vocabulary books. The time given to complete the test was one hour.

The second stage of data collection was interviewing some students. Six students were selected and separately interviewed for about 3 minutes. The interviews were recorded and analyzed for possible main sources of errors in collocations.

#### Data Analysis

Quantitative and qualitative data were obtained from the collocation completion test and retrospective think-aloud protocols. The quantitative data collected from the collocation completion test were analyzed using SPSS. Firstly, students' answers were analyzed to explore any difference in difficulty among the three different parts of speech (i.e. V+N, Adj+N, and V+Prep). Kruskal-Wallis tests were used to address this issue. Secondly, the data obtained from the test were analyzed to explore the relationship between the participants' correct answers to particular collocations and the frequency,



Mutual Information (MI), and frequency of components that include in those particular collocations. These analyses were carried out through using Correlations. The third part of the analysis used Mann-Whitney tests to find out whether the number of students answering the question correctly changes according to whether the answer is the same in L1.

### Conclusion

This chapter presented information about the participants and instructional setting, and instruments. The participants were 40 Kurdish seniors studying EFL at Koya University. Two instruments were used to collect data. A collocation completion test was administered to the learners to explore major sources of collocational errors and a think-aloud protocol was used to find out if there could be other possible major source(s) of errors in collocations. This chapter also showed information about data analysis and data collection procedures. In the following chapter, the process of data analysis is presented.

## CHAPTER IV: DATA ANALYSIS

### Introduction

The objective of the present study was to explore the main sources of collocational errors made by EFL learners at Koya University. This chapter presents the results of both quantitative and qualitative data analyses performed to address this issue.

### The Results of Quantitative Data Analysis

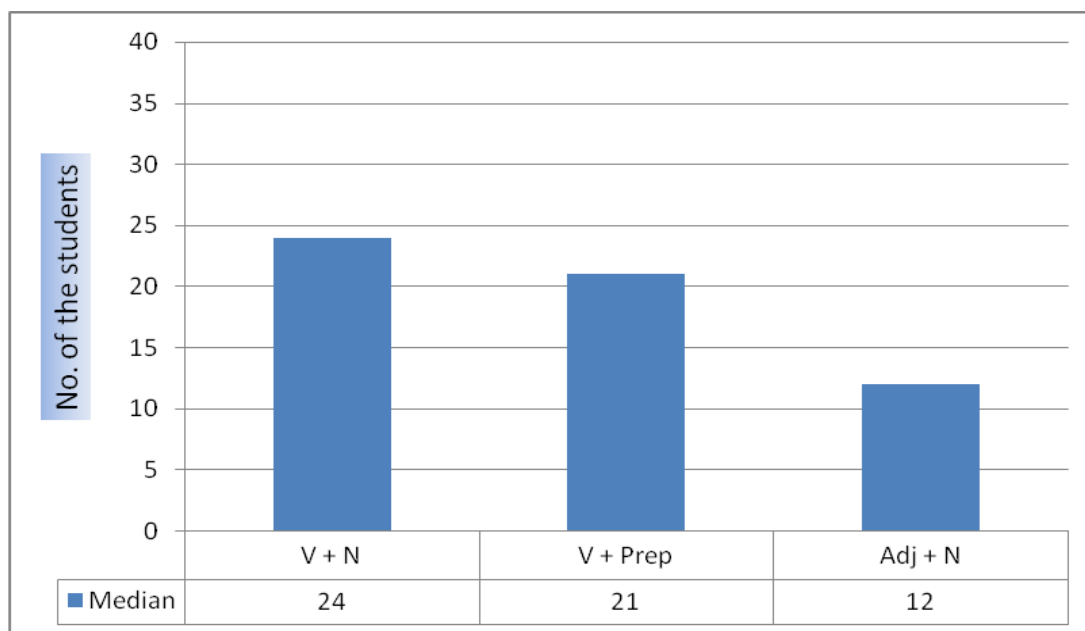
The quantitative data were obtained from the collocation completion test. To analyze the data, 7 variables were created. The variables were:

1. parts of speech of the collocations;
2. frequency of the collocations;
3. Mutual Information (MI) of the collocations;
4. frequency of the first component of the collocations;
5. frequency of the second component of the collocations;
6. whether the collocations are similar or dissimilar in L1;
7. the number of students giving correct answer to an item.

The quantitative data obtained were analyzed in three main steps in order to yield many results. As the first step of the analysis, the data were examined to explore whether the number of students answering an item correctly changes according to different groups of parts of speech of the collocations. Descriptive statistics (see figure 1) showed that V + N collocations were apparently the easiest type, being answered correctly by a median of 24 out of 40 students. V + Prep collocations were the second

easiest type ( $Mdn = 21$ ) and Adj + N collocations were the most difficult ( $Mdn = 12$ ) type for the students.

**Figure 1: Median of the students who correctly answered the collocation types**



Since both Kolmogorov-Smirnov and Shapiro-Wilk tests showed that the data were not normally distributed, non-parametric Kruskal-Wallis tests were used to test for the significance of the differences between different groups of parts of speech of the collocations. The test scores revealed that the number of students who correctly answered an item changes significantly according to different groups of parts of speech,  $H(2) = 6.958, p < .05$ . Since the test scores were significant, Mann-Whitney tests were used to find out the differences in difficulty between pairs of groups of parts of speech. Significant differences were found between V + N and Adj + N collocations ( $U = 196.5, p(2\text{-tailed}) < .05, r = -.39$ ) and between Adj + N and V + Prep collocations ( $U = 194.00, p(2\text{-tailed}) < .05, r = -.33$ ). It is worth stating that both effect size scores (i.e.  $r = -.39$  and  $r = -.33$ ) are moderate indicating that these significant differences found between

these collocation types are moderately large. However, no significant difference was found between V + N and V + Prep collocations ( $U = 309.00$ ,  $p$  (2-tailed)  $> .05$ ,  $r = -.325$ ).

The second step of the quantitative data analysis aimed to explore the relationships among frequency, MI, and frequency of components of the collocations and the number of the students answering an item correctly. To examine these relationships, non-parametric correlation was used, since both Kolmogorov-Smirnov and Shapiro-Wilk tests showed that the data in these variables were not normally distributed. It is worth noting that there are two types of non-parametric correlation, namely, Spearman's and Kendall's correlations. In the following analyses Kendall's correlation was used because more accurate generalizations can be drawn from Kendall's correlation than from Spearman's (Field, 2005). Table (4:1) shows the results of non-parametric correlations.

**Table 4.1: The results of non-parametric correlations**

	The number of students answering an item correctly
Frequency of the collocations	Correlation Coefficient .134
	Sig. (one-tailed) .047
MI of the collocations	Correlation Coefficient -.113
	Sig. (two-tailed) .156
Frequency of the 1 <sup>st</sup> component of the collocations	Correlation Coefficient .036
	Sig. (two-tailed) .654
Frequency of the 2 <sup>nd</sup> component of the collocations	Correlation Coefficient .084
	Sig. (two-tailed) .294

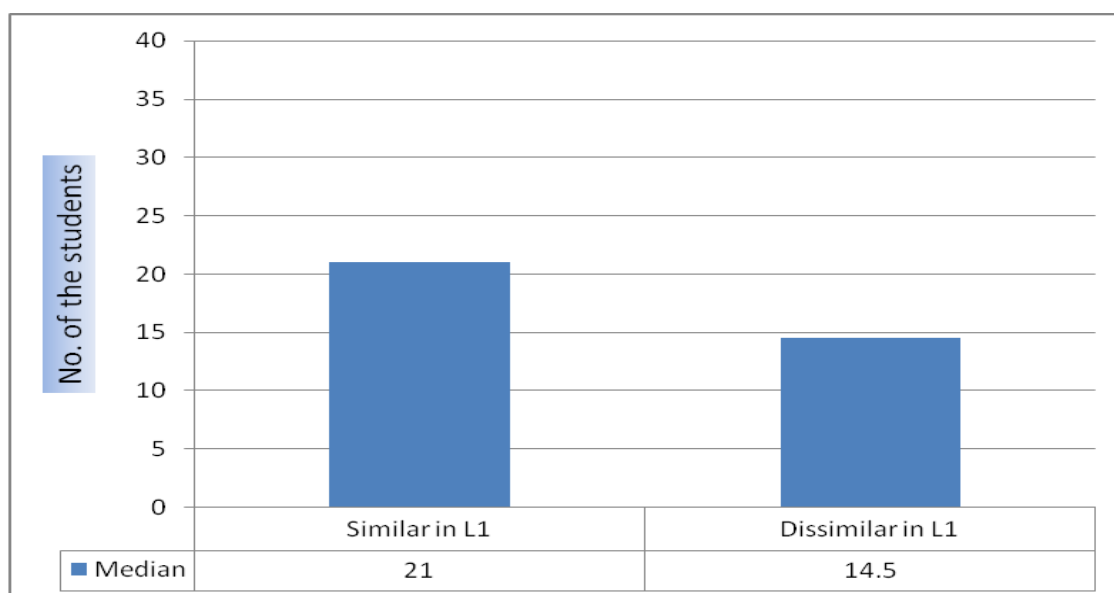
As shown in table (4.1), the frequency of the collocations correlated significantly with the number of students giving correct answers to an item,  $r = .134$ ,  $p$  (one-tailed)  $< .05$ . It is worth saying that the correlation coefficient score (.134) is very small indicating that there is only a weak relationship between these two variables. However, MI of the collocations ( $r = -.113$ ,  $p$  (two-tailed)  $> .05$ ), frequency of the first component of the collocations ( $r = .036$ ,  $p$  (two-tailed)  $> .05$ ), and frequency of the second component of the collocations ( $r = .084$ ,  $p$  (two-tailed)  $> .05$ ) did not significantly correlate with the number of students answering an item correctly.

We have seen so far that Adj + N are the hardest parts of speech for the learners, and that lower frequency collocations are more difficult than higher-frequency ones. However, it will be remembered from Chapter 3 (table 3.3) that Adj + N were the least frequent collocations in the corpus. With these facts in mind, it is not clear whether Adj + N collocations are the hardest because of their parts of speech or simply because of their lower frequency, or if both these factors have effects on the production of correct collocations. If the learners' problems with Adj + N collocations were due to part of speech, then it can be concluded that Adj + N collocations are the most problematic type for the learners, and that parts of speech could be a major source of collocational errors. However, if Adj + N were the most problematic collocations simply because of their lower frequency, then part of speech probably did not have effect on making errors mostly in this collocation type.

The final step of the quantitative data analysis aimed to find out whether the number of students who correctly answered an item changes according to whether the collocations are the same in L1. To address this issue, Mann-Whitney tests were used, as

both Kolmogorov-Smirnov and Shapiro-Wilk tests showed that the data in these two variables were not normally distributed. Descriptive statistics (see figure 2) showed that collocations that are similar in L1 were apparently easier, being answered correctly by a median of 21 out of 40 students than collocations which are dissimilar in L1 ( $Mdn = 14.5$ ).

**Figure 2: Median of the students who correctly answered the collocations which are similar and dissimilar in L1**



The results of Mann-Whitney tests revealed that the number of students answering an item correctly changes significantly according to whether the collocations are similar ( $Mdn = 21$ ) or dissimilar ( $Mdn = 14.5$ ) in L1,  $U = 513.00$ ,  $p$  (2-tailed)  $< .05$ ,  $r = -.23$ . However, the effect size score ( $r = -.23$ ) is rather low, indicating that this significant change is small.

#### The Results of Qualitative Data Analysis

The qualitative data were obtained from retrospective think-aloud protocols. For these protocols, 6 students were selected on the basis of their scores in the collocation

completion test (i.e. two top, two average and two poor scores). These students were interviewed and asked to give their reasons for choosing a particular option while completing the collocations in the collocation completion test. The interviews were conducted in the students' first language (i.e. Kurdish). Their speech was audio taped and their reasons were translated into English. The main questions asked to the students were the following:

- Why did you choose this particular option but not the others?
- What did you think while choosing this particular option?

Students' answers to these questions were analyzed to explore whether there is/are other possible main source(s) of errors in collocations. As a result, the researcher reached two main conclusions:

Firstly, it can be inferred from the data obtained from the interviewees that in many cases students were unable to choose the correct collocation from between synonyms. For instance, an answer of one interviewee was that "I was not sure about choosing *round* or *circular* with *face*; later, I decided to choose *circular*, because I thought that *circular face* is more suitable than *round face*". In this item, both *circular* and *round* are synonymous, and the correct collocation was *round face*. However, the interviewee made error in producing the correct collocation, because he substituted *round* with *circular* unintentionally. Another example was quoted from another interviewee, who said "in fact I am still uncertain whether this item is correct, because *get*, *achieve* and *obtain* [i.e. the distracters] are very similar in meaning, and I thought that all of them could be possible with *goals*. Later I decided to choose *obtain*".

The second conclusion drawn from the qualitative data analyses supports what has been already found in the quantitative data analyses. This conclusion stems from two main facts reflected in the interviewees' answers. First of all, all interviewees thought in their first language while choosing particular options. While giving answers to the questions, the interviewees clarified the options and sentences in their first language, and they tried to associate the meaning of the English words to the meaning of the Kurdish words. For instance, one of the answers of the interviewees was that "this option was not possible because if we choose this option the sentence would not have meaning in Kurdish". Collocations such as *ease pain*, *dry wine*, and *keep eye* are not possible in Kurdish; for this reason, the students chose *break* with *pain*, *bitter* with *wine* and *look* with *eye* because *\*break pain*, *\*bitter wine* and *\*look eye* are possible collocations in Kurdish. This attributes to the fact that the majority of the interviewees said "*high winds*, *dry wine*, *ease pain* and *keep eye* are not possible because they do not have meaning in Kurdish". Those students appear to have committed collocational errors just because they tried to choose the best option to find L1 equivalent collocations without thinking about the meaning of the collocations in English. Thus, the results of qualitative data analyses reconfirmed that the influence of L1 is one of the major problems for EFL students in producing correct collocations. Secondly, lack of exposure to collocations leads students to produce incorrect collocations. Some students who had not encountered some collocations produced them incorrectly. For instance, an interviewee said "this is the first time I see this collocation (i.e. *dry wine*); therefore, I did not know which of these options is correct. Later, I randomly chose *bitter* (i.e. *bitter wine*)". This student made an error in this collocation because he had not been exposed



to it before. However, collocations which had been often met before were produced correctly. For instance, when an interviewee was asked about a collocation that he had correctly answered, he said “I saw this collocation (i.e. *dramatic changes*) on TV; one day while I was watching BBC channel, I saw this collocation on the channel”. When another interviewee was also asked about a collocation that she had answered correctly, she said “I am 100% sure that this option was the right choice because one day while I was reading one of the literary subjects I encountered this “phrase” and I directly checked its meaning in a dictionary”. It can be inferred from these quotations that the frequency of collocations has a direct influence on producing correct collocations, that is, the more collocations are presented, the more they are correctly produced.

#### Conclusion

This chapter presented the results of both quantitative and qualitative data analyses. The quantitative data were obtained from the collocation completion test. Kruskal-Wallis and Mann-Whitney tests and non-parametric correlations were used to analyze the data. The qualitative data, on the other hand, were obtained from the interviewees’ responses to the interview questions and were analyzed to find out whether there is/are other possible main source(s) of collocational errors. In the following chapter, the findings confirmed in the data analyses procedures will be discussed.

## CHAPTER V: DISCUSSION AND CONCLUSION

### Introduction

The purpose of this study was to explore the main sources of collocational errors made by EFL learners at Koya University. Both the quantitative and qualitative data obtained from the collocation completion test and think-aloud protocol, respectively, were analyzed to address this issue. This chapter consists of some sections in which discussion of the findings, pedagogical implications, limitations of the study, suggestions for further studies and overall conclusion are presented.

### Discussion of the Findings

The findings obtained from the results of the analyses were discussed in terms of both the quantitative and qualitative data. The first major finding was that Adj + N collocations were the most difficult type for the participants. This is evidenced by the fact that the median number of students answering Adj + N items correctly was significantly lower than that for either V + Prep items or V + N items. At the same time, no significant difference was found between V + Prep and V + N items. It is worth noting that Adj + N collocations were the least frequent type in both the collocation completion test and language. Therefore, it should be kept in mind that learners may not know some particular collocations just because of their low frequency. Thus, it is not obvious whether the students' Adj + N errors are due to part of speech or to low frequency.

Some previous studies have also concluded that Adj+N collocations were the most difficult type for the learners. In Boonyasaquan's (2005) study, the Thai

participants were given a translation task and the results showed that Adj + N collocations were the most frequent type that was violated by the participants. Boonyasquan (2005) attributed these collocational violations mostly to the participants' indulgence in over-literal translation. In Chang's (1997 cited in Hsueh, 2005) study, English compositions written by college freshmen were investigated. The focus of the investigation was on various types of lexical and grammatical collocations such as Adj+N, V+N, Prep+N and V+Prep. From the results, many collocational errors were found, but Adj + N collocations had the highest frequency of error, that is, they were considered as the most difficult type. V+Prep collocations were found to be the easiest type due to having the lowest frequency of error.

Interestingly, the finding of Kuo (2009) regarding collocation types goes counter to the findings of these studies. Kuo (2009) collected data from 98 free writing samples written by 49 intermediate college students majoring in English. The results showed that the participants made many collocational errors, but that Adj + N collocation type was easier for the students than V + N type. This contradictory finding can perhaps be attributed to the fact that in Kuo's (2009) study, while identifying the V + N collocations, the central focus was on de-lexical verbs such as *make*, *do*, *give* and *get*. In fact, these verbs are more likely to be misused by the students because these verbs can form various collocations and each verb can have varied meanings when it is combined with different words. For instance, in each of the following collocations, *make* has a different meaning: *make a mistake* (*make* = doing something), *make tea/breakfast* (*make* = prepare a drink or food), *make money/a profit* (*make* = to earn money), *make a hole/mark* (*make* = to cause something). However, while identifying Adj + N

collocations, this criterion was not applied. Given this, it is not surprising that the participants of Kuo's (2009) study had more problems with V + N collocations.

All in all, collocations may differ in difficulty according to their parts of speech, that is, the number of correct collocations produced probably changes according to different groups of parts of speech. The reason behind the ambiguity of this factor is that Adj + N collocations were the least frequent type in both the test and language. Therefore, it is hard to conclude that parts of speech were different in difficulty, because it is not clear whether this difficulty resulted from the difference in parts of speech or the lower frequency. If Adj + N collocations are the most difficult type for EFL learners, they can be helped by providing descriptive language materials such as descriptive essays and dialogues in order to help the learners to be exposed to as many Adj + N collocations as possible. As a result, the students' knowledge of Adj + N collocations will possibly be increased.

The second major finding was that frequency of the collocations was significantly correlated with the number of students answering an item correctly. From this analysis, it can be inferred that the high frequency of collocations leads learners to retain these collocations. Interestingly, the results of the qualitative data analyses also confirmed this. In these analyses, it was concluded that the students produced correctly those collocations that had previously been encountered. According to the answers of the interviewees, some students made errors in collocations because the students had not encountered these collocations before. However, correct collocations were produced as a result of high exposure to them. Therefore, high frequency of collocations leads students to get these collocations right. This is also claimed by Durrant (2008). One of the major

findings of his thesis was that adult learners can learn collocations from input that provides repeated exposure to collocations, and that frequent exposure to collocations can dramatically improve learning. Learners are likely to learn collocations that are encountered regularly during learning sessions. If the students are often exposed to collocation, they often recognize it and automatically get it right because they have already stored it in their minds and they use it as a ready-made chunk whenever is needed. From this, it seems that low exposure to collocations or input including collocations lead to collocational unawareness, which causes insufficiency of collocational knowledge; as result, learners make many collocational errors. Therefore, it can be concluded that low frequency of collocations is a major source of errors in collocations.

The third major finding was that the number of students who correctly answered an item changes significantly according to whether the collocations are similar or dissimilar in L1. From this analysis, it can be seen that the collocations which are similar in L1 were easier for the students. This can be attributed to the fact that collocations that are similar in L1 can be translated into L2 without resulting in any errors. Similarly, Baker (1992) stated that collocations that carry similar meaning in learners' L1 lead students to rarely make mistakes in these collocations. Previous research has also reached the conclusion that L1 is an effective factor that leads students to make errors in collocations (e.g. Nesselhauf, 2003; Jing, 2008; Mahmoud, 2005, Zinkgraf, 2008; Koosha & Jafarpoor, 2006; Boonyasquan, 2005; Huang, 2001; Bahns & Eldaw ,1993; Sadeghi, 2009). For instance, Sadeghi (2009) indicated that differences between L1 and L2 collocations lead substantially to errors in producing L2 collocations. Nesselhauf

(2003) concluded that the effect of L1 on the production of L2 collocations is relatively high, since in her study, 56% of collocational errors in L2 written production was attributed to L1 interference. In contrast to these, in the study of Kuo (2009), the L1 did not have such an influential impact on the collocations, as negative transfer had the lowest ratio. According to Kuo (2009), this finding could be attributed to the students' awareness of the L1 and L2 differences and the effect of collocation instruction which focused on bilingual collocations. The results of the qualitative data analyzed in the present study also confirmed that L1 interference affected the students' correct answers. This can be seen from the fact that the interviewees during the interviews explained the meaning of the items in their first language, and they tried to make a connection between the meanings of English and Kurdish collocations. For instance, the students preferred to choose *\*break pain*, *\*bitter wine* and *\*look eye* instead of *ease pain*, *dry wine* and *keep eye*, because the former collocations are the possible ones in Kurdish, not the latter ones. Those students appear to have committed collocational errors just because they tried to choose the best option to find L1 equivalent collocations without thinking about the meaning of the collocations in English. If learners do not know the target collocation, the first strategy they rely on is mostly their L1, because usually students think in their first language while producing their L2 without being aware of whether this is acceptable in L2. Therefore, students should be made aware of the differences between the L1 and L2; otherwise, students often make errors in collocations. Thus, it was concluded from the present study that the influence of L1 is another main source that leads students to commit collocational errors.

A fourth finding was that a major type of error involved students' substituting one of the collocation components for its synonym. Previous researchers (e.g. Jing, 2008; Boonyasquan, 2005; Farghal & Obiedat, 1993; Park, 2003; Kuo, 2009) have concluded that many EFL learners' collocational errors resulted from using the synonym of the components of the collocations. Similarly, in the present study, some incorrect collocations were produced because the learners chose the synonym of the target (i.e. correct) collocate. For instance, incorrect collocations such as *\*circular face* and *\*obtain goals* were produced because the participants substituted *round* for *circular* and *achieve* for *obtain*, as the correct collocations were *round face* and *achieve goals*. Synonyms are words that are similar (Jing, 2008) or close (Palmer, 1981) in meaning; however, when they co-occur with different words they can form various collocations which are different in meaning. In the current study, the learners seem to be not aware of that synonyms can have varying collocational restrictions. As mentioned in Chapter 2, collocation consists of base and collocate where the choice of the collocate depends on the base. For instance, in a collocation such as *round face*, the choice of the adjective (i.e. *round*) depends on the noun (i.e. *face*), since the noun is the base. This shows that the learners were not conscious about the fact that *face* requires an adjective that has to describe physical appearance. Instead, they chose an adjective *circular*, which is not used to describe physical appearance; rather, it is used with some nouns such as *motion*, *argument*, *flow*, *letter* and *orbit* (Longman Dictionary of Contemporary English, 2003). For instance, in English, it is possible to say *round face*, *round head* and *round neck*, but not *\*circular face*, *\*circular head* and *\*circular neck*.

This type of error (i.e. replacing the collocates with their synonyms) could also be attributable to the teachers' tendency to teach words individually rather than collocationally. This may have led the learners to memorize many words with many synonyms without being conscious about the usage of these words. Some solutions for this problem could be that learners should be alerted to the fact that words have various collocational restrictions. Additionally, they need to learn that components of collocations cannot be replaced by other words even if they are synonyms. Further, collocations should be taught as an indivisible part of language. To do this, teachers should be encouraged to use the idiom principle, that is, to teach words in phrases and chunks. This is also claimed by Farghal and Obiedat (1993) and Liu (1999 cited in Kuo, 2009), who state that learners [and teachers] should rely more on the idiom principle because the accumulation of such a principle in vocabulary will decrease the amount of collocational errors.

A fifth finding was that no significant correlation was found between the MI of the collocations and the number of the students answering an item correctly. This could be because MI is not linked to the knowledge of collocations; rather it is about to what extent a strong relationship exists between the components of collocations. Thus, it can be inferred from this that the MI of collocations is not an important factor affecting correct production of collocations and that it is not such a good predictor of what collocations the students know.

The last finding was that no significant relationship was found between the frequency of the components (i.e. both the first and second components) of the collocations and the number of the students answering an item correctly. From this it can



be inferred that the frequency of the component words of the collocations did not affect the students' correct answers, and that knowledge of collocations was not linked to the frequency of their components. Therefore, students are likely to make collocational errors even if the component words are frequent. Zingraf (2008) in her study concluded that the 1000 most frequent words in English appeared in atypical collocations produced by advanced EFL learners. This shows that the component words of the collocations were very frequent, but the collocations were incorrect. From these it can be concluded that knowledge of collocations is far behind the frequency of their components. For this reason, frequency of component words of collocations is another factor that does not affect producing correct collocations.

In conclusion, the results of both the quantitative and qualitative data analyses showed that the learners' collocational errors resulted from two major sources, namely, low frequency of collocations and the influence of L1. Parts of speech could also be another major source if the lower frequency of Adj + N collocations in the collocation completion test and language is not taken into account. However, factors such as MI and frequency of the components of the collocations were found ineffective in the production of correct collocations because they do not cause collocational errors and they are not linked to the learners' collocational knowledge.

#### Pedagogical Implications

The results of the present study showed that even seniors majoring in English make errors in collocations. Therefore, this study confirms the conclusions of numerous studies in which the authors have reached a consensus that collocations do merit special attention in language teaching.

All collocations are necessary for native-like language production. However, since collocations are too numerous to list, the focus should be on some particular collocations. According to the results of the present study, the majority of the students answered V + N and V + Prep collocations correctly; whereas Adj + N were only answered correctly by the minority. However, Adj + N collocations were unintentionally the least frequent type in the test. This probably was a factor leading the learners to make errors mostly in this type. Thus, regardless of the lower frequency of Adj + N collocations, the central focus should be on this particular part of speech, because they were the hardest for the participants. However, V + N and V + Prep should not be ignored, as some participants also had problems with these collocations.

In the present study, it was confirmed that collocations that are congruent in the L1 are easier for learners because the majority of the participants answered congruent collocations correctly. For this reason, there should also be a focus on non-congruent collocations. A stronger claim was made by Bahns and Eldaw (1993) who stated that for teaching collocations, L1 equivalent collocations can be ignored because learners can automatically get these collocations right. Differences between L1 and L2 collocations lead learners to make substantial errors in collocations (Sadeghi, 2009). Therefore, learners should be alerted to these differences; otherwise, in spite of having learnt the correct collocations, learners are likely to produce L1 equivalent collocations. For instance, it will be pointless to teach Kurdish EFL learners collocations such as *make mistake* and *firm friends* without alerting the learners to the fact that *\*do mistake* and *\*near friends* (Kurdish equivalent collocations) are not possible in English. One of the recommendations for avoiding the risk of these differences is that teachers can

encourage students to use bilingual collocation dictionaries. This is also claimed by Bahns (1993) and Nesselhauf (2003), who stated that providing collocation instruction focusing on bilingual collocations can help students to avoid collocational errors caused by the influence of mother tongue.

It would be highly desirable to create a list of target collocations for learners to study which takes into account the various factors which appear to influence their difficulty, because compiling such a list may help the learners to overcome many problems that the learners have with collocations. Such a list should be compiled on the basis of some certain principles. As the first major principle, focusing on specific parts of speech is highly recommended. As shown in the current study, regardless of their lower frequency, Adj + N were the most problematic parts of speech for the learners. For this reason, incorporating these problematic parts of speech possibly helps the learners to overcome many problems the learners have with this type of collocation. The second main principle should be that the intended collocations should not be congruent with the L1. Many researchers claim that non-congruent collocations often cause problems (e.g. Nesselhauf, 2003; Huang, 2001), and that congruent collocations should be ignored because learners do not commit errors in these collocations (e.g. Bahns & Eldaw, 1995). Similarly, the present study confirmed that non-congruent collocations were more problematic than congruent ones; therefore, the main focus should be on non-congruent collocations. For these reasons, including non-congruent collocations in the future collocation list could help learners to decrease the amount of their errors resulted from collocations that are not congruent with the L1. The last major principle recommended by the present study is that low frequency of the collocations should be

highly considered. While selecting infrequent collocations, it should be kept in mind that a careful balance needs to be maintained between choosing collocations which are infrequent enough to be difficult but frequent enough to be useful. For instance, collocations such as *wear perfume* (Freq. 79/ 4 million words), *controversial figure* (Freq. 102/4 million words) and *rapid decline* (Freq. 118/4 million words) seem to be infrequent collocations which often cause difficulty for learners. However, these collocations are likely to be useful because they are often used in the learners' daily life. Including this kind of infrequent collocations in such a list helps the learners to be exposed to these collocations; as a result, the learners' collocational knowledge would be increased. Thus, compiling a list of target collocations for learners to study taking into account these major principles may solve a lot of problems that EFL learners have with collocations.

In conclusion, to enhance EFL learners' knowledge of collocations, students should be presented with a lot of language teaching materials including the problematic collocations, particularly Adj + N type as a central subject matter, and students should be alerted to the L1 and L2 differences. Furthermore, teachers should provide various collocation tasks, activities and assignments with regard to authentic language materials, culture and opportunities for practicing collocations in order to help learners gain an understanding of words associations, specifically collocations.

#### Limitations of the Study

In the present study some limitations were identified. Firstly, the participants were small in number and limited to a particular university. Therefore, the results

probably could not be generalized to all EFL learners, and the instructional setting could not represent all teaching and learning situations.

Secondly, due to semantic and syntactic differences between Kurdish and English languages, some distracters, especially prepositions in the collocation completion test, did not have the exact equivalent translation in the target language.

Thirdly, the qualitative data were obtained from few selected participants; better results could have been achieved if the data had been collected from all participants.

Finally, it was not possible to distinguish the effects of frequency from the effects of parts of speech. Therefore, no precise information about the effects of parts of speech on the production of correct collocations was obtained.

#### Suggestions for Further Studies

The following ideas could be used in further studies in the domain of collocations:

1. In this study, the focus was on some specific types of collocations (i.e. V+N, Adj+N and V+Prep); further studies can consider EFL learners' sources of errors in other types of collocations such as Adv+Adj, N+Prep and N+V. As a result, we can gain a deeper understanding of what leads EFL learners to make errors in collocations.
2. In the present study, the students made errors in L1 congruent collocations; in future research, the difference in difficulty in L1 equivalent collocations or the reasons for committing errors in L1 congruent collocations could be investigated.
3. Since this study was limited to a small group of a particular proficiency level at Koya University, learners from different proficiency levels at various universities in Northern Iraq could participate in further research. Consequently, detailed

information about Kurdish EFL learners' collocational knowledge would be achieved.

4. The aim of conducting further research could be compiling a list of target collocations for EFL students to study with taking into account some basic factors that substantially cause collocational errors.
5. Further research could work to provide precise information about the effects of parts of speech on the production of correct collocations.

### Conclusion

The objective of the present study was to explore the main sources of collocational errors made by EFL learners at Koya University in Northern Iraq. The participants were Kurdish EFL seniors. For data collection, two different instruments were used. The first instrument was the collocation completion test, which was used to explore the main sources of the learners' collocational errors. A think-aloud protocol, which was the second instrument was used to find out whether there is/are other possible main source(s) of collocational errors. The data obtained were analyzed to address these issues. The results of both the quantitative and qualitative data analyses revealed many results. Firstly, a significant relationship was found between frequency of the collocations and the students' correct answers. Secondly, Adj + N collocations were the most difficult type for Kurdish EFL students. Thirdly, the number of students answering an item correctly changed significantly according to whether the collocations are similar in L1. Fourthly, the substitution of the collocate for its synonyms was a type of collocational errors produced by the learners. Fifthly, no significant correlations were found between MI and frequency of the components of the collocations and the

students' correct answers. The findings confirmed that low frequency collocations and L1 interference are major sources that lead the learners to make collocational errors. Since it was impossible to distinguish the effects of part of speech from the effects of frequency in the collocation completion test, it is not clear whether part of speech have any effects on producing correct collocations. The findings also showed that the learners' collocational error often involve use of synonym(s) of the collocates. Furthermore, MI and frequency of the components of collocations are not important factors in the production of correct collocations because these factors did not cause collocational errors. With these findings in mind, materials designers and teachers, especially those in Northern Iraq can provide better language teaching materials and activities relating to collocations.

## REFERENCES

- Bahns, J. (1993). Lexical collocations: a contrastive view. *ELT*, 47(1), 56-63.
- Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocation? *System* 21(1), 104-114.
- Baker, M. (1992). *In other words: a course book on translation*. London: Routledge.
- Benson, M., Benson, E., & Ilson, R. (1986a). *Lexicographical description of English*. Amsterdam: John Benjamins.
- Biskup, D. (1992). L1 influence on learner's renderings of English collocations: A Polish/German empirical study. In Arnaud and Béjoint (Eds.). *Vocabulary and Applied Linguistics* (pp. 85-93). Caroli, M. T: Macmillan.
- Boonyasaquan, S. (2005). An analysis of collocational violations in translation. *Journal of Humanities*, 27(2), 79-91.
- Boonyasaquan, S. (2006). The lexical approach: An emphasis on collocations. *Journal of Humanities*, 28(1), 98-108.
- Brown, D. (1974). Advanced vocabulary teaching: the problem of collocation. *RELC Journal*, 5(2), 1-11.
- Carter, R. & McCarthy, M. (1988). *Vocabulary and language teaching*. New York: Longman.
- Coady, J. & Huckin, T. (Eds.) (1997). *Second language vocabulary acquisition*. Cambridge: Cambridge University Press.
- Conzett, J. (2000). Integrating collocation into a reading and writing course. In M. Lewis (Ed.), *Teaching collocation: further developments in the lexical approach* (70-87). London: Language Teaching Publications.
- Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2, 223-235.
- Durrant, P. (2008). High frequency collocations and second language learning. A thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy.
- Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33-68). Cambridge: Cambridge University Press.



Fan, M. (2009). An exploratory study of collocational use by ESL students-a task based approach. *System*, 37, 110-123.

Farghal, M. & Obiedat, H. (1995). Collocations: a neglected variable in EFL. *International Review of Applied Linguistics in Language Teaching*, 33(4), 315-331.

Field, A. (2005). *Discovering statistics using SPSS*. London: SAGE publications.

Gass, S. M. & Selinker, L. (2001). *Second language acquisition: An introductory course*. NJ: Lawrence Erlbaum Associates.

Hill, J. & Lewis, M. (Eds.) (1997). *Dictionary of selected collocations*. London: Language Teaching Publications.

Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.). *Teaching collocation: Further developments in the lexical approach* (pp. 47-67). London: Language Teaching Publications.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24-44.

Hsu, J. (2007). Lexical collocations and their relationship to the online writing of Taiwanese college English majors and non-English majors. *Electronic Journal of Foreign Language Teaching*, 4(2), 192-209.

Hsueh, S. (2005). A study on the relationship between collocations and English writing. *Kaohsiung journal*, 8(98), 161-184.

Huang, L. (2001). Knowledge of English collocations: An analysis of Taiwanese EFL learners. *Texas Foreign Language Education Conference*, 6(1), 113-132.

Jaaskelainen, R. (2002). Think-aloud protocol studies into translation. *Target*, 14(1), 107-136.

Jing, M. (2008). Erroneous collocations caused by language transfer in Chinese EFL writing. *US-China Foreign Language*, 6(9), 57-61.

Kjellmer, G. (1990). A mint of phrases. In K. Aijmer and B. Altenberg (Eds.), *English corpus linguistics* (111-127). London: Longman, Harlow.

Koosha, M. & Jafarpour, A. A. (2006). Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal*, 8(4), 192-209.

Kuo, C. (2009). An analysis of the use of collocation by intermediate EFL college students in Taiwan. *ARECLES* 6, 141-155.

- Lee C. & Liu J. (2009). Effects of collocation information on learning lexical semantics for near synonym distinction. *Computational Linguistics and Chinese Language Processing* 14(2), 205-220.
- Lennon, P. (1996). Getting 'easy' verbs wrong at the advanced level. *IRAL* 34(1), 23-36.
- Lewis, M. (1993). *The Lexical Approach: the state of ELT and a way forward*. London: Language Teaching Publications.
- Lewis, M. (1997). *Implementing the lexical approach*. England: Language Teaching Publications.
- Lewis, M. (Ed.). (2000). *Teaching collocation: Further developments in the lexical approach*. London: Language Teaching Publications.
- Longman Dictionary of Contemporary English* (4<sup>th</sup> ed.). (2003). England: Pearson Education Limited.
- Mahmoud, A. (2005). Collocation errors made by Arab learners of English. *Asian EFL Journal, Teaching Articles*, 5/2, 117-126.
- McCarthy, M. & O'Dell, F. (2005). *English collocations in use*. Cambridge: Cambridge University Press.
- McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.
- Meara, P. (1984). The study of lexis in interlanguage. In A. Davies, C. Criper, & A. Howatt (Eds.). *Interlanguage* (225- 235). Edinburgh: University of Edinburgh Press.
- Mel'cuk, I. (1998). Collocations and lexical functions. In Cowie. A. P. (Ed.). *Phraseology: Theory, analysis and application* (pp. 23-53). New York: Oxford University Press.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nattinger, J. R. & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford 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. (2004). What are collocations? In D. J. Allerton, Nadja Nesselhauf and Paul Skandera (Eds.). *Phraseological units: Basic concepts and their application* (pp. 1-21). Basel: Schwabe.
- Nesselhauf, N. (2005). *Collocation in learner corpus*. Amsterdam: John Benjamins B.V.
- Palmer, F. R. (1981). *Semantics* (2<sup>nd</sup> ed.). Cambridge: Cambridge University Press.
- Park, S. (2003). Lexical collocation use by Korean EFL college learners. *Seoul National University Department of English Language and Literature, SNU Working Papers in English Linguistics and Language 2*, 19-40.
- Pawley, A. & Syder, F. H. (1983). Two puzzles for linguistic theory: Native like selection and native like fluency. In J. C. Richards and R. W. Schmidt (Eds.), *Language and communication* (pp. 191-225). London: Longman.
- Rankin, J. Mark (1988). Designing thinking aloud studies in ESL reading. *Reading in a Foreign Language 4*(2), (119-132).
- Richards, J. C. & Rogers, T. S. (2001). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.
- Sadeghi, K. (2009). Collocational differences between L1 and L2: Implications for EFL learners and teachers. *TESL Canada 2*(62), 100-124.
- Sinclair, J. (1991). *Corpus concordance collocation*. Oxford : Oxford University Press.
- Smadja, F. A. (1989). Lexical co-occurrence: The missing link. *Literary and Linguistic Computing, 4*, 163-168.
- Woolard, G. (2000). Collocation- encouraging learner independence. In M. Lewis (Ed.). *Teaching collocation: Further developments in the lexical approach* (pp. 28-46). London: Language Teaching Publications.
- Yang, Y. & Hendricks, A. (2004). Collocation awareness in the writing process. *Reflection of English Language Teaching 3*, 51-78.
- Zinkgraf, M. (2008). V + N miscolllocations in the written production of university level students. *Estudios de Lingüística Inglesa Aplicada, 8*, 91-116.

## APPENDICES

## Appendix A: Collocation Completion Test

NAME: .....

GENDER: .....

GIVEN TIME: **1 HOUR**

.....

**Please circle the most appropriate option to complete the following sentences. For each item, please choose ONLY ONE option.**

1. Like all other mammals, whales ..... air and give milk.  
A) receive      B) take      C) consume      D) breathe
2. Kurda is so calm that he never ..... his temper.  
A) loses      B) misses      C) becomes      D) lacks
3. I asked my neighbor to ..... an eye on my house.  
A) keep      B) look      C) store      D) hold
4. Teachers are cautioned to use great care in writing test items because tests ..... time if they are not effective for evaluation.  
A) spend      B) mislay      C) waste      D) lose
5. Morphine is a kind of drug used to ..... pain.  
A) decrease      B) ease      C) quit      D) break

6. The lawyer took some pictures of the crime scene in order to ..... evidence for the court.
- A) make            B) gather            C) form            D) raise
7. The company plans to ..... internet access for its customers.
- A) supply            B) establish            C) provide            D) bring
8. Redgrave has ..... two gold medals and he is one of the most successful current Olympic sportsmen in Britain.
- A) achieved            B) succeeded            C) won            D) received
9. Most researchers start from the assumption that it is morally wrong to ..... research on people who do not know that they are being studied.
- A) write            B) investigate            C) conduct            D) make
10. They need 4 million dollars this year to ..... the budget.
- A) stabilize            B) steady            C) balance            D) equalize
11. In Twelfth Night, William Shakespeare ..... light on different types of love.
- A) sheds            B) puts            C) underlines            D) flashes
12. One of the reasons for sending children to kindergartens is to help them..... knowledge about interaction with other people.
- A) get            B) learn            C) achieve            D) gain
13. The referee glanced at his watch and ..... the whistle for half time.
- A) hit            B) blew            C) breathed            D) exhaled
14. Negotiating is an activity that seeks to ..... agreement between two or more different starting positions.
- A) arrive            B) reach            C) meet            D) attain

15. Every year on March 16, Kurdish people ..... candles in commemoration of 5000 martyrs in Halabja.
- A) fire            B) light            C) switch on            D) torch
16. The organizations need to find the best methods to ..... their goals.
- A) get            B) achieve            C) meet            D) obtain
17. Robinho, who is a Brazilian football player, quickly ..... the attention of several clubs.
- A) attracted            B) pulled            C) tempted            D) bring
18. The earthquake ..... great damage in Haiti.
- A) caused            B) created            C) let            D) helped
19. Jim has no sense of humor and cannot ..... jokes.
- A) say            B) do            C) tell            D) speak
20. Sarah is trying to draw a picture in order to ..... her message.
- A) express            B) say            C) communicate            D) convey
21. The Secretary of State for Industry said that robots would ..... jobs in Britain, when he opened an unmanned factory last November.
- A) spread            B) create            C) make            D) build
22. The government is planning to ..... some money to homeless children.
- A) present            B) spread            C) award            D) donate
23. His girlfriend likes dancing and would like to ..... a horse.
- A) drive            B) get on            C) ride            D) steer

24. Her parents are working hard, because they have to .....bills for themselves and five kids.
- A) spend            B) stump up            C) pay            D) tip
25. His mom ..... birth to a baby boy.
- A) made            B) gave            C) donated            D) had
26. At the party there were different types of wine; many people preferred sweet wine, but we preferred ..... wine.
- A) bitter            B) dry            C) tasteless            D) parched
27. Nina has ..... face, pale skin and short-cut hair.
- A) a circular            B) an elliptical            C) a round            D) an egg-shaped
28. A witness who saw the incident described the driver as white, about 25 years old, and of slim build with ..... shoulders.
- A) long            B) broad            C) rough            D) distant
29. A free-fighting match is held under ..... rules.
- A) strict            B) harsh            C) exact            D) tight
30. It is a beautiful night, a full moon and a few ..... stars against the black sky over the farm.
- A) light            B) bright            C) colorful            D) sparkling
31. Hurricane damage through ..... winds and tidal surges causes an immense amount of destruction and poses a major threat to many coastal communities in the USA.
- A) shrill            B) fast            C) high            D) soaring

32. Hungary's capital Budapest symbolizes the ..... changes in the country since the communist collapse.
- A) influential      B) exotic      C) dramatic      D) strong
33. Priority for assistance will be given to senior citizens, disabled and ..... parent families.
- A) single      B) alone      C) separated      D) bachelor
34. During recent ..... disasters such as earthquakes and floods, the Federal Emergency Management Agency has quickly sent out relief checks to thousands of residents.
- A) divine      B) authentic      C) natural      D) real
35. She suffered a ..... injury and was in a coma for three months after the accident.
- A) somber      B) serious      C) difficult      D) solemn
36. He gave her a ..... glance and smiled.
- A) fast      B) sudden      C) quick      D) single
37. It was believed that the accident happened as an indirect result of ..... rain and snow storms in the city.
- A) weighty      B) heavy      C) tough      D) strong
38. The nearest supermarket is next door where you can buy different types of ..... drinks such as Pepsi and Orange juice.
- A) chilly      B) velvety      C) soft      D) gassy
39. On the day of our departure all our ..... friends came to the railway station to see us off.
- A) near      B) approachable      C) close      D) nearby



40. The interaction between members of the ..... family is likely to be important especially in eastern societies.
- A) extended      B) detailed      C) complete      D) full
41. From the 1950s to the 1990s radical changes in teaching styles reflect ..... changes in social and cultural values.
- A) major      B) underlying      C) finest      D) optimum
42. According to the ..... wisdom, successful presidential candidates had to come from the middle of the political spectrum.
- A) routine      B) conformist      C) conventional      D) old-fashioned
43. We should work for a global ban on chemical and ..... weapons in order to reduce the risk of genocide.
- A) bacterial      B) fatal      C) evolutionary      D) biological
44. The number of Pandas and white Tigers became so depleted that they were placed on the ..... species list.
- A) risky      B) under-served      C) nonexistent      D) endangered
45. Each year in Birmingham, The National Exhibition Centre hosts a lot of ..... scale trade and public exhibitions, covering the whole spectrum of industry, commerce and leisure.
- A) large      B) robust      C) broad      D) bulky
46. Graduates in the ..... sciences are well equipped to enter a large variety of occupations after leaving the University.
- A) national      B) public      C) social      D) human

47. Kurds are ..... part of forming new federal Iraq.  
A) a historic      B) a prime      C) an integral      D) a landmark
48. The courts have ruled that school boards can impose ..... sanctions on teachers who go on strike.  
A) financial      B) fiscal      C) monetary      D) economic
49. It was their second get-together, and the two have now become ..... friends.  
A) solid      B) near      C) tight      D) firm
50. I usually squeeze lemon over sandwiches and I add .....pepper to them.  
A) ground      B) earthy      C) squashed      D) smashed
51. For the school competition, our teacher chose Lewis and Jim because she knows them personally and can vouch ..... their reliability.  
A) to      B) towards      C) for      D) from
52. Last year she came to the school for the first time, and every eye turned to gape ..... her long red hair and golden earrings.  
A) at      B) on      C) to      D) in
53. The Secretary of State can insist ..... changes to any plans that do not make sufficient use of independent suppliers.  
A) of      B) towards      C) upon      D) to
54. Telling lies is going to detract ..... your personality.  
A) against      B) in      C) from      D) for
55. The girl next door winked ..... me and smiled.  
A) for      B) at      C) about      D) to

56. More than 30 police clashed ..... rival gangs when they wanted to steal goods from a supermarket.
- A) With regard to      B) along with      C) towards      D) with
57. In general, males tend to compete ..... one another for females.
- A) versus      B) against      C) along with      D) along
58. It's important to differentiate ..... fact and opinion.
- A) of      B) beside      C) between      D) in
59. Carnival gives the poor a chance to mingle ..... the rich.
- A) to      B) at      C) with      D) from
60. With the help of a detective, she had to sift ..... the papers on every desk and the rubbish in every drawer.
- A) into      B) via      C) among      D) through
61. The governor warned ..... thieves at stations and advised that possessions should not be left near carriage windows.
- A) towards      B) versus      C) against      D) regarding
62. If my children are rude, that reflects ..... me as a parent.
- A) of      B) over      C) above      D) upon
63. Poetry doesn't usually translate well ..... another language.
- A) through      B) to      C) into      D) for
64. The first recipient of the original scholarship was Norcross Burrowes in 1880, who went on to graduate ..... the Victoria University of Manchester in 1884.
- A) from      B) in      C) at      D) of

65. He often complains ..... not being appreciated at work.  
A) into                      B) about                      C) for                      D) towards
66. Few political interest groups are transformed ..... successful political parties.  
A) for                      B) through                      C) within                      D) into
67. Anxiety can interfere ..... children's performance at school.  
A) about                      B) with                      C) within                      D) along with
68. People differ ..... one another in their ability to handle stress.  
A) from                      B) of                      C) at                      D) to
69. The immune system interacts ..... both the nervous system and the hormones.  
A) into                      B) within                      C) together (prep.)                      D) with
70. Afro-Americans account ..... 12% of the USA population.  
A) of                      B) to                      C) for                      D) at
71. Doctors are aiming to concentrate more ..... prevention than cure.  
A) on                      B) onto                      C) at                      D) for
72. Many English words are derived ..... Latin.  
A) for                      B) of                      C) from                      D) at
73. Some countries such as Turkey and Cyprus depend ..... tourism for much of their income.  
A) on                      B) of                      C) over                      D) onto
74. I am really worried ..... her small brother; he has been missing since last night.  
A) concerning                      B) regarding                      C) about                      D) of
75. Many advertisements try to associate drinking ..... fun.  
A) to                      B) with                      C) at                      D) along with

## Appendix B: Transcript of a Sample Student's Think-aloud Protocol in English

The researcher: good morning!

The student: good morning

The researcher: I would like to ask you some questions regarding the test you took. Are you ready?

The student: yes, I am ready.

The researcher: in the first item, you chose the option "breathe"; why did you choose this?

The student: first I looked at the word after the blank (i.e. air), then I realized that animals need air to breathe.

The researcher: so, you chose this option on the basis of the word after the blank.

The student: yes.

The researcher: what about the second item? What did you think while choosing this option?

The student: in fact I do not know what "temper" means. Therefore, I looked at the words in the sentence, and I saw "calm" in the sentence and I realized that when someone is "calm", s/he does not lose something. As a result, I chose "lose".

The researcher: can you tell me why did you choose "look" for this item?

The student: because in this sentence, someone asks his neighbor to look at his house while he is not at home.

The researcher: as you see, in these items you chose "circular" "exact" and "near"; can you tell me the reason behind choosing these options?

The student: ok. In this item, I chose “circular” because I saw the word “face” after the blank. So, I realized that this sentence is about describing someone who has a “circular face”. For the following item, I chose “exact” because the rules have to be “exact”.

About this item which says “on the day of our departure all our .....friends, blab la...”, I chose “near” for the blank because according to the sentence, someone leaves a place and all his “near friends” are with him to say goodbye.

The researcher: let’s take a look at these items which requires choosing correct prepositions. Why did you choose these prepositions?

The student: to be honest, I am not good at prepositions. Therefore, I chose most of these prepositions at random. However, there were some prepositions such as “upon” “about” and “from” I chose correctly.

The researcher: how did you know that they are the correct options?

The student: because there were the verbs “insist”, “complain” and “differ” that requires prepositions “upon”, “about” and “from”, respectively, and I have seen these verbs with these prepositions many times.

The researcher: thank you for your time.

The student: you are welcome.

## Appendix C: Transcript of a Sample Student's Think-aloud Protocol in Kurdish

تویژەر: به یانیت باش

خویندکار: به یانیت باش

تویژەر: ده مهووت هه ندیك پرسیرت لیبكه م له باره ی نه وتیسته ی كردمان. تۆناده ی ؟

خویندکار: به ئی نامادم

تویژەر: له برگه ی یه كه م تو ( breathe ) ت هه ئبژاردوو، بوچی نه وت هه ئبژارد؟

خویندکار: له پیش هه مو شندا من سه یری وشه ی دوا ی بوشایی یه كه م كرد كه ( air ) ه , دوا ی بوم دهركهوت كه گیانداران پیویستیان به (ههوا) یه بو هه ناسه دان.

تویژەر: كه واته تو له سه ر نه ساسی وشه ی دوا ی بوشاییه كه نه وشه یه ت هه ئبژارد .

خویندکار: به ئی

تویژەر: نه ی دهرباره ی برگه ی دووم، بیرت نه چی دهردهوه كاتیك نه م ئو پشنه ت هه ئبژارد ؟

خویندکار: له راستیدا نازانم ( temper ) مانای چی ، بویه سه یری وشه ی ناو رسته كه م كرد بینیم كه ( calm ) هه یه كه مانای ( له سه ر خو ) ، بویه ( lose ) م هه ئبژارد چونكه هه ستم كرد كه هه ركه سیك ( calm ) بیته نه وا هه رگیز شه ت له ده ست نادات .

تویژەر: ده توانیت پیم بلیی بوچی نه م برگه یه دا ( look ) ت هه ئبژارد؟

خویندکار: چونكه له م رسته یه دا كه سیك داوا له دراوسیكه ی ده كات كه چاویکی له مائه كه بیته كاتیك كه له مال نیه .

تویژەر: وهكو ده بیینی له و برگانه ی خوارتر تو ( circular ) ، ( exact ) و ( near ) ت هه ئبژاردوو، ده توانی هوكاری نه

وهه ئبژاردنانه م پی بلییت ؟

خویندکار: باشه ، له و برگه یه من ( circular ) م هه ئبژارد چونكه بینیم كه ( face ) له دوا ی بوشاییه كه هه یه بویه بوم دهركهوت كه نه ورسته یه باسی كه سیك ده كات كه ده موچاوی خره . بو برگه ی دواتر ( exact ) م هه ئبژارد چونكه له دوا ی بوشاییه كه وشه ی ( rules ) هه یه كه مانای ( یاساکان ) دیت بویه یاساکان پیویسته ده قیق بن . دهرباره ی نه و برگه یه كه ده بیته ( on the day of our departure all our ..... Friends etc. ) من ( near ) م هه ئبژارد چونكه به گویره ی رسته كه یه كیک هه یه كه نه وشینه به جیده هیلیت و ( هاورى نزیكه كانی ) له گه ئی دان بوخوا حافیزی كردن .

تویژەر: با سه یری نه و برگانه بکهین كه داوا ی نامرازی په یوهندی دروست ده كهن، بوچی نه و نامرازانه ت هه ئبژاردن ؟

خوینکار: له راستیدا من زور باش نیم له دانانی نامرازی دروست، له به رنه وه، زوربه ی نامرازه كانم هه رنه خووه هه ئبژارد. به لام له هه ندیکیان دنیا بوم كه راستن وهكو ( upon ) ، ( about ) وه ( from ) .

تویژەر: چونت زانی نه وانه راستن ؟

خوینکار: چونكه له رسته كاندا کاری ( insist ) ، ( complain ) و ( differ ) كه وتبونه پیش بوشاییه كان، وه نامرازه كانی ( upon ) ، ( about ) وه ( from ) م زور به یه كه وه بینیه له گه ل نه و کارانه .

تویژەر: زور سوپاس بو کانت.

خوینکار: شایه نی نیه .